

Bias Corrected Estimates of GED Returns

James J. Heckman and Paul A. LaFontaine¹

First draft, March 2003

Final draft, March 22, 2005

¹This project was supported by the Mellon Foundation, the Joyce Foundation, the Pew Foundation and NICHD R01-34598-03. We discussed our findings with Barry Hirsch at Trinity College, March 2003. The website for this paper is http://jenni.uchicago.edu/ged_imputation.

Abstract

Currently, over 30% of all wage earners refuse to answer one or more income related questions in the Current Population Survey (CPS) Monthly data. The CPS attempts to correct for any resulting non-response bias by allocating wages for non-respondents through a “hot deck” imputation procedure. The “hot deck” method matches wage non-respondents to wage respondents according to certain demographic and labor force characteristics. This paper shows that this imputation procedure can lead to serious bias in estimating GED returns using CPS data. Although all estimates of returns to education are affected, we focus on estimating the returns to General Educational Development (GED) certification. The imputation bias arises from the limited number of educational categories used to impute wages and is independent of any potential non-response bias which may also be present. After correcting for this bias in various ways, we show that the CPS data yields estimated returns to GED certification similar to those previously found in Cameron and Heckman’s (1993) study using the National Longitudinal Survey of Youth (NLSY) data which do not suffer from imputation bias. We further show that all positive wage returns to GED certification in the CPS appear to be from an unobserved ability bias. This holds true for both the native and foreign born, as well as older versus more recent birth cohorts.

JEL Code: C61

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
jjh@uchicago.edu
fax: 773-702-8490

Paul A. LaFontaine
Center of Social Program Evaluation
Harris School
1155 E. 60th Street
Chicago, IL 60637
klmpaul@lily.src.uchicago.edu

1 Introduction

There has been rapid growth in the fraction of persons who achieve high school certification by means of an equivalency exam rather than through the traditional route of classroom attendance and high school graduation. The primary vehicle for high school equivalency certification is the General Educational Development (GED) program. In 1960, only two percent of all new high school certificates were awarded through equivalency exams in the United States. By 2001, over twenty percent of all new high school credentials were produced through GED certification (See Figure 1).

This rapid growth in exam certification occurred despite apparently low direct economic returns to it. Using data from the National Longitudinal Study of Youth (NLSY), Cameron and Heckman (1993) find that, controlling for differences in ability, male exam-certified high school equivalents are statistically indistinguishable from high school dropouts who are uncertified in terms of their hourly wages. Any differences in wages among exam-certified equivalents and uncertified dropouts are completely accounted for by differences in ability. There is no demonstrable causal effect of GED certification on wages.¹ Cameron and Heckman conclude that whatever economic return there is to GED certification must come through the access to further post-secondary education and training that certification provides. However, GEDs are much less likely than ordinary high school graduates to complete two or four year colleges. A large body of subsequent work, summarized in Boesel, Alsalam, and Smith (1998), confirms that GED recipients and high school graduates are not equivalent.

Advocates of the GED testing program raised some potentially valid criticisms of the Cameron and Heckman analysis following its publication (Murnane, Willett, and Boudett, 1999; Boudett, 2000; Jaeger and Clark, 2002). First, their research only considered labor market outcomes at ages 25 and 28. If GED certification opens up access to occupations that are closed to high school dropouts then the effect of certification may not manifest itself until later in the life cycle. A second concern was the small sample sizes available in the NLSY data. Some argued that it would not be possible to assess the entire GED program based on a few hundred NLSY participants.

¹Later work by Cameron (1994) found similar results for NLSY females.

Finally, there may be a disparate impact of the GED program across different race groups or other subpopulations. For instance, a GED may send a different signal for recent immigrants who acquire the credential than it does for native born dropouts. This paper addresses these questions.

In 1998, the Current Population Monthly Survey (CPS) began distinguishing between the two types of high school completion statuses. The large sample sizes for various racial and ethnic groups, as well as the wide range of available ages, appear to make the CPS ideal for addressing some of the limitations of the Cameron and Heckman analysis. However, four potentially serious problems and limitations plague these data. First, the CPS contains no measure of ability. Cameron and Heckman found that the GED program is selective because higher ability dropouts obtain GED certification. Once differences in ability between GED recipients and uncertified dropouts are accounted for, wage differentials disappear. Second, as found by Hirsch and Schumacher (2004) in the context of estimating union-nonunion wage differentials, “match bias” can result from the CPS method of imputing missing wages. We find that the estimated returns to GED certification are substantially upward biased because GED respondents who either refuse or fail to report their wage information are frequently assigned (matched to) the wages of traditional high school graduates. Third, CPS data show that a large fraction of workers have no reported earnings because they are unemployed or out of the labor force. Finally, potential bias may arise from low and high income earners refusing to report earnings.

This paper addresses the first three of these problems. We show that when estimation is performed carefully, the returns to GED certification and other educational estimates using CPS data are similar to those obtained from other, cleaner, data sources. We find that GED recipients who do not continue on to college earn the same wages as uncertified dropouts after correcting for differences in ability. This result applies to both males and females across the age spectrum. We find no evidence of post-certification life cycle wage growth attributable to the program. The apparent return to GED certification for older age groups in the raw data is due to a greater unobserved ability bias for older birth cohorts rather than from a causal effect of GED certification. After correcting for problems with the CPS data, the estimated GED-dropout difference in wages is the same in comparable NLSY and CPS cohorts. The positive wage returns to GED certification found in unadjusted CPS data arise from unobserved ability bias and improper allocation of GED missing

wages. We also show that ability bias is greater when comparing foreign born GED recipients and foreign born dropouts. After adjusting for ability, no statistically significant effect of the GED on wages is discernible for both native and foreign born males and females of all race and ethnic groups.

The plan of this paper is as follows. In section 2, we discuss the CPS and compare evidence from it with evidence from the NLSY. In section 3, we discuss the issues of age and cohort effects. In section 4, we replicate NLSY-CPS results by using data from the National Adult Literacy Survey (NALS). In section 5, we consider GED returns among immigrants. Section 6 concludes.

2 The Importance of Wage Imputation and Non-response

We use the monthly outgoing rotation groups from the CPS for the period January 1998 to December 2003. Our sample consists of civilian males and females age 20-64 who are either in their fourth or eighth month in the sample. We use a sample of dropouts, GED recipients and high school graduates who have completed no college along with a sample of four year degree holders for whom we cannot determine what type of high school certificate they hold. For our wage analysis, we exclude those people who are enrolled in school; are self-employed; reported their ethnicity as Native American, Aleut, or Eskimo; or had their education status or years of schooling responses imputed. The self-employed are excluded because earnings are not available for these individuals. All regressions also exclude those who earn less than \$.50 or more than \$200 an hour (in 2000 dollars). Data loss due to these exclusions are listed in Table 1(a). The main exclusion is due to self employment status, while other sample restrictions only account for a small fraction of lost data.

2.1 CPS Problems and Limitations

Due to its large sample sizes, the long period over which it is collected, and its perceived quality, the CPS has become the primary data source for understanding a host of important economic issues, including the U.S. earnings structure, racial wage gaps and returns to education. The growing non-response to income related questions calls into question the quality of the data and its comparability across time. Figure 2 shows that prior to 1994 the percentage of those who chose not to report earnings was relatively stable at around 15%. After 1994, earnings non-response rose from a low of

24% in 1995 to nearly 34% in 2003.²

Increasing rates of non-response, greater numbers of workers selectively withdrawing from the labor force, and the CPS practice of not collecting wage information from the self-employed, have resulted in substantial fractions of respondents with missing wage data among certain race, sex and age groups. Table 1(b) reveals that only about 50% of White and Hispanic males in each outgoing rotation group report earnings due to the combination of these factors. Wage data for Black males are only available for around 38% of the sample due to higher rates of income non-response among the employed and higher incidence of unemployment among this population. The situation is worse for women, due to their lower labor force participation rates.

Unlike the NLSY data that surveys each person individually, the CPS questionnaire is administered by telephone to one person who responds for his or her entire household. Potentially exacerbating the non-response problem is that the accuracy of the available wage information may also be questionable. For males, over 60% of the wage and labor force information is given by a proxy respondent and these respondents may not be privy to all income related information. The amount of available self-reported wages is extremely low- only around 25% for males and 30% for females. The propensity to report earnings varies across race groups. Specifically, black males and females are 10% more likely not to report earnings than either their white or Hispanic counterparts.

Unfortunately, the CPS survey does not provide enough information to determine the nature of this response bias. We present some evidence on the severity of this potential bias using NLSY data. Non-response bias may not be large since the estimates obtained from CPS data closely track those estimates from cleaner data sources.

2.2 CPS Imputation Strategy

To avoid computing national statistics based on a sample with a large proportion of missing data, and in an attempt to correct for possible non-response bias caused by missing wage data, the CPS allocates missing earnings using a “hot deck” imputation method. A hot deck assigns the wages of

²The dramatic increase in allocation after 1994 is primarily due to the implementation of the newly redesigned CPS questionnaire. The new questionnaire asks a longer, more complex series of questions in order to determine weekly wages, and the new data processing procedures set weekly wages to missing if even one of these questions is met with either a refusal or “don’t know” response.

respondents to non-respondents based on a limited set of demographic, education and occupational characteristics.³ A common practice among researchers is to treat allocated values as observed when using CPS survey data. In a widely cited paper, Angrist and Krueger (1999) claim that CPS wage allocation is empirically unimportant. This paper shows that CPS allocation methods and the resulting match bias are of first order economic importance in estimating returns to GED certification.

“Match bias”, a phrase due to Hirsch and Schumacher (2004), arises from the limited number of categories used to impute non-respondent wages. Of particular interest to this paper, the matching of wage non-respondents to wage respondents is based on only three levels of educational attainment: high school dropouts, high school graduates with up to but not including a Bachelors degree, and those with a Bachelors degree and above. Given these education categories, it is clear that estimated returns for those who graduate high school and do not attend college tend to exhibit upward bias since non-respondents will frequently be matched to those who complete some college. On the other hand, estimated returns for those who complete above a Bachelors degree should be biased downward as a result of non-respondents being assigned the wages of those with only Bachelors degree. Clearly, all CPS educational estimates will be affected by this type of educational mismatching within allocation cells. Bollinger and Hirsch (2005) and Hirsch and Schumacher (2004) present a more detailed discussion of the CPS hot deck procedure and the resulting attenuation bias in estimates among various educational categories.

GED allocated wages exhibit a particularly severe form of this type of misallocation bias since non-respondents who hold GED credentials are frequently assigned the wages of high school graduates who may have post-secondary education up to but not including a Bachelors degree. If wage differentials exist between GED recipients and high school graduates, then this differential will tend to zero as the proportion of GED non-respondents increases. As non-response has grown from less than 15% to over 30% in recent years, the upward bias in GED estimated returns has increased proportionally.

Table 2(a) shows that for native males the estimated return to GED certification is overstated

³Currently, the CPS matches non-respondents to respondents in the monthly data based on the following categories: gender (2), race (2), age (6), occupation (13), hours worked (8), education (3), and tips and overtime receipt (2).

by over 35% when CPS allocated wages are included in the sample. After dropping these poorly imputed wages, the estimated return to GED certification drops from .14 log points to .09. For females, shown in Table 2(b), the bias tends to be generally smaller in magnitude but is still over 25%. The estimated return decreases from .15 log points to just under .11 for the full sample of females. As predicted, excluding allocated earners also decreases estimated returns to high school graduation and college completion. However, this decrease is not of the same magnitude that we find for GED recipients. The resulting reduction for the full sample of males is just over 5% for college graduates and just under 12% for high school graduates who did not attend college. The observed effects of CPS allocation for the female sample are similar. Overall, imputation tends to increase the estimated college-dropout and high school-dropout wage differentials and leaves the college-high school differential largely unaffected. The most serious bias is observed in the GED category.

Tables 2(a) and 2(b) show that the returns are different across racial, sex and ethnic groups although not dramatically so. Returns to certification are always higher for females compared to males, and minorities have higher returns than whites. Both Hispanic males and females show the highest returns to GED acquisition among all racial groups. However, the differences across groups are not dramatic. The largest estimated difference between pooled and separated race estimates is only .04 log points.

In order to assess how sensitive these estimates are to non-response and match bias, we implement a hot deck imputation procedure that differs from the CPS hot deck only in that it matches using more precisely defined educational groups. This is done both to show that it is the exclusion of GED status as a match criterion in the CPS hot deck that causes the match bias and to correct for possible non-response bias in our final estimates.

We impute wages using the CPS hot deck with an added GED educational category. In order to account for the uncertainty associated with the imputed wage estimates of non-respondents, we use the bootstrapping algorithm of Shao and Sitter (1996). This procedure produces unbiased estimates of standard errors by re-imputing missing wages for the bootstrap replicates.

The last columns of Tables 2(a) and 2(b) show that the results obtained from either reallocating wages or dropping those who do not report earnings are nearly identical. This is entirely consistent

with the findings of Hirsch and Schumacher who show similar results comparing the wages of union and nonunion workers. This is not conclusive evidence that non-response is not present but it is strongly suggestive.

The primary focus of this paper is on the false conclusions that can be reached regarding the value of a GED credential using CPS data when CPS allocations are included. For the rest of this paper we choose the most expedient method by dropping employed workers who do not report earnings rather than imputing missing wages. Due to the richer set of conditioning variables available in the NLSY compared to the CPS, we correct NLSY-based estimates for sample selection bias using both parametric and semiparametric selection correction models described in more detail in the next section.

2.3 Using the NLSY to Assess CPS Ability Bias

Even though the exclusion of allocated earners dramatically reduces the size of the estimated return to a GED credential, the resulting return relative to dropouts is both positive and statistically significant for both males and females across all race groups. Cameron and Heckman (1993) found that positive returns to GED certification could be attributed entirely to ability bias. Those who choose to take the GED examination are a select group from the dropout pool. The distributions of measured ability of the people who choose to take the GED and those who do not are very different. The CPS data do not include any measures of ability. Unobserved ability may be driving the observed wage differences between education categories.

In 1980, the Armed Services Vocational Aptitude Battery (ASVAB) was administered to all NLSY respondents, with a completion rate of about 94% for the sample. We use the AFQT test score as our measure of ability.⁴ Figures 3 present the distributions of AFQT scores by education and race for the NLSY. The differences in ability between GED recipients and dropouts for both males and females of all races are large and statistically significant.⁵ In fact, GED recipients have

⁴The ASVAB consists of a battery of ten tests: general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension, and electronics information. The Armed Forces Qualification Test (AFQT) is the sum of word knowledge, arithmetic reasoning, paragraph comprehension, and numeric operations components of the ASVAB and is a general measure of trainability used by the military for enlistment screening and job assignment.

⁵Wilcoxon rank sum tests of stochastic dominance show strong differences.

nearly the same measured ability as high school graduates who do not continue on to college across all races.

In order to determine the potential importance of ability bias in generating the estimated returns to GED certification using CPS data, we compare CPS estimates to those obtained in the NLSY both including and excluding the AFQT score. Tables 3(a) and 3(b) show that the estimated returns to certification across race groups using NLSY data, when respondents are between 20 and 39 years of age, are similar to those we obtained from the CPS. The exception is for black males—where the CPS generated estimate is much higher. Returns to GED certification are also positive and, in all but one case, statistically significant across all race groups using standard significance levels. However, when the AFQT score is added to the model the estimated GED effect is zero for males. The estimated effect for females is still slightly positive across all race groups but is always statistically insignificant. Again, we see that all wage differentials between GED recipients and dropouts can be accounted for by including ability in the equation. The positive estimates obtained in the CPS are similarly due to an unobserved ability bias that results from high ability dropouts self-selecting into the GED program.

To test the robustness of the NLSY results to potential sample selection bias problems that may arise from excluding workers on the basis of their labor force status, we estimate a parametric selection correction model due to Heckman (1979).⁶ As shown in the last columns of Tables 3(a) and 3(b), accounting for selective participation in the workforce does not overturn our conclusion that GEDs are paid the wages of high school dropouts at the same ability level.

3 Cohort Versus Age Effects

Proponents of the GED program argue that a GED title may confer little initial benefit but that after time, GED holders will experience higher wage growth than dropouts who do not certify. If this hypothesis is true we would expect to see higher wage differentials between GED recipients and high school dropouts at older ages. Tables 4(a) and 4(b) shed light on this question by estimating the return to GED certification in the CPS by age groups for white males and females. We focus

⁶At our website, we report estimates based on a semiparametric factor model structure (see Carneiro, Hansen, and Heckman, 2003). The parametric and semiparametric estimates agree (see tables A1–A4).

on whites because the minority comparison samples in NALS and the NLSY are too small.⁷ We consider only GED recipients who get no further education. For white males, we find evidence that apparently supports this claim. GED recipients in each successive age category have higher estimated returns to certification than the previous one. For white females, the pattern of returns is quite different, being nearly constant across age groups.

It is not clear whether the higher returns to GED certification at older ages are due to age or cohort effects. It is not possible to answer the age vs. cohort question using cross-sectional data such as the CPS (see Heckman and Robb, 1985). It may be that the acquisition of the GED title causes the wage differential to increase between male GED recipients and dropouts at older ages or it may be that older birth cohorts exhibit higher returns due to unobservable differences in quality between GED recipients and dropouts that are not present in more recent birth cohorts. Comparing CPS to NLSY data and data from the National Adult Literacy Survey (NALS) discussed further below, we find that higher estimated returns for older groups are due to differences in cohorts and not increased wage growth resulting from GED acquisition.

By comparing GED estimates of the NLSY cohort in the CPS to estimates reported by Cameron and Heckman at younger ages, Jaeger and Clark (2002) claim to find evidence of strong GED life cycle wage growth. They report that estimated returns to GED certification in the monthly CPS data for the NLSY cohort—those born between 1957 and 1964—far exceed the estimates reported at age 25 and 28 in Cameron and Heckman’s analysis. They conclude that by the time GED recipients are in their late 30’s to early 40’s, the GED title has helped them “catch up” to high school graduates and to far exceed the wage growth exhibited by high school dropouts who do not exam certify.

Tables 4(a) and 4(b)⁸ show that this conclusion arises as an artifact of inclusion of allocated earners in the Clarke and Jaeger samples. After excluding those who do not report their earnings, the CPS estimated GED returns for the NLSY cohort, at comparable ages of 35–46, are nearly identical to the estimates obtained from the NLSY when the sample is in their 20’s and again in their 30’s.

⁷The estimates for minorities are consistent with those for whites but the cells are small and the standard errors are large. See tables A5 and A6 in the table appendix on the website for these results.

⁸The log hourly wage regressions in the NLSY and CPS comparisons include similar covariates and are based on the same sample restrictions to make the estimates comparable.

Both data sources show that GED recipient wage growth is not stronger than that exhibited by high school dropouts. Furthermore, the positive wage differences between GED recipients and uncertified dropouts is completely accounted for by the inclusion of the ability measure for males and females of all ages.

Tables 5(a) and 5(b) strengthen this conclusion by comparing male and female estimates of the CPS-NLSY cohort with cross sectional estimates obtained from the NLSY sample at ages 25, 28, 30, 35 and 38. We again see that the estimated returns to GED certification and high school graduation for this cohort are remarkably similar between the two data sources and across ages. The estimated GED-dropout difference at ages 35 and 38 are no different than those previously found by Cameron and Heckman at ages 25 and 28. According to official published statistics from the GED testing service, over 75% of GED recipients acquire the degree before the age of 25. Therefore, the majority of the wage sample at 35 and 38 have had their diplomas for over 10 years, ample time for any positive net benefits to accrue. If GED recipients have not shown positive wage growth within ten years of obtaining the title it is highly unlikely that they will do so later. Both the NLSY and CPS data strongly reject the hypothesis of post-certification life cycle wage growth posited by Clarke and Jaeger as well as Murnane, Willett, and Boudett (1999); Boudett (2000), once match bias is accounted for and estimation is made on comparable cohorts.

4 Evidence on Cohort Effects

Controlling for ability differences in the NLSY data produces no statistically significant differences in wages between GED recipients and dropouts who do not certify for both males and females at all ages. It is possible that the differences in wages between GED recipients, high school graduates and dropouts observed in the CPS can be completely accounted for by unobserved ability differences as well. Given that the NLSY cohort shows little life cycle wage growth, it is also plausible that the higher returns to GED certification seen for older birth cohorts in CPS data are due to a growing difference in this ability bias between GED recipients and dropouts. Two not necessarily mutually exclusive possibilities may explain the data. The first is that as the GED program has expanded rapidly over the last 30 years, the quality of GED recipients may have declined. Second, the quality

of dropouts may have improved. Figure 4 supports the hypothesis that the quality of dropouts has improved across cohorts while GED quality has remained roughly constant. Male and female dropouts of all races have obtained greater levels of schooling while GED completed secondary schooling levels are nearly constant across all birth cohorts. The greater schooling attainment of dropouts may indicate that the skill gap between GED recipients and dropouts is closing across cohorts or it may be the consequence of social promotion. Both factors may be at work. We now turn to the National Adult Literacy Survey (NALS) data to explore this issue.

4.1 NALS Data

The National Assessment of Literacy (NALS) is a decennial survey administered by the NCES to a random sample of the U.S. adult population to determine their literacy skills. The 1992 sample used in this section consists of a random sample of 13,600 adults age 16 and over and a state supplement of 11,344 adults. The NALS testing battery consists of three separate tests designed to measure three types of skills: prose, document, and quantitative skills. Unlike the CPS, the NALS sample does not ask respondents to report their hours of work. Therefore all comparisons between CPS and NALS data are based on weekly wage regressions. These regressions are restricted to those individuals who: have weekly wages less than \$100 or more than \$4000 (2000 dollars); are younger than 20 years of age or older than 64; or are Aleut, Eskimo or Native American. Controls for central city status, married with spouse present, year of survey, region of residence, a quadratic in age and race dummies, where appropriate, are included in each regression.⁹

4.2 NALS Test Scores

As measured by the NALS test scores, people who choose to take the GED test are as capable in their basic cognitive skills as high school graduates and more capable than high school dropouts who choose not to certify. Figure 5 shows the distributions of total NALS test scores, derived from the average over all three components of the NALS battery, by race and education status for the native born. The distributions of NALS scores for high school graduates and GED recipients are

⁹The amount of data lost due to these exclusion restrictions for the NALS sample is comparable to data loss generated from similar restrictions on the CPS sample (See Figure 1(b)).

nearly identical across all races, while dropouts have lower scores. In terms of basic skills, the GED exam effectively sorts between those who pass the exam and those who do not.

Since the gap in years of schooling completed between dropouts and GED recipients is narrowing across birth cohorts, we might expect to find the cognitive skill gaps between the groups to be narrowing as well. Figure 6, which presents NALS score distributions across different birth cohorts, shows that this is indeed the case. The distributions of scores for GEDs have remained nearly identical to those of high school graduates across all birth cohorts for males and females. As dropouts have obtained more years of schooling, their test score distributions are becoming more similar to GEDs across birth cohorts, but they are still statistically significantly different, even in the most recent cohort. This result not only shows that the growing ability bias across cohorts needs to be controlled for in wage estimation, but also that the rise in GED certification may be due in part to diminishing participation costs of preparing for the exam by uncertified dropouts. Whereas passing the GED examination for a 6th grade high school dropout in 1930 would have required substantial investment and skill acquisition, the average dropout from today's public school system with 10 years of education may require only minor preparation to pass the exam.

4.3 OLS Estimation

The returns to GED certification found in the NALS92 sample for males and females ages 20–64 closely match those found in the CPS 1998–2003. Tables 6(a) and 6(b) show that male GED recipients have 6.6% higher weekly wages than dropouts before controlling for ability. Female GED recipients earn 9.4% more than dropouts. However, these positive returns to certification are completely accounted for when we control for the NALS test score. As with the male NLSY sample, GED recipients earn less than dropouts at the same level of ability. Once again, this effect is not statistically significant. Female GED recipients show a small but statistically insignificant positive return to certification adjusting for ability, much as we saw in the NLSY data. It is evident that the lack of ability proxies in CPS data overestimate wage returns to GED certification and that all positive returns to certification can be completely accounted for by selection into the GED program based on ability.

The NALS distribution of test scores across birth cohorts shows that the ability differential between GED recipients and dropouts is diminishing. In a cross-section, this results in the pattern of wage estimates across ages observed for males in the CPS. Older age groups show a higher return to certification. This is a spurious difference due solely to a greater ability gap between GEDs and dropouts in earlier cohorts.

Table 7 makes this point clearly by comparing estimated weekly wage returns in both the CPS and NALS for two birth cohorts. The first is the pre-NLSY cohort (those born before 1957), and the second includes the NLSY cohort (those born 1957–1964) and those born afterwards. Once again, we see the pattern of higher returns for the older cohort in both the NALS and CPS data. However, controlling for the NALS test score, both male and female GED recipients across all birth cohorts show no statistically distinguishable wage benefits from certification. The available evidence suggests that the GED program has always selected the most able from the dropout pool and that the direct wage benefits across all certification cohorts is small to non-existent once this selection on ability is accounted for.

5 GED Returns Among Immigrants

Jaeger and Clark (2002) argue that the GED has an even greater signaling effect for immigrants than for the native born. Their study does not control for either the cognitive differences between education groups or the amount of education GED recipients obtained in their home countries. It is possible that the GED program is even more selective among the immigrant population than it is for natives, so that only the most able immigrants with higher skills GED certify. Failure to control for these factors would cause an even wider disparity between the GED and dropout literacy and cognitive distributions than is found in native born populations which would result in a higher perceived return for this sub-population if these differences were not accounted for in estimation. Figure 7 reveals that the distributions of literacy levels for foreign-born dropouts, GED recipients, and high school graduates are dramatically different. While GED recipients and high school graduates are nearly identical in terms of literacy, immigrant dropouts have extremely low literacy and quantitative skills. In fact, foreign born high school graduates and GED recipients more

closely resemble their native born educational counterparts in literacy than foreign born dropouts. These vast differences in basic skills among foreign born educational groups call into question the comparability of wage returns between them, since the types of jobs available to them will be very different as well. This evidence suggests that it is even more important to adjust for literacy and cognitive skill differences among the foreign born in order to accurately determine the value of an immigrant GED credential.

Immigrants who take the GED also come into the country with higher levels of completed schooling in their home countries than immigrants who do not take the GED. Table 8 shows that GED recipients and high school graduates are far more likely than dropouts to have attended secondary schooling in their native country. The majority of immigrant dropouts complete elementary school or less. Both high school graduates and GED recipients are also more likely to have been schooled solely in the U.S., as evidenced by the percentage who did not attend school before arriving in the U.S. GED recipients also have the highest probability of entering the country having completed a post-secondary vocational training program. All of these factors point towards the possibility that the GED program is even more selective among immigrants than for natives and that large wage differences exist between foreign GED recipients and foreign dropouts before they certify.

5.1 OLS Estimation

We now present CPS and NALS estimates of the returns to GED certification among the foreign born. We estimate the same regression model as is used to analyze the native-born except that we also add controls for country of birth, citizenship status, and cohort of entry into the United States. Table 9 shows that the CPS match bias that results from matching foreign born non-respondent GED recipients and high school graduates to native wage donors by the hot deck overstates the value of both degrees by about .05 log points for males and .06 log points for females. In contrast to the results for the native born, we cannot reject the hypothesis that GED certification is equivalent to high school graduation for both males and females. The data reject the null hypothesis that there are no direct wage benefits of obtaining a GED compared to staying in the dropout state.

The positive estimated returns to GED certification among the foreign born in the CPS appears

to be driven by unobserved ability bias. We have already established that, in the NALS data, GED recipients and those dropouts who choose not to certify have very different skill distributions. Table 9 shows that these differences account for all differences between GED recipients and uncertified dropouts and that the positive wage returns to certification estimated in CPS data are spurious due to selection on ability.

Given the small immigrant sample available in the NALS data, we must be cautious in drawing any firm conclusions about the value of GED certification among the foreign born. However, the evidence is strongly suggestive that those immigrants who choose to GED certify are very different from those who do not and that any study of the value of GED certification among this population needs to be able to account for this selection.

6 Conclusion

This paper shows the importance of accounting for the CPS hot deck procedure in order to obtain unbiased estimates of the return to education using CPS data. Misallocation of non-respondent GED recipients to high school status results in a sizeable overestimate of the value of GED certification. This bias is not due to non-response bias and is even higher among certain populations such as the foreign-born. Correcting for match bias is important in order to have conceptually comparable estimates of the returns to the GED across different data sources. Researchers should pay closer attention to how missing wages are allocated. Alternative allocation procedures may dramatically affect their conclusions. The importance of this warning is highlighted by the low direct wage returns and zero life cycle wage growth that we find for GED certification, in contrast to the evidence presented by Jaeger and Clark (2002) using a biased sample.

Our evidence suggests that direct returns to GED certification are low. Selection into the GED program by agents based on cognitive ability can account for all wage differentials between those dropouts who do not certify and those who choose to do so. This gap in cognitive skills appears to be greater for older birth cohorts and it is this greater ability bias that causes the higher estimated returns for older ages in CPS data. No empirically significant life cycle wage growth can be attributed to the GED title itself. These cognitive skill differences also account for the positive

effects found for GED certification among immigrants in the CPS. This evidence highlights the importance of using data with a rich set of family background and cognitive conditioning variables in order to evaluate the true impact of social programs. While the CPS data provide a foundation from which to begin an analysis of the GED program, it cannot be considered a definitive data source. For this reason, we are currently engaged in a deeper analysis of NLSY data and other data sources, to determine the treatment effects of GED certification among different groups, and to expand on the analysis of differences in GED certification across cohorts reported here.

While the direct benefits of GED certification appear low, there may still be an economic value to GED certification in opening post-secondary schooling and training opportunities. However, all available evidence suggests that GED certification for those who do not obtain post-secondary schooling has little or no direct effect on wages among men, women, older and more recent cohorts, and the foreign born. All perceived differences between GED recipients and dropouts who do not certify can be accounted for by cognitive skill differences, and these are highly correlated with schooling. What is true today was true 60 years ago when the GED program first started. There are no cheap substitutes for classroom instruction and training and the benefits they afford.

References

- Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, pp. 1277–1366. New York: North-Holland.
- Boesel, D., N. Alsalam, and T. M. Smith (1998). *Educational and Labor Market Performance of GED Recipients*. Washington, DC: U.S. Dept. of Education, Office of Educational Research and Improvement, National Library of Education.
- Bollinger, C. and B. Hirsch (2005). Match bias due to earnings imputation: When does it matter? Unpublished manuscript, Trinity University, Department of Economics.
- Boudett, K. P. (2000, December). “Second chance” strategies for women who drop out of school. *Monthly Labor Review*, 19–31.
- Cameron, S. (1994). Assessing high school certification for women who dropout. Unpublished manuscript, University of Chicago, Department of Economics.
- Cameron, S. V. and J. J. Heckman (1993, January). The nonequivalence of high school equivalents. *Journal of Labor Economics* 11(1, Part 1), 1–47.
- Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44(2), 361–422. 2001 Lawrence R. Klein Lecture.
- Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica* 47(1), 153–162.
- Heckman, J. J. and R. Robb (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations. In W. M. Mason and S. E. Fienberg (Eds.), *Cohort Analysis in Social Research: Beyond the Identification Problem*. New York: Springer-Verlag.

- Hirsch, B. T. and E. Schumacher (2004, July). Match bias in wage gap estimates due to earnings. *Journal of Labor Economics* 22(3), 689–722.
- Jaeger, D. and M. Clark (2002, April). Natives, the foreign-born, and high school equivalents: New evidence on the returns to the ged. Technical Report 462, Princeton University Industrial Relations Section Working Paper.
- Murnane, R. J., J. B. Willett, and K. P. Boudett (1999, October). Do male dropouts benefit from obtaining a GED, postsecondary education, and training? *Evaluation Review* 22(5), 475–502.
- Shao, J. and R. R. Sitter (1996, September). Bootstrap for imputed survey data. *Journal of the American Statistical Association* 91(435), 1278–1288.

Table 1(a): Exclusion Restrictions by Data Source

	Native Males			Native Females			Foreign Males		Foreign Females	
	CPS	NLSY	NALS	CPS	NLSY	NALS	CPS	NALS	CPS	NALS
Potential Observations	352,858	55,057	5,412	371,222	54,101	7,058	65,004	821	68,688	886
Not Working	64,302	12,358	872	117,363	19,873	2,306	10,061	109	29,377	354
Working and Enrolled	1,681	1,612	311	2,227	1,862	425	305	60	251	59
Self Employed *	40,311	3,334	0	21,064	2,107	0	5772	0	2,921	0
Other Race	3,065	0	30	2,761	0	38	124	...	128	...
Zero Years of Education	385	17	0	280	34	0	886	0	482	0
Imputed Education	988	0	0	780	0	0	298	0	166	0
Earnings Outliers	286	130	137	298	81	380	61	26	36	48
Total Observations	239,400	37,961	4,106	225,517	30,621	3,952	47,295	629	35,174	429
% Not Working	.182	.224	.161	.316	.367	.327	.155	.133	.428	.400
% Working and Excluded	.170	.111	.096	.112	.105	.168	.139	.117	.105	.194

Note: The total excluded observations is not the sum of the column since many individuals fall into multiple categories. Calculations based on a sample of employed dropouts, GED recipients and high school graduates with no college plus four year college graduates. The sample ages are: 20-64 for the CPS; 20-39 for the NLSY; and 20-64 for the NALS.

*It is not possible to determine years of schooling or self employment in the NALS data

Table 1(b): Sources and Extent of CPS Missing Wage Data by Race for the Full Sample

	White Males				Black Males				Hispanic Males				Foreign Males
	20-29	30-39	40-49	50-59	20-29	30-39	40-49	50-59	20-29	30-39	40-49	50-59	
Potential Wage Obs.	95,928	122,760	140,283	109,744	12,956	14,819	15,032	10,654	21,697	20,714	14,283	8,049	87,019
Unemployed	4,580	3,529	3,748	2,648	1,311	846	763	392	1,286	884	561	323	3,523
OLF	11,003	6,734	10,341	16,999	2,851	1,891	2,663	2,947	2,109	1,366	1,375	1,509	8,696
Self Employed	4,132	14,577	22,713	19,102	265	749	834	719	611	1,331	1,288	714	8,493
Military	1,595	2,006	854	221	239	330	114	17	238	154	65	6	493
Non-Response	21,753	27,912	33,282	24,692	3,384	4,433	4,790	3,087	5,034	4,781	3,367	1,875	21,714
Wage Observations	52,865	68,002	69,345	46,082	4,906	6,570	5,868	3,492	12,419	12,198	7,627	3,622	44,100
% Reporting Wages	.551	.554	.494	.420	.379	.443	.390	.328	.572	.589	.534	.450	.507
Proxy Responses	30,731	37,080	37,766	24,655	1,780	1,794	1,918	1,547	8,046	7,133	4,445	2,110	25158
% Self Reporting Wages	.231	.252	.225	.195	.241	.322	.263	.183	.202	.245	.223	.188	.218
	White Females				Black Females				Hispanic Females				Foreign Females
	20-29	30-39	40-49	50-59	20-29	30-39	40-49	50-59	20-29	30-39	40-49	50-59	
Potential Wage Obs.	99,672	128,051	145,846	113,599	17,926	20,086	19,987	13,893	20,726	20,741	15,244	8,826	89359
Unemployed	3658	3210	3,139	2,006	1,650	1,173	855	341	1,115	966	600	240	3236
OLF	21677	28971	28,539	32,172	4,620	3,932	4,389	4,687	7,513	6,786	4,422	3,684	30779
Self Employed	2727	8900	11,819	9,710	216	515	564	407	276	634	686	441	4540
Military	188	116	86	11	59	53	31	0	23	10	9	0	49
Non-Response	18930	23214	30,358	22,277	4,023	5,682	6,047	3,741	3,111	3,295	2,887	1,411	16784
Wage Observations	52,492	63,640	71,905	47,423	7,358	8,731	8,101	4,717	8,688	9,050	6,640	3,050	33,971
% Reporting Wages	.527	.497	.493	.417	.410	.435	.405	.340	.419	.436	.436	.346	.380
Proxy Responses	23,274	19,702	21,934	14,447	2,337	2,114	2,101	1,145	3,981	3,281	2,553	1,259	14983
% Self Reporting Wages	.293	.343	.343	.290	.280	.329	.300	.257	.227	.278	.268	.203	.212

Based on CPS 1998-2003 Monthly outgoing rotation groups. Potential wage obs. are those people in their 4th or 8th month in samples who are in the civilian labor force. These are the individuals for whom wage and job information questions are asked.

Table 2(a): CPS OLS Log Hourly Wage Regressions for Males by Race

	Model (1) Including Allocated Earners				Model (2) Excluding Allocated Earners				Model (3) Relocating Missing Wages			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.137 (.005)	.135 (.006)	.146 (.016)	.163 (.016)	.088 (.006)	.083 (.007)	.105 (.020)	.117 (.018)	.086 (.007)	.080 (.008)	.092 (.021)	.109 (.019)
High School no college	.209 (.003)	.209 (.004)	.207 (.009)	.209 (.010)	.184 (.004)	.180 (.005)	.195 (.012)	.197 (.012)	.183 (.005)	.181 (.005)	.191 (.013)	.203 (.012)
College Graduate	.571 (.004)	.570 (.004)	.584 (.012)	.591 (.015)	.540 (.004)	.534 (.005)	.590 (.015)	.573 (.017)	.546 (.005)	.540 (.005)	.585 (.017)	.584 (.016)
HS - Dropout	.209	.209	.207	.209	.184	.180	.195	.197	.183	.181	.191	.203
College - Dropout	.571	.570	.584	.591	.540	.534	.590	.573	.546	.540	.585	.584
College - HS	.362	.360	.377	.382	.357	.354	.395	.375	.362	.359	.394	.381
Adjusted R-squared	.287	.272	.221	.282	.321	.306	.278	.313	.314	.312	.248	.299
Observations	236666	203012	21182	11824	158314	137892	11868	8100	236666	203012	21182	11824
F-test: Prob>F												
GED=Dropout	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
GED=HS	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

**All dummy variables are defined exclusively. Dropouts are the excluded category. Persons enrolled in school at each age are deleted as are those people who: have wages less than \$.50 or more than \$200 an hour (2000 dollars); are self-employed; are not born in the U.S.; are younger than 20 years of age or older than 64; did not complete at least one year of schooling; are Aleut, Eskimo or Native American; or had their completed schooling or GED status imputed by the CPS. Controls for central city status, married with spouse present, year of survey, region of residence, a quadratic in age and race, where appropriate, are included in each regression but not shown. Reported standard errors are corrected for heteroscedasticity and clustering with the Huber-White sandwich estimator except when reimputing wages. Standard errors after reimputation are calculated using the method outlined in Shao and Sitter (1998).

Table 2(b): CPS OLS Log Hourly Wage Regressions for Females by Race

	Model (1) Including Allocated Earners				Model (2) Excluding Allocated Earners				Model (3) Relocating Missing Wages			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.150 (.005)	.140 (.006)	.174 (.013)	.196 (.016)	.110 (.006)	.102 (.007)	.121 (.016)	.157 (.018)	.108 (.007)	.099 (.007)	.117 (.015)	.144 (.017)
High School no college	.237 (.003)	.236 (.004)	.217 (.007)	.257 (.010)	.215 (.004)	.216 (.005)	.191 (.009)	.234 (.012)	.210 (.005)	.205 (.005)	.199 (.009)	.226 (.013)
College Graduate	.673 (.004)	.666 (.004)	.712 (.009)	.708 (.014)	.647 (.004)	.639 (.005)	.698 (.011)	.700 (.016)	.639 (.004)	.629 (.005)	.689 (.010)	.683 (.017)
HS - Dropout	.237	.236	.217	.257	.215	.216	.191	.234	.210	.205	.199	.226
College - Dropout	.673	.666	.712	.708	.647	.639	.698	.700	.639	.629	.689	.683
College - HS	.437	.430	.494	.450	.432	.423	.508	.466	.429	.424	.490	.457
Adjusted R-squared	.277	.263	.308	.313	.307	.291	.355	.351	.305	.287	.342	.347
Observations	223046	185465	26160	10866	154742	130817	15716	7815	223046	185465	26160	10866
F-test: Prob>F												
GED=Dropout	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
GED=HS	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

See table 2(a) for sample definitions and regression controls

Table 3(a): NLSY OLS and Parametric Selection Corrected Hourly Wage Regressions for Males by Race*

	Model (1) No selection or AFQT				Model (2) Including AFQT				Model (3) Controlling for AFQT and Selection**			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.065 (.020)	.068 (.035)	.049 (.030)	.092 (.040)	-.004 (.021)	-.004 (.034)	.003 (.031)	.006 (.042)	-.008 (.021)	.004 (.034)	.000 (.031)	-.020 (.044)
High School no college	.131 (.014)	.165 (.021)	.080 (.024)	.140 (.031)	.044 (.015)	.071 (.023)	.026 (.025)	.032 (.032)	.035 (.015)	.065 (.022)	.029 (.024)	.001 (.034)
College Graduate	.477 (.018)	.472 (.024)	.500 (.037)	.523 (.055)	.274 (.022)	.276 (.031)	.312 (.043)	.253 (.057)	.257 (.022)	.261 (.031)	.307 (.043)	.207 (.057)
AFQT Score113 (.008)	.109 (.012)	.113 (.015)	.125 (.016)	.110 (.008)	.104 (.012)	.111 (.015)	.123 (.018)
HS - Dropout	.131	.165	.080	.140	.044	.071	.026	.032	.035	.065	.029	.001
College - Dropout	.477	.472	.500	.523	.274	.276	.312	.253	.257	.261	.307	.207
College - HS	.346	.307	.420	.383	.230	.205	.286	.221	.221	.197	.279	.206
Adjusted R-squared	.303	.299	.261	.212	.331	.324	.296	.250
Observations	33573	18199	9009	6365	32054	17351	8735	5968	36706	19126	11168	6412
F-test: Prob>F												
GED=Dropout	.001	.055	.107	.022	.842	.899	.925	.882	.701	.909	.993	.650
GED=HS	.000	.004	.257	.219	.010	.018	.390	.516	.021	.053	.284	.605

*All dummy variables are defined exclusively. Dropouts are the excluded category. Persons enrolled in school at each age are deleted as are those people who: have wages less than \$.50 or more than \$200 an hour (2000 dollars), are younger than 20 years of age or older than 39, or are self-employed. Controls for central city status, married with spouse present, year of survey, region of residence, a quadratic in age and race, where appropriate, are included in each regression but not shown. Reported standard errors are corrected for heteroscedasticity and clustering with the Huber-White sandwich estimator.

**We use a parametric model selection correction model due to Heckman (1979). For both males and females the participation equation includes: race dummies, family income in 1979, mother's and father's education, broken home status at 14, urban status at 14, south at 14, number of siblings, local unemployment rate age and age squared. For the female model, spouses income, number of children in the household, and dummies for the presence of a baby or toddler in household are also included.

Table 3(b): NLSY OLS and Parametric Selection Corrected Hourly Wage Regressions for Females by Race

	Model (1) No selection or AFQT				Model (2) Including AFQT				Model (3) Controlling for AFQT and Selection***			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.113 (.021)	.093 (.029)	.122 (.039)	.111 (.041)	.027 (.021)	.012 (.030)	.033 (.035)	.027 (.043)	.017 (.021)	.000 (.031)	.015 (.034)	.032 (.045)
High School no college	.225 (.016)	.199 (.023)	.248 (.032)	.247 (.030)	.130 (.016)	.123 (.024)	.141 (.029)	.123 (.034)	.101 (.016)	.096 (.027)	.107 (.028)	.116 (.036)
College Graduate	.651 (.019)	.607 (.026)	.667 (.037)	.769 (.041)	.429 (.023)	.413 (.032)	.415 (.041)	.507 (.052)	.376 (.023)	.372 (.038)	.345 (.039)	.475 (.054)
AFQT Score131 (.009)	.118 (.118)	.151 (.016)	.146 (.019)	.126 (.009)	.123 (.012)	.135 (.016)	.131 (.021)
HS - Dropout	.225	.199	.248	.247	.130	.123	.141	.123	.101	.096	.107	.116
College - Dropout	.651	.607	.667	.769	.429	.413	.415	.507	.376	.372	.345	.475
College - HS	.426	.408	.419	.522	.299	.290	.274	.384	.276	.276	.238	.359
Adjusted R-squared	.309	.298	.312	.307	.339	.323	.349	.349
Observations	28489	16225	7341	4923	27567	15645	7195	4727	42707	22186	12923	7598
F-test: Prob>F												
GED=Dropout	.000	.002	.002	.007	.187	.689	.268	.534	.428	.999	.673	.394
GED=HS	.000	.000	.000	.001	.000	.000	.001	.013	.000	.001	.002	.049

***See table 3(a) for discussion of the selection model and exclusions.

Table 4(a): CPS-NLSY Comparison- OLS and Selection Corrected Hourly Wage Regressions for White Males

	CPS				CPS NLSY	NLSY		NLSY		NLSY	
	20-29	30-39	40-49	50-59	Cohort 35-46	Excluding AFQT		Including AFQT		AFQT and Selection	
						20-29	30-39	20-29	30-39	20-29	30-39
GED no college	.031 (.011)	.082 (.013)	.104 (.014)	.130 (.017)	.076 (.014)	.052 (.035)	.067 (.043)	-.031 (.035)	-.040 (.042)	-.024 (.035)	-.031 (.042)
High School no college	.112 (.008)	.173 (.008)	.234 (.010)	.220 (.012)	.195 (.009)	.152 (.020)	.206 (.027)	.057 (.022)	.062 (.029)	.047 (.022)	.057 (.029)
College Graduate	.363 (.009)	.544 (.009)	.615 (.010)	.589 (.012)	.598 (.010)	.387 (.024)	.584 (.030)	.198 (.031)	.318 (.038)	.175 (.031)	.305 (.037)
AFQT Score111 (.012)	.153 (.015)	.104 (.012)	.149 (.015)
HS - Dropout	.112	.173	.234	.220	.195	.152	.206	.057	.062	.047	.057
College - Dropout	.363	.544	.615	.589	.598	.387	.584	.198	.318	.175	.305
College - HS	.250	.371	.381	.369	.403	.235	.377	.142	.256	.128	.248
Adjusted R-squared	.246	.283	.267	.228	.294	.214	.278	.244	.317
Observations	29120	40190	38916	24418	34184	10625	8284	10180	7930	11795	8501
F-test: Prob>F											
GED=Dropout	.001	.000	.000	.000	.000	.134	.113	.367	.344	.487	.459
GED=HS	.000	.000	.000	.000	.000	.003	.001	.006	.009	.026	.025

See tables 2(a) and 3(a) for sample definitions and controls.

Parametric selection model estimates are shown. See table 3(a) for details of the estimation procedure.

Table 4(b): CPS-NLSY Comparison- OLS and Selection Corrected Hourly Wage Regressions for White Females

	CPS				CPS NLSY	NLSY		NLSY		NLSY	
	20-29	30-39	40-49	50-59	Cohort 35-46	Excluding AFQT		Including AFQT		AFQT and Selection	
						20-29	30-39	20-29	30-39	20-29	30-39
GED no college	.095 (.013)	.102 (.014)	.119 (.014)	.105 (.018)	.108 (.015)	.084 (.033)	.119 (.044)	.011 (.034)	.011 (.046)	-.004 (.034)	-.001 (.046)
High School no college	.164 (.009)	.229 (.010)	.251 (.011)	.229 (.011)	.243 (.011)	.172 (.026)	.222 (.033)	.092 (.028)	.118 (.036)	.029 (.027)	.108 (.035)
College Graduate	.527 (.010)	.703 (.010)	.683 (.011)	.619 (.012)	.704 (.012)	.483 (.029)	.732 (.035)	.298 (.036)	.510 (.044)	.194 (.036)	.501 (.044)
AFQT Score126 (.013)	.142 (.019)	.120 (.013)	.151 (.018)
HS - Dropout	.164	.229	.251	.229	.243	.172	.222	.092	.118	.029	.108
College - Dropout	.527	.703	.683	.619	.704	.483	.732	.298	.510	.194	.501
College - HS	.363	.474	.432	.391	.461	.310	.510	.207	.393	.165	.393
Adjusted R-squared	.323	.321	.261	.230	.290	.217	.300	.244	.327
Observations	26307	35136	38342	25211	31642	9442	6914	9110	6671	13182	9307
F-test: Prob>F											
GED=Dropout	.000	.000	.000	.000	.000	.010	.007	.738	.811	.899	.991
GED=HS	.000	.000	.000	.000	.000	.001	.006	.002	.005	.213	.004

See tables 2(a) and 3(a) for sample definitions and controls.

Parametric selection model estimates are shown. See table 3(a) for details of the estimation procedure.

Table 5(a): CPS-NLSY Comparison- OLS Log Hourly Wage Regressions for Males by Age

	NLSY Excluding AFQT Score					CPS NLSY Cohort	NLSY Including AFQT Score				
	25	28	30	35	38	38-46	25	28	30	35	38
GED no college	.059 (.038)	.043 (.035)	.015 (.034)	.050 (.041)	.079 (.061)	.085 (.015)	-.034 (.039)	-.037 (.038)	-.065 (.036)	-.043 (.041)	-.084 (.062)
High School no college	.170 (.024)	.141 (.023)	.161 (.024)	.157 (.028)	.194 (.041)	.208 (.010)	.059 (.026)	.044 (.024)	.056 (.025)	.029 (.030)	-.015 (.041)
College Graduate	.373 (.033)	.443 (.029)	.496 (.029)	.650 (.034)	.714 (.048)	.611 (.011)	.134 (.040)	.235 (.035)	.269 (.037)	.382 (.040)	.333 (.053)
AFQT Score134 (.013)	.124 (.013)	.133 (.013)	.149 (.014)	.192 (.019)
HS - Dropout	.170	.141	.161	.157	.194	.208	.059	.044	.056	.029	-.015
College - Dropout	.373	.443	.496	.650	.714	.611	.134	.235	.269	.382	.333
College - HS	.203	.302	.335	.493	.520	.403	.074	.191	.214	.353	.348
Adjusted R-squared	.168	.229	.258	.314	.358	.308	.207	.269	.294	.352	.415
Observations	2247	2367	2400	2287	1088	30549	2165	2254	2298	2196	1039
F-test: Prob>F											
GED=Dropout	.119	.217	.650	.219	.193	.000	.390	.330	.067	.287	.173
GED=HS	.001	.003	.000	.003	.032	.000	.006	.018	.000	.038	.189

See tables 2(a) and 3(a) for sample definitions and controls.

Table 5(b): CPS-NLSY Comparison- OLS Log Hourly Wage Regressions for Females by Age

	NLSY Excluding AFQT Score					CPS NLSY Cohort	NLSY Including AFQT Score				
	25	28	30	35	38	38-46	25	28	30	35	38
GED no college	.096 (.047)	.117 (.048)	.109 (.048)	.114 (.046)	.149 (.062)	.107 (.015)	.014 (.048)	.007 (.049)	-.014 (.049)	.028 (.047)	.022 (.065)
High School no college	.210 (.033)	.234 (.035)	.275 (.034)	.272 (.035)	.315 (.051)	.237 (.011)	.113 (.034)	.123 (.037)	.125 (.036)	.161 (.038)	.160 (.055)
College Graduate	.489 (.036)	.640 (.039)	.728 (.037)	.799 (.038)	.858 (.062)	.700 (.011)	.277 (.042)	.417 (.048)	.432 (.045)	.573 (.048)	.570 (.073)
AFQT Score137 (.014)	.142 (.016)	.180 (.017)	.142 (.018)	.146 (.023)
HS - Dropout	.210	.234	.275	.272	.315	.237	.113	.123	.125	.161	.160
College - Dropout	.489	.640	.728	.799	.858	.700	.277	.417	.432	.573	.570
College - HS	.279	.405	.452	.527	.543	.463	.164	.294	.307	.412	.411
Adjusted R-squared	.176	.261	.318	.319	.311	.297	.213	.295	.361	.350	.342
Observations	1855	1832	1873	1857	913	29452	1803	1782	1812	1800	885
F-test: Prob>F											
GED=Dropout	.041	.014	.025	.013	.017	.000	.765	.879	.783	.554	.733
GED=HS	.003	.002	.000	.000	.001	.000	.011	.003	.001	.001	.004

See tables 2(a) and 3(a) for sample definitions and controls.

Table 6(a): NALS-CPS Comparison- OLS Log Weekly Wage Regressions for Males by Race

	CPS				NALS Excluding Test Score				NALS Including Test Score			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.085 (.007)	.079 (.007)	.107 (.023)	.115 (.021)	.066 (.043)	.079 (.050)	.057 (.108)	.008 (.138)	-.022 (.043)	-.003 (.050)	-.028 (.110)	-.079 (.140)
High School no college	.193 (.004)	.190 (.005)	.200 (.013)	.212 (.013)	.221 (.024)	.241 (.030)	.163 (.046)	.191 (.079)	.126 (.025)	.147 (.031)	.092 (.049)	.092 (.080)
College Graduate	.577 (.005)	.571 (.005)	.616 (.016)	.619 (.019)	.658 (.026)	.664 (.031)	.688 (.063)	.639 (.091)	.441 (.032)	.443 (.038)	.514 (.076)	.407 (.111)
NALS Score148 (.013)	.156 (.016)	.115 (.031)	.129 (.047)
HS - Dropout	.193	.190	.200	.212	.221	.241	.163	.191	.126	.147	.092	.092
College - Dropout	.577	.571	.616	.619	.658	.664	.688	.639	.441	.443	.514	.407
College - HS	.384	.381	.415	.407	.437	.423	.525	.447	.315	.297	.422	.316
Adjusted R-squared	.316	.301	.261	.303	.389	.352	.354	.316	.407	.371	.371	.337
Observations	158603	136796	11704	8026	4077	3236	589	245	4077	3236	589	245
F-test: Prob>F												
GED=Dropout	.000	.000	.000	.000	.122	.115	.596	.955	.603	.948	.799	.574
GED=HS	.000	.000	.000	.000	.000	.000	.314	.182	.000	.001	.264	.215

All dummy variables are defined exclusively. Dropouts are the excluded category. Persons enrolled in school at each age are deleted as are those people who: have weekly wages less than \$100 or more than \$4000 (2000 dollars); are not born in the U.S.; are younger than 20 years of age or older than 64; or are Aleut, Eskimo or Native American. Controls for central city status, married with spouse present, year of survey, region of residence, a quadratic in age and race dummies, where appropriate, are included in each regression but not shown. Robust standard errors shown.

Table 6(b): NALS-CPS Comparison- OLS Log Weekly Wage Regressions for Females by Race

	CPS				NALS Excluding Test Score				NALS Including Test Score			
	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics	All	Whites	Blacks	Hispanics
GED no college	.127 (.007)	.110 (.009)	.142 (.020)	.158 (.021)	.094 (.037)	.088 (.047)	.083 (.085)	.087 (.083)	.023 (.037)	.019 (.047)	.008 (.084)	.054 (.086)
High School no college	.241 (.005)	.234 (.006)	.235 (.011)	.266 (.014)	.229 (.023)	.215 (.031)	.252 (.046)	.233 (.067)	.158 (.024)	.149 (.032)	.179 (.046)	.192 (.070)
College Graduate	.704 (.005)	.686 (.006)	.783 (.013)	.766 (.019)	.737 (.026)	.706 (.033)	.860 (.056)	.731 (.099)	.561 (.032)	.530 (.039)	.678 (.065)	.637 (.119)
NALS Score145 (.015)	.154 (.018)	.135 (.029)	.064 (.051)
HS - Dropout	.241	.234	.235	.266	.229	.215	.252	.233	.158	.149	.179	.192
College - Dropout	.704	.686	.783	.766	.737	.706	.860	.731	.561	.530	.678	.637
College - HS	.463	.453	.548	.499	.508	.492	.607	.497	.403	.382	.499	.445
Adjusted R-squared	.252	.235	.336	.309	.304	.279	.371	.379	.320	.295	.387	.384
Observations	150841	126097	15272	7577	3952	2950	750	238	3952	2950	750	238
F-test: Prob>F												
GED=Dropout	.000	.000	.000	.000	.011	.059	.330	.295	.528	.671	.927	.533
GED=HS	.000	.000	.000	.000	.000	.002	.041	.094	.000	.002	.034	.109

See tables 2(a) and 6(a) for sample definitions and regression controls

Table 7: NALS-CPS Comparison- OLS Log Weekly Wage Regressions by Cohort of Birth

	NALS Males						NALS Females					
	Males 1940-1956			Males 1957-1969			Females 1940-1956			Females 1957-1969		
	CPS	NALS	NALS	CPS	NALS	NALS	CPS	NALS	NALS	CPS	NALS	NALS
GED no college	.126 (.013)	.108 (.069)	.003 (.071)	.084 (.011)	.067 (.062)	.013 (.062)	.136 (.013)	.106 (.057)	.048 (.057)	.128 (.012)	.073 (.061)	-.009 (.062)
High School no college	.226 (.008)	.254 (.040)	.139 (.043)	.209 (.007)	.196 (.038)	.136 (.038)	.253 (.009)	.244 (.039)	.180 (.041)	.259 (.009)	.243 (.040)	.171 (.040)
College Graduate	.616 (.009)	.757 (.042)	.515 (.052)	.636 (.008)	.523 (.041)	.353 (.049)	.691 (.010)	.833 (.041)	.675 (.051)	.751 (.009)	.662 (.044)	.476 (.050)
NALS Score154 (.020)132 (.023)121 (.024)165 (.025)
HS - Dropout	.226	.254	.139	.209	.196	.136	.253	.244	.180	.259	.243	.171
College - Dropout	.616	.757	.515	.636	.523	.353	.691	.833	.675	.751	.662	.476
College - HS	.390	.503	.376	.427	.327	.217	.438	.589	.495	.491	.419	.305
Adjusted R-squared	.248	.360	.380	.310	.319	.336	.206	.304	.314	.248	.281	.302
Observations	51798	1730	1730	61594	1530	1530	53104	1754	1754	55810	1432	1432
F-test: Prob>F												
GED=Dropout	.000	.119	.964	.000	.283	.838	.000	.063	.405	.000	.234	.882
GED=HS	.000	.019	.030	.000	.022	.027	.000	.007	.009	.000	.002	.001

See tables 2(a) and 6(a) for sample definitions and regression controls

Table 8: NALS Foreign Years of Schooling Completed Before Entering U.S

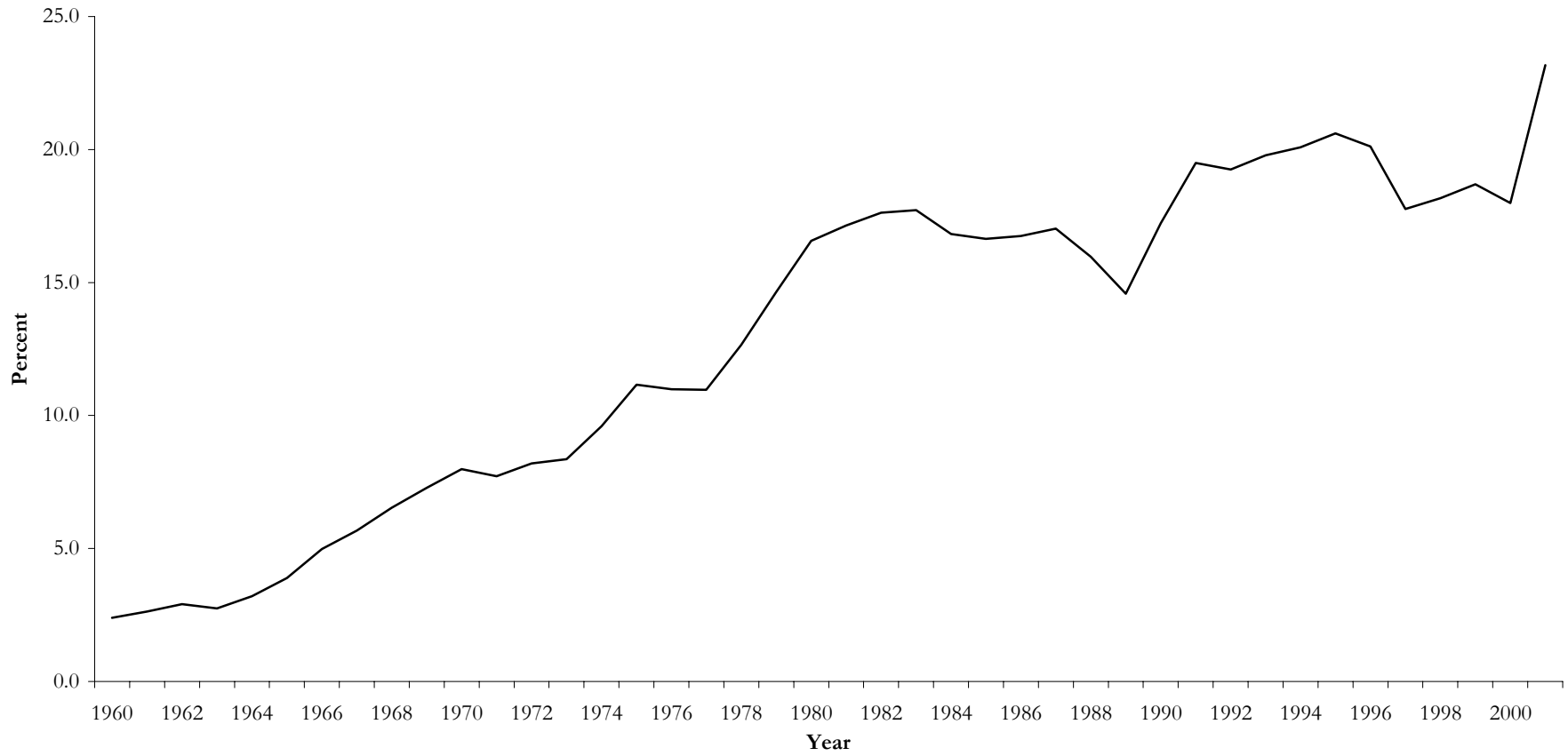
	Males			Females		
	Dropouts	GED	HS	Dropouts	GED	HS
Did not attend school	.104	.154	.145	.084	.158	.126
Primary (Grades K-3)	.151	.039	.039	.158	.000	.049
Elementary (Grades 4-8)	.494	.115	.089	.524	.263	.113
Secondary (Grades 9-12)	.223	.577	.648	.197	.474	.635
Vocational Training	.002	.077	.011	.009	.053	.014
College	.007	.000	.017	.006	.000	.005
Other	.000	.000	.006	.004	.000	.005
NA	.019	.039	.045	.018	.053	.054
Observations	431	26	179	513	38	162

Table 9: NALS-CPS Comparison- OLS Log Weekly Wage Regressions for the Foreign Born

	CPS Including Allocated Values		CPS Excluding Allocated Values		NALS Excluding Test Score		NALS Including Test Score	
	Males	Females	Males	Females	Males	Females	Males	Females
GED no college	.186 (.016)	.157 (.018)	.134 (.019)	.090 (.020)	.109 (.113)	.086 (.112)	.012 (.110)	-.045 (.111)
High School no college	.159 (.006)	.189 (.007)	.100 (.007)	.138 (.008)	.093 (.057)	.095 (.067)	-.024 (.058)	-.049 (.070)
College Graduate	.603 (.009)	.641 (.009)	.574 (.011)	.591 (.012)	.614 (.064)	.659 (.071)	.319 (.076)	.397 (.084)
NALS Score155 (.024)	.153 (.029)
HS - Dropout	.159	.189	.100	.138	.093	.095	-.024	-.049
College - Dropout	.603	.641	.574	.591	.614	.659	.319	.397
College - HS	.445	.452	.474	.453	.521	.564	.343	.446
Adjusted R-squared	.337	.309	.376	.325	.508	.350	.540	.391
Observations	46912	33996	31498	22747	629	429	629	429
F-test: Prob>F								
GED=Dropout	.000	.000	.000	.000	.333	.446	.914	.685
GED=HS	.084	.086	.063	.027	.885	.935	.751	.973

In addition to the regression controls listed in Table 6(a), all regressions include additional controls for cohort of entry, world region of birth, and whether or not the person is a citizen of the U.S.

Figure 1:
GED Credentials Issued as a Percentage of Public and Private High School Graduates, USA
1960-2001



Note:

Source: (1) The Department of Education National Center for Education Statistics Digest of Educational Statistics, 2001, and (2) American Council on Education, General Educational Development Testing Service Statistical Report 1989 and 2002

Figure 2:
CPS Monthly Outgoing Rotation Groups Percentage of Allocated Earners, 1979-2003

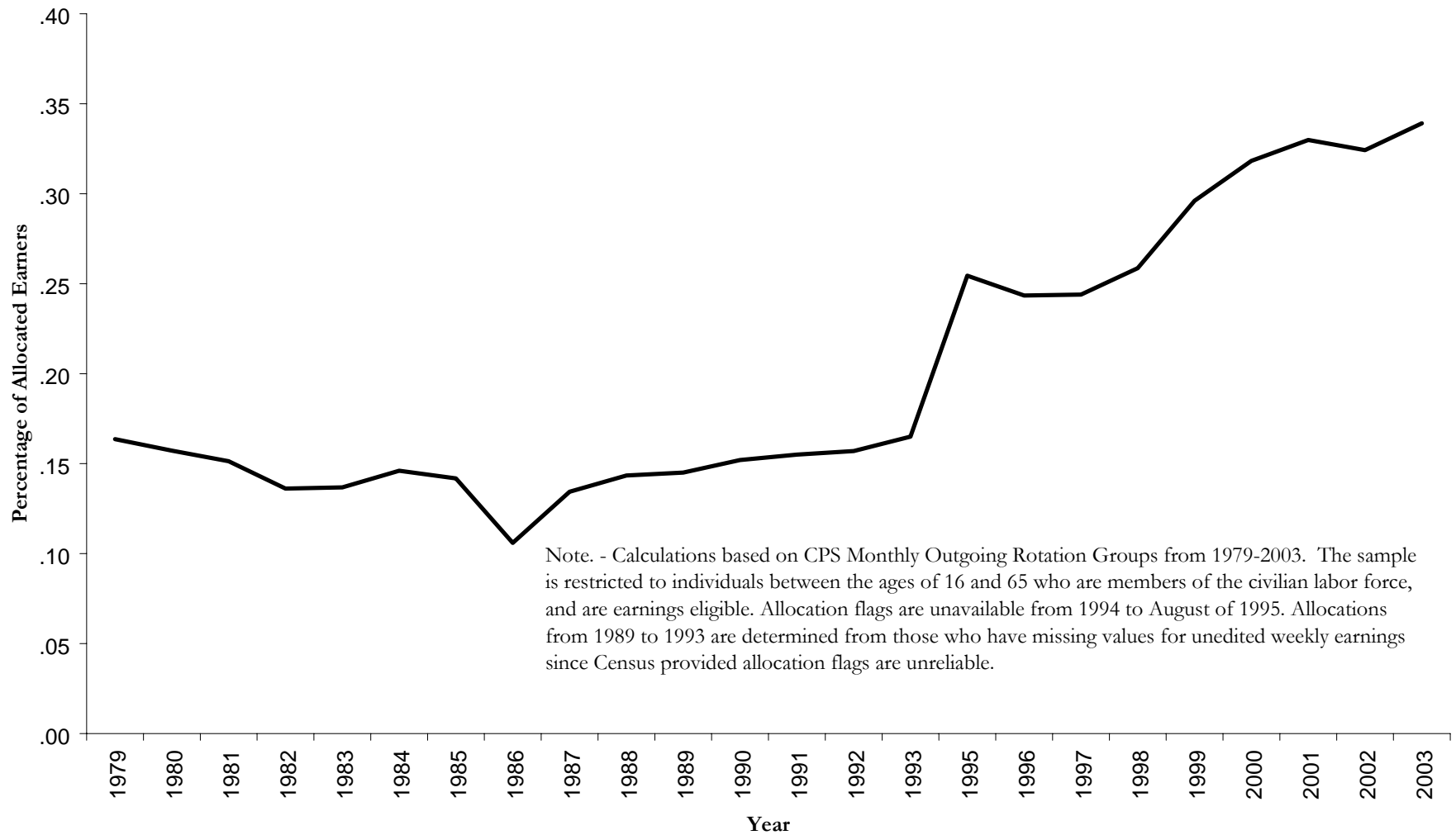
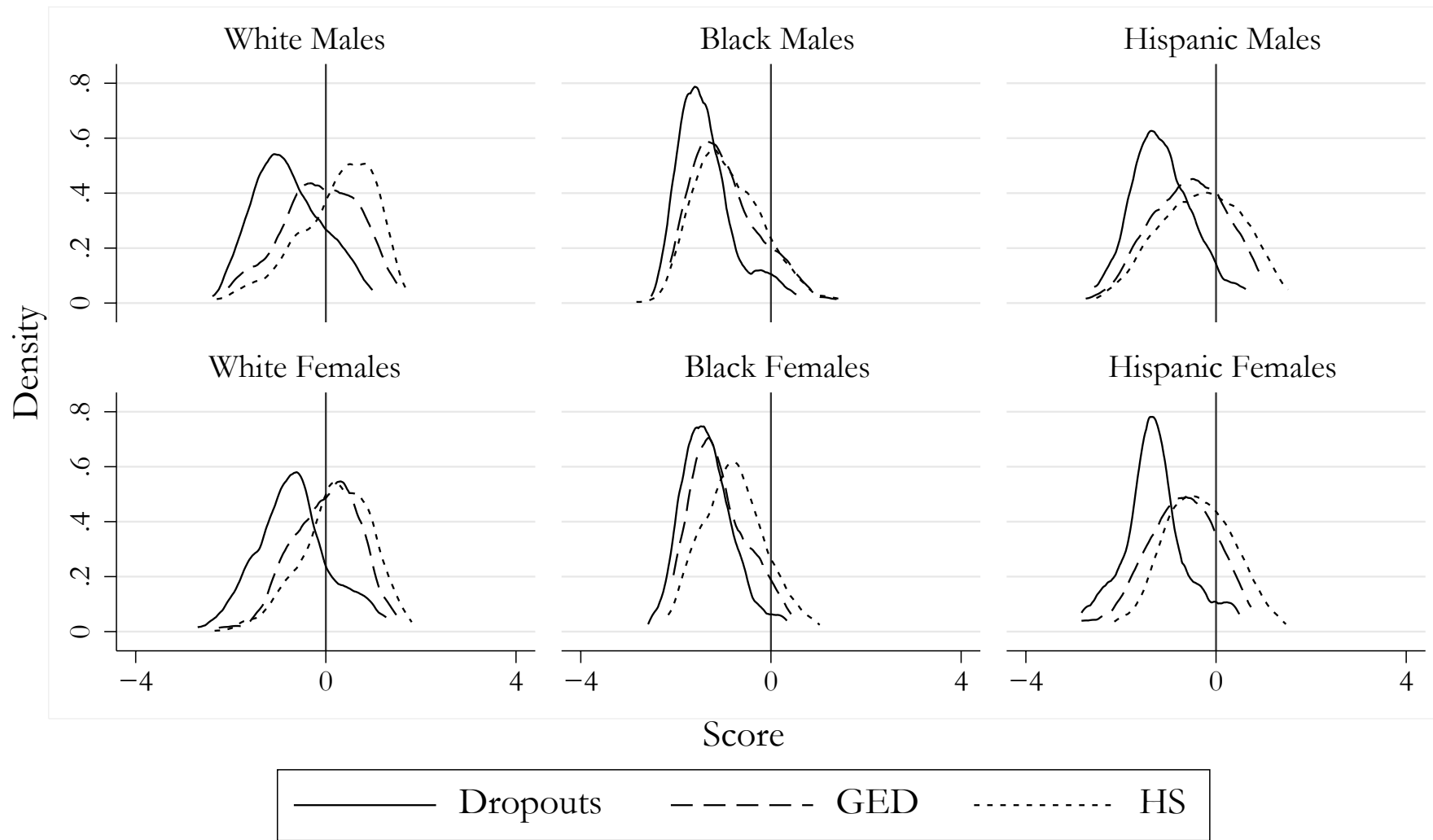
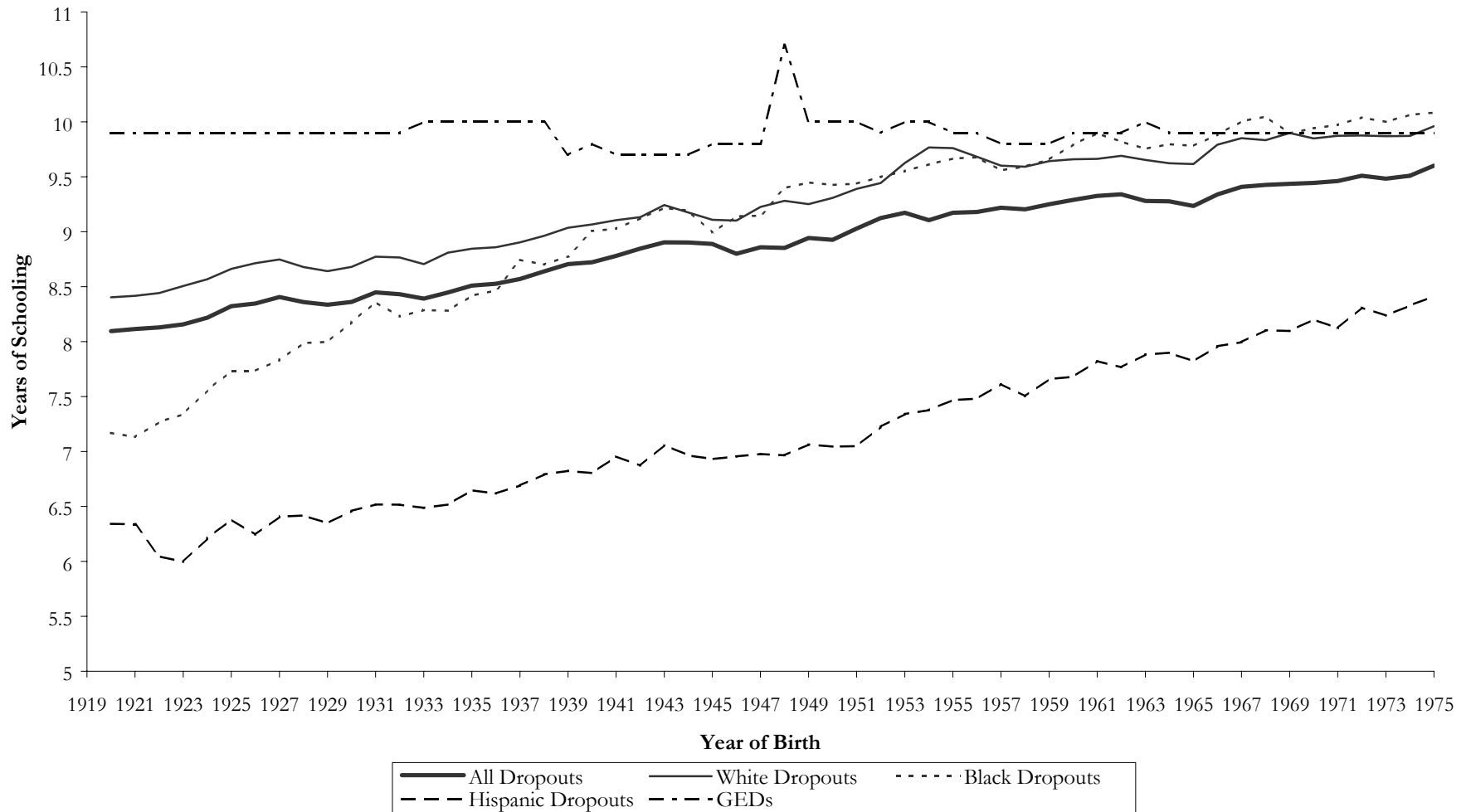


Figure 3
Density of NLSY AFQT Scores by Race and Gender



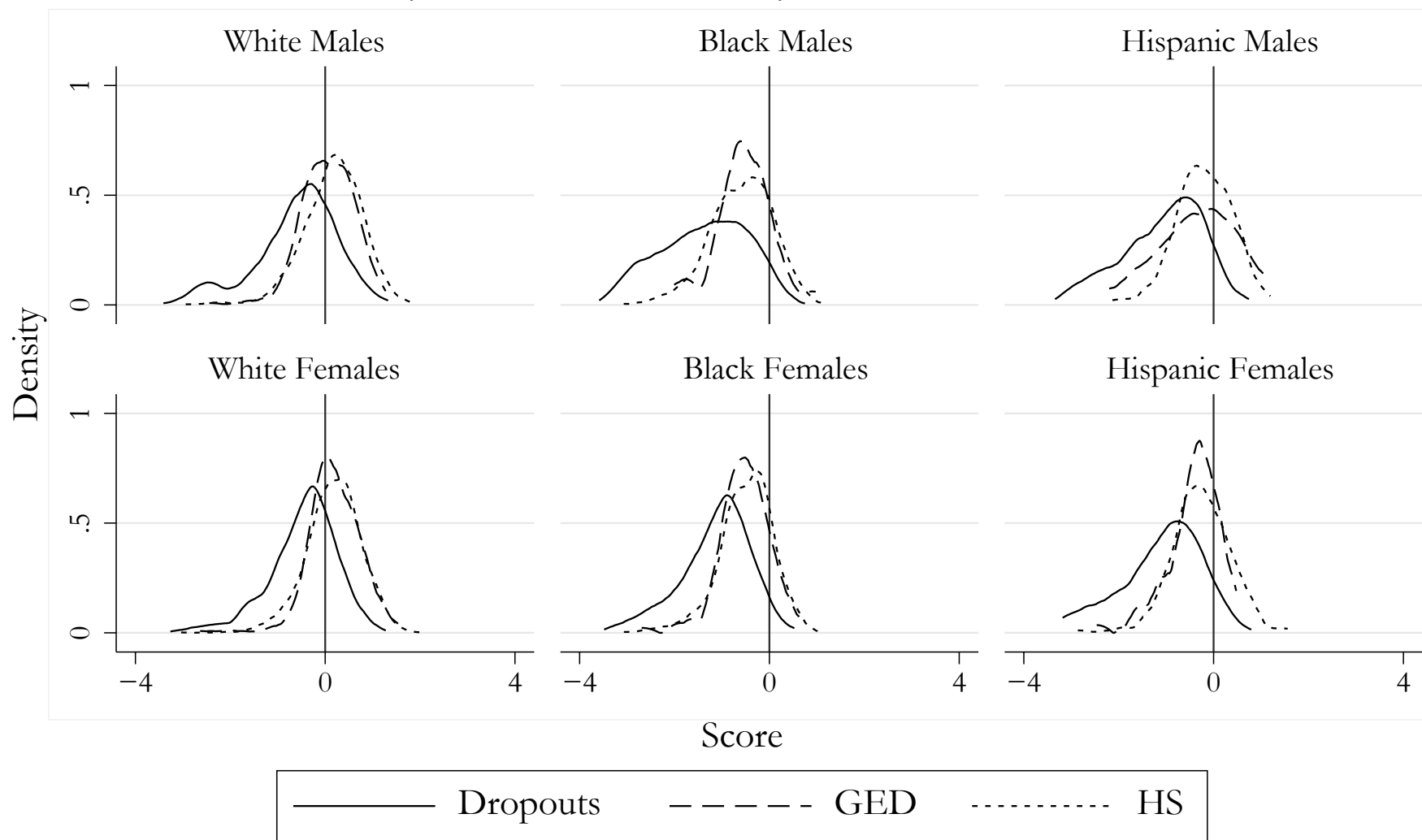
Dropouts, GED holders and high school graduates without post-secondary education. AFQT is a subset of 4 out of 10 ASVAB tests used by the military for enlistment screening and job assignment. It is the summed score from the word knowledge, paragraph comprehension, mathematics knowledge, and arithmetic reasoning ASVAB tests.

**Figure 4:
Average Years of Secondary Schooling for Dropouts and GEDs By Year of Birth**



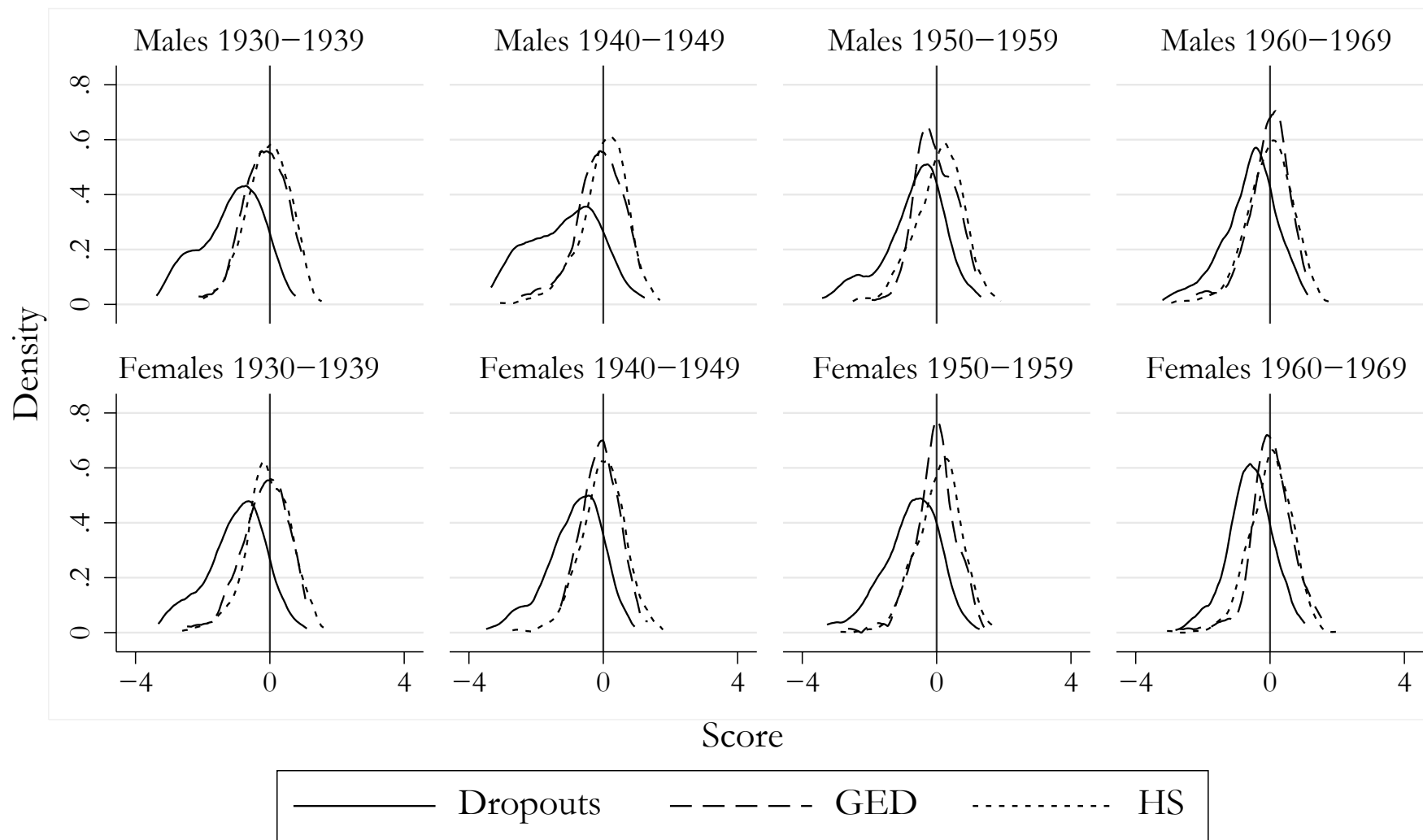
Note: The year-of-birth of GED recipients is imputed as the year that they obtained the GED minus 25 years.
Source: CPS March Data, 1968-2000; GED Statistical Report (various years)

Figure 5
Density of NALS Test Scores by Race for the Native Born



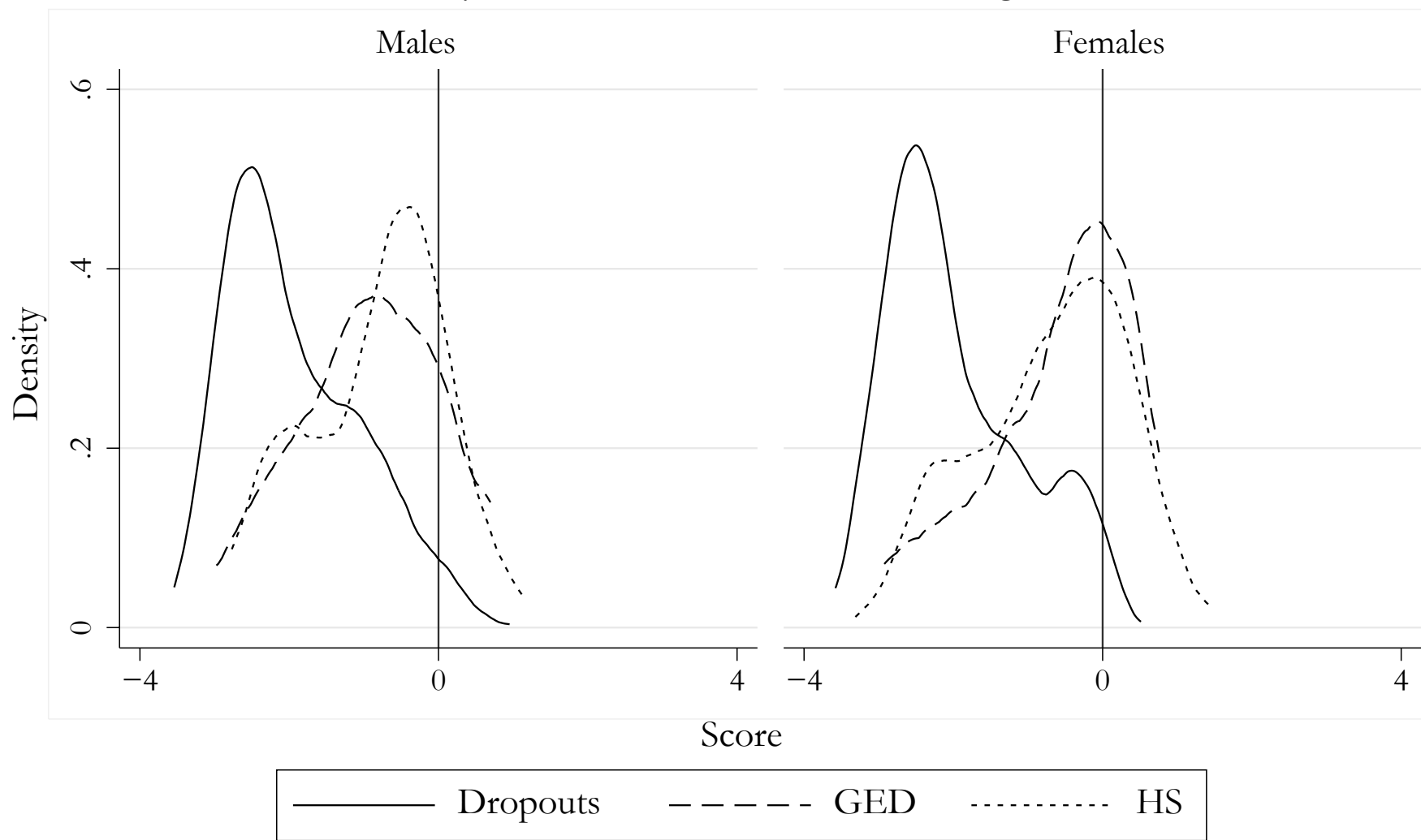
Education categories are dropouts, High School Graduates and GED holders without post-secondary schooling. Test scores are the average over the Prose, Document, and Quantitative examinations administered to all NALS respondents. This is a measure of basic literacy and computational skills. They have been standardized to mean zero and variance one in the population.

Figure 6
Density of NALS Test Scores by Birth Cohort for the Native Born



Education categories are dropouts, High School Graduates and GED holders without post-secondary schooling. Test scores are the average over the Prose, Document, and Quantitative examinations administered to all NALS respondents. This is a measure of basic literacy and computational skills. They have been standardized to mean zero and variance one in the population.

Figure 7
Density of NALS Test Scores for the Foreign Born



Education categories are dropouts, High School Graduates and GED holders without post-secondary schooling. Test scores are the average over the Prose, Document, and Quantitative examinations administered to all NALS respondents. This is a measure of basic literacy and computational skills. They have been standardized to mean zero and variance one in the population.