Patterns of Selection in Labor Market Participation

Mariesa Herrmann*

Cecilia Machado†

May 1, 2012

PRELIMINARY DRAFT

Abstract

Understanding how selection into employment has changed over time is crucial for interpreting the convergence of the measured gender wage gap. However, selection correction methods, which rely on questionable assumptions, have arrived at conflicting conclusions about selection. This paper directly measures patterns of selection of men and women using measures of pre-labor market cognitive ability (high school test scores) from longitudinal educational surveys. Using data from Project Talent, NLS-72, High School and Beyond, and NELS-88, we are able to follow four cohorts from high school until labor market entry and investigate how selection on ability changes over time and over the life cycle. We find that differential selection on the ability of female labor market participants is unlikely to explain the entire convergence of the measured gender wage gap and explore why our results differ from those of previous work.

JEL: J31, J16, J24

Keywords: Selection, Gender Wage Gap

*Columbia University. E-mail: mariesah@gmail.com

†Getulio Vargas Foundation and IZA, EPGE. E-mail: machadoc@gmail.com

1 Introduction

The last several decades have marked a period of rapid economic transformation for women. Over this period, women's participation in the labor force has increased significantly, and their wages have rapidly approached those of men. Figure 1 illustrates these trends using data on white high school graduates between the ages of 25 and 60 from the March Current Population Survey (CPS); participation is measured as the proportion of women working full-time full-year (FTFY) and female relative wages are measured as the difference between the log hourly wages of women and men working FTFY.[1] This figure shows that between 1960 and 2010, female participation increased from around 0.2 to 0.5 and the gender wage gap fell from approximately 0.40 to 0.25 log points. We also see that the trends in relative wages and participation roughly coincide, with steep increases in both series occurring between 1970 and 1990, followed by slower increases thereafter.

Potential reasons for the decline in the gender wage gap are myriad. A leading explanation is that women have made large improvements in human capital (e.g., education, actual experience) and occupational choices, which reflect their changed expectations regarding labor force participation.[2] Other explanations include changes in labor market institutions or demand that favor women relative to men, such as deunionization, increasing returns to soft- rather than hard-skills, and reductions in discrimination [Blau and Kahn, 2006, Weinberg, 2000]. However, since wages are only observed for individuals who select into employment, a lingering question is to what extent the observed changes in the gender wage gap reflect or conceal unobservable changes in the selectivity of the female workforce [Blundell et al., 2007, Blau and Kahn, 2006, Mulligan and Rubinstein, 2008].

The impact of selection bias on women's measured wages has been the subject of much study. Historically, selection bias has been thought to be particularly problematic for women's wages due to women's low rates of labor force participation. Recent increases in fe-

[1] We follow the literature and define FTFY participants as those working more than 35 hours a week and more than 50 weeks in a year. We focus on high school graduates, since our later analysis is restricted to individuals with at least a high school degree. More details on the CPS sample is found in the Appendix.

[2] For example, controlled fertility allowed young women to pursue higher education in degree-fields traditionally dominated by men [Goldin and Katz, 2002, Edlund and Machado, 2011].

male labor force participation have ameliorated worries about selection, but researchers still disagree about the bias's sign. The conventional wisdom is that labor market participants are positively selected, so average observed wages exceed average potential wages [Blundell et al., 2007, Smith and Ward, 1989, Olivetti and Petrongolo, 2008]. However, the opposite is true if participants are negatively selected, for example, due to assortative mating [Neal, 2004].

Of course, there is also no reason to believe that selection has remained constant over time. Somewhat conflicting accounts of changes in selection are provided by Blau and Kahn [2006] and Mulligan and Rubinstein [2008], who base their findings on imputed wages and a parametric selection model, respectively. While the former study finds that selection was more positive in the 1980s and less positive in the 1990s, the latter reports that selection changed from negative to positive between the late 1970s and late 1990s.

Since these accounts rely on different methods and assumptions, it is unclear to what extent such differences may be driving their conclusions. This paper takes a different approach and attempts to provide direct evidence on patterns of selection into the labor market. Specifically, we examine how cognitive ability - a trait that has been shown to directly affect wages [Murnane et al., 1995, Neal and Johnson, 1996, Kuhn and Weinberger, 2005, Fortin, 2007] and which is unobserved in most data sets - affects selection into employment and how this relationship has changed over time. Our approach is also closely related that of Corcoran et al. [2004] who use these data to examine how female selection into the teaching profession changed over time. It complements Mulligan and Rubinstein [2008], who supplement their analysis by examining test scores in a sample of white women from the NLSW, by disentangling age and cohort effects on selection for both men and women.[3]

Our analysis is based on data from four nationally representative surveys of high school

[3]Mulligan and Rubinstein [2008] find that women with above average IQs were more likely to participate in the labor force in the 1990s than in the 1970s, but their results are based on a sample of women who were all between the ages of 14 and 24 in 1968 and thus may confound cohort with age effects. Our data allow us to disentangle these effects, since we compare multiple cohorts of women at approximately the same age. This distinction is potentially important, since many women drop out out of the labor force during their child-bearing years and return when their child-rearing is complete [Smith and Ward, 1989]. Our sample also includes men, so we can assess whether these patterns merely reflect the generally higher ability of labor market survivors, regardless of gender.

students - Project Talent, NLS-72, High School and Beyond, and NELS-88 - which contain test scores that are measured in high school and, thus, are not contaminated by labor market choices or investments. We scale test scores to represent each individual's rank within the distribution for their cohort and gender, holding the distribution of ability fixed over time.

We analyze selection on ability into labor market participation separately for individuals at younger ages (23-25) and at older ages (29-32). Comparing all four cohorts at younger ages, we find that selection for women has became less positive between 1966 and 1999. Comparing the two earlier cohorts - Talent and NLS-72 - at older ages, we find that selection changed from zero to positive between 1972 and 1986, although this positive relationship is no longer statistically significant once we control for experience. For men, at younger ages, we find evidence of positive selection in the 1970s and 1980s but not in the 1960s or 1990s. However, at older ages, we find that men are more positively selected in 1986 than in 1972, and this result is robust to the inclusion of controls for experience. These findings suggest that worries that selection bias - at least on ability - contaminates measured wages may be exaggerated for women and understated for men.

This paper proceeds as follows: In Section 2, we describe the literature on the gender wage gap and selection, Section 3 describes the data, Section 4 presents our methodology and results, Section 5 investigates the possible explanations for our results, and Section 6 concludes.

2 Background

Economists have long been concerned about selection bias in observed wage gaps. Correction methods for selection date back to Gronau [1974] and Heckman [1974]. Recently, several well-regarded studies have used different methods to correct for selection bias in the estimation of gender wage gaps, only to reach starkly different conclusions about its evolution.

Using a parametric selection model (among other methods), Mulligan and Rubinstein [2008] estimate that the gender wage gap in the U.S. remained stable at around -0.338 log points between the late 1970s and late 1990s. They find that female selection into employment switched from negative to positive over this time period, which they interpret as evidence that growing wage inequality within gender differentially induced more able women to work. Based on this finding, they argue that the narrowing of the gender wage gap can be entirely attributed to unobservable changes in labor force composition. However, it is worth noting that their parametric model relies upon the validity of somewhat questionable instruments for female labor force participation - interactions between the number of children under the age of 6 and marital status.

In contrast, using imputation methods, Blau and Kahn [2006] estimate that the median gender wage gap closed 0.15 log points per year in the 1980s and 0.10 log points per year in the 1990s. Their selection-corrected estimate is smaller than their uncorrected estimate of 0.17 log points for the 1980s but larger than their uncorrected estimate of 0.09 log points for the 1990s, suggesting that selection became more positive in the 80's and less positive in the 1990's. Of course, these estimates depend on the validity of their imputation procedure, which uses part-time or recent wage observations when they are available, and, when they are not available, relies on assumptions about women's potential wages should fall above- or below- the median.

Taking a different approach, Blundell et al. [2007] assume that selection is positive and correct for it using bounding procedures. They find that the pay gap in the U.K. closed at least 0.23 log points between 1978 and 1998; Machado [2009] replicates their procedures for the U.S. and obtains similar results. However, the preceding discussion suggests that positive selection is not necessarily a trivial assumption. Neal [2004] also argues that white women may be negatively selected into employment due to assortative matching in the marriage market. Yet, the assumption of positive selection is crucial for the results in Blundell et al. [2007] as the bounds become uninformative when this assumption is relaxed.

From these studies, it is evident that there is large disagreement in the literature regarding the nature of selection, despite its importance for drawing conclusions about the evolution of the gender wage gap.

3 Data

We use data from four nationally representative longitudinal surveys of high school students: Project Talent (Talent) for the class of 1961, the National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HS&B) for the class of 1980, and National Education Longitudinal Study of 1988 (NELS-88) for the class of 1992.[4] Since the latter three studies belong to the National Education Longitudinal Studies (NELS) program of the National Center for Education Statistics (NCES), for the sake of simplicity, we refer to all four surveys as the NELS.[5]

Each of these studies administered a background questionnaire and a battery of aptitude tests while students were in high school. The studies also conducted a number of follow-up surveys to obtain information about students' educational attainment, labor market experiences, and family formation after high school, though the length of time between the follow-up surveys and high school graduation varied across studies.[6] We focus only on the follow-up surveys taken when respondents were at least 23 years old, and we use retrospective information when available to aid comparability between surveys. For Talent, we have follow-up surveys from 1967, when they were 23, and from 1972, when they were 29. We also have two follow-up surveys for NLS-72, taken in 1979 when they were 25 and in 1986 when they were age 32. HS&B and NELS-88 were only followed for a short period of time, so the last HS&B follow-up was conducted in 1986 when students were 24. The final NEL-88 follow-up was conducted in 2000, but we use outcomes from 1999, when respondents were 25 years old. Figure 2 compares the availability of data across NELS.

These data span a crucial time period for understanding changes in female labor force participation and gender wage gaps. It is well-known that many of the changes observed in

[4]We focus on the Talent class of 1961, who were juniors in 1960 when the exams were administered. Since the other cohorts were tested in their senior year, we believe the juniors to be more comparable than the sophomores or freshmen. Unfortunately, the 5-year follow up for the Talent seniors (the class of 1960) lacks data on full-year employment. However, results using the 11-year follow up for the Talent seniors are qualitatively similar.

[5]Talent is a predecessor of this program and was administered by the American Institutes for Research, in conjunction with other organizations, such as the University of Pittsburgh.

[6]More information about these surveys can be found in the Appendix.

the cross-section (e.g., Figure 1) reflect differences in the characteristics and behaviors of the cohorts entering versus those exiting the labor force [Weinberger and Kuhn, 2010]. Figure 3 uses CPS data to graph these trends over the life-cycle for the cohorts in our sample. From Panels A and B, we see that within cohorts, the relative wages and participation of women decline as they enter childbearing age. After women complete childbearing, relative female participation rises, but relative female wages rebound slightly at best. In contrast, once individuals complete post-secondary education, the participation rate for men remains quite steady (Panel C). The figure also shows that female wage and participation trends begin to diverge between cohorts as workers reach their mid-twenties, suggesting that even though our analysis is limited to young workers, its occurs at critical age. Moreover, there is also a particularly large divergence in trends between the cohort born in the early 1940s and the one born in the 1950s, the two cohorts for whom labor force information is available at later ages in NELS, around age 30.

Our main sample consists of high school graduates who responded to the selected follow-up surveys and have test scores available.[7] To abstract away from changes in race inequality, we focus on non-Hispanic whites, and to avoid life-cycle effects on labor market outcomes, we compare individuals in the follow-up surveys when they are approximately the same age. We compare all four cohorts close to the age of labor market entry, when respondents are between 23 and 25 years old, but due to the limited follow up periods of our surveys, we are only able to compare the two earlier cohorts - Talent and NLS-72 - at older ages, around 29 to 32. The multiple observations from Talent and NLS-72 also allow us to examine how selection changes within cohorts over time.

As our measure of cognitive ability, we use students' score on the math exam, which has been found by other studies to have a stronger effect on wages than verbal test scores [Murnane et al., 1995, Kuhn and Weinberger, 2005, Fortin, 2007]. Since exams between

[7] We focus on high school graduates to avoid bias caused by differences in students' ages across the baseline surveys. While most of the baseline surveys were taken during the senior year, some were taken earlier, in 9th through 11th grades. Direct comparisons of all students from the baseline surveys would fail to account for ability differences between students who drop out of high school in 9th grade and students who drop out of high school in their senior year.

the surveys differ in content and scaling, we follow other studies and construct comparable measures by calculating each student's centile rank in the distribution of white high school graduates of the same gender [Kuhn and Weinberger, 2005, Corcoran et al., 2004]. We interpret this rank as a measure of students' cognitive ability, assuming that the distribution of ability in this population has remained stable over time.[8] This allows us to examine changes in selection for individuals at the same rank in the ability distribution, although it is worth noting that relative to men of the same rank, women have made significant gains in skills over this time period [Goldin et al., 2006, Fortin et al., 2012].

Our measure of labor force participation is whether individuals were employed full-time, full-year (FTFY), which we define as working at least 35 hours per week and at least 50 weeks in the previous year.[9] We generate this measure from information on the number of hours and weeks worked; we also calculate real hourly wages (in 2000 dollars).

Table 1 displays average FTFY participation and log hourly wages for men and women in each longitudinal survey. For the purposes of comparison, the table also includes averages from the CPS for high school graduates who are the same age as our NELS cohorts in each survey year. There are some differences between the estimates in the NELS and CPS, but these tend to be small, and overall, surveys generate similar patterns for participation and wages. The top panel documents trends in labor force participation; comparisons between Talent and NELS88 show that the participation rates of women in their early to mid-twenties rose dramatically, from approximately 23% to 54% between 1966 and 1999. Participation rates of women in their late twenties to early thirties also grew, from about 23% in the early 1970s to 37% in the late 1980s. On the other hand, the NELS data show that male participation rates are relatively high but have slightly declined over time. The bottom panel documents the trend of the decline in the gender wage gap. For workers at approximately age 30, the gender wage gap fell from approximately 0.34 to 0.25 log points between the early 1970s and late 1980s. Differences in log wages for younger workers are smaller and seem to

[8] Our results are robust to using alternative measures of test scores (e.g., verbal scores) or transformations (e.g., z-scores with mean 0 and standard deviation 1, indicators for ability quintiles, scaling test scores for men and women within the same distribution).

[9] Results that use full-time employment as a measure of participation are qualitatively similar.

primarily reflect differences in the age profile of gender wage gaps.

In addition to information on labor market outcomes, the follow-up surveys also provide information on marital status and the number and age of children, which we use to examine heterogeneity in selection across subgroups of the population. Table 2 presents a full set of summary statistics for the NELS data. These data illustrate several well-known demographic trends: men and women are attaining higher levels of education, they are marrying less and at older ages, and conditional on age, they are less likely to have young children.

4 Patterns of Labor Market Selection

To examine how cognitive ability affects the likelihood of labor market participation, we estimate the following linear probability model - separately by cohort, age, and gender:

$$\text{FTFY}_{igca} = \alpha_{gca} + \beta_{gca}\text{score}_{igca} + \gamma_{gca}X_{igca} + \varepsilon_{igca} \tag{1}$$

where $FTFY_{igca}$ is an indicator that equals 1 if individual i of gender g from cohort c was employed FTFY at age a and equals 0 otherwise. score_{igca} is a measure of individual's ability, X_{igca} is a vector of controls (education, parents' education, and high school region), and ε_{igca} is the error term.[10]

The coefficient of interest is β_{gca}, which measures the relationship between ability and labor force participation for a given gender, cohort, and age. Since our measure of ability ranges from 0 to 1, for individuals who are 1 percent higher in the ability distribution, β_{gca} measures the difference in percentage points in participation. The sign of β_{gca} thus provides evidence on whether average selection on ability into employment is positive, negative, or zero for a cohort, while the changes across cohorts in β_{gca} provide evidence that this relationship is changing over time.

We begin by investigating patterns of selection for all four cohorts in the years 1966, 1979,

[10]Education controls include indicators for some college and college or more. Parents' education controls are indicators for less than high school, high school, some college, college or more, or missing. High school region includes controls for south, west, midwest, and northeast.

1986, and 1999, when respondents were approximately 23 to 25 years old. Table 3, Panel A displays the results for women. For the three earliest cohorts, who enter the labor market between the late 1960s and late 1980s, ability is positively associated with labor market participation, but there is no relationship between ability and labor market participation in the most recent cohort. Although we cannot reject that average selection has remained constant over time, moving from left to right, the estimates are suggestive that at young ages, average selection became more positive for women in late 1980s and less positive in the late 1990s. The increase in participation from 23% to 44% between Talent and HS&B appears to increase positive selection; the estimates suggest that a 10% increase in ability corresponds to a 1.4% increase in the probability of participation in 1966 versus a 2.0% increase in this probability in 1986. On the other hand, when participation increases to 54% in NELS-88, and the coefficient on ability becomes smaller in magnitude and is no longer statistically significant. In addition, the lack of selection in NELS-88 seems surprising, given that participation is still relatively low.

Results for men are displayed in Table 3, Panel B. Similar to our results for women, we find evidence of positive selection for men in the NLS-72 and HS&B cohorts. This may be due to low-ability men dropping out of the labor market in response to the declining labor market opportunities for low-skill men in the late 1970s and 1980s [Juhn, 1992]. These findings also suggest that even though male labor force participation is relatively high, male selection is not necessarily negligible.

Since Table 3 compares cohorts at slightly different ages, one concern is that these results may reflect changes in selection over the life-cycle rather than changes in selection across cohorts. Table 4 addresses this concern by presenting results for the Talent and NLS-72 cohorts for a much wider range of ages, spanning ages 23 through 32. This table provides some evidence on how selection on ability changes over the life-cycle, as well as between these cohorts at older ages. From Panels A and B, we see that for the Talent cohort, ability is not related to participation at age 29 for either men or women, even though female participation rates are relatively low. This suggests that for this cohort, the measured gender wage gap of 0.319 log points (Table 1) may not be biased by selection - at least on ability.

On the other hand, in the NLS-72 cohort, the positive relationship between ability and participation strengthens as participants age. While this relationship is increasing for men, it remains stable for women after they reach their thirties. Comparing the Talent and NLS-72 cohorts around age 30 (Columns 2, 6), we find that positive selection increased between the early 1970s and mid-1980s for both men (p-value = 0.50) and women (p-value = 0.52).[11] These results also confirm the results from the previous table that selection on ability appears to be more positive in the 1980s for both men and women.

Our results point to the following findings: First, across cohorts, women at young ages (around 23-25) become less positively selected between 1960 and 1990. Because FTFY participation increases across cohorts, this pattern suggests that lower ability women were pulled into the labor force in later cohorts. This evidence accords with what might be expected under a fixed positive selection rule, in which the earliest participants have the highest ability and adding more workers reduces the average ability of participants. However, the comparison between the Talent and NLS-72 cohorts reveals that women at older ages (around 30) become increasingly positively selected into work between 1972 and 1986. Participation also increases between these two cohorts, but average selection becomes even stronger (and positive). An interesting feature of these findings is that the relationship between average selection and female participation is not monotonic, as might be expected from either a uniformly positive or negative selection rule.

Most interestingly, the within cohort comparison reveals a dramatic change in women's life-cycle profiles of selection into employment. In Talent, more able women who were working in their early twenties are no longer participating by their thirties: since participation remains stable throughout the life-cycle, these women must have been replaced by less able women in the labor force. In NLS-72, however, participation is also stable, but selection increases over the life-cycle. This suggests that not only do the more able survive, but they replace some of the low ability women who were working at younger ages.

[11] This is not a feature of selective attrition between follow-up surveys; estimates from regressions that control for individual fixed effects and common weights in each survey are even larger in magnitude for the NLS-72 cohort.

What explains these patterns? Women's increased investments and higher attachment to labor force emerge as natural candidates. Increasing returns to skill and wage inequality could also be behind these changing patterns of selection. In addition, marriage and fertility, which have been shown to directly affect women's labor supply, have been undergoing substantial changes across these cohorts. We turn then to investigate those explanations.

5 Investments and Marriage

5.1 Investments

In this paper, we have focused on test scores as a measure of the average selection of labor market participants. This focus is driven by the demonstrated importance of test scores for labor market productivity [Murnane et al., 1995, Neal and Johnson, 1996] and the fact that test scores are uncontaminated by worker's choices in the labor market. Of course, there are many other dimensions of ability [Fortin, 2007, Kuhn and Weinberger, 2005] on which workers could be selected, and while in the labor market, workers make additional decisions about investments and human capital accumulation that affect participation and wages.

In this section, we provide results which control for one measure of workers' labor market investments: past full-time labor market experience. This allows us to investigate the sensitivity of our results to this control and further illustrates how participants and non-participants may differ on dimensions other than ability. Labor market experience is also an important determinant of labor market outcomes, and much of the convergence in the gender wage gap has been attributed to improvements in women's actual labor market experience [Blau and Kahn, 1997, O'Neill and Polachek, 1993, O'Neill, 2003].

Results that control for years of full-time work experience are presented in Table 5. Unfortunately, results for ages 23-25 are limited to the NLS-72 and HS&B cohorts (columns 1 and 2) because experience is not consistently reported for either the Talent and NELS-88 cohorts.[12] As expected, previous full-time experience has a positive relationship with current

[12] The question in Talent 5-year follow up (age 23) is whether the individual has worked continuously since

FTFY participation, suggesting that labor market participants tend to have more experience than non-participants, and though one should be cautious about comparing the coefficients on experience for individuals of different ages (who thus have different number of years of potential experience), for women, it does not appear that the gap in experience is increasing between participants and non-participants over time.

While the inclusion of a control for experience does not affect our findings of positive selection on ability at young ages in NLS-72 or HS&B for either women or men (Columns 1 and 2), the control does reduce the estimated amount of positive selection on ability at age 32 for NLS-72, suggesting that more able individuals accumulate more labor market experience. For women, the coefficient on ability is no longer statistically significant once we control for experience, suggesting that selection on ability into employment is negligible once experience is taken into account. Men, however, remain positively selected on ability and have more full-time experience relative to non-participants.

5.2 Marriage

We have already seen that female labor force participation rose dramatically across the cohorts we study. The majority of this rise comes from the increased participation of married women. For women in their twenties, female labor force participation increased from 18% to 48% for married women and from 37% to 61% for unmarried women between 1966 and 1999. For women in their thirties, participation increased from 18% to 31% for married women between 1972 and 1986, while it dropped slightly - from 58% to 52% - for unmarried women. Other observers have noted that the increase in married women's labor force participation was concentrated among the wives of middle and high-income men [Juhn and Murphy, 1997, Mulligan and Rubinstein, 2008]. If spouses are assortatively matched, this could suggest that married women have become more positively selected into the labor force over time. On the other hand, if assortative matching is increasing over time [Pencavel, 1998, Qian, 1998], changes in the labor force participation of married women could simply reflect changes in

leaving school, and the NELS-88 surveys only ask about experience prior to 1994 and whether the individual worked for at least 6 months in 1997 and 1998.

the characteristics of married individuals.

In this section, we examine whether there is heterogeneity in selection by marital status, by modifying equation 1 to include an indicator for whether the individual was married and an interaction between this indicator and the test score measure. Table 6, Panel A displays the results. The first four columns correspond to the results at ages 23-25. Across cohorts, ability has no impact on the FTFY employment of unmarried women, with the exception of HS&B. Marriage decreases the probability of employment in all four cohorts, but it only differentially affects selection into employment for the NLS-72 and the NELS-88 cohorts. In both of these cohorts, married women are positively selected into FTFY employment, though average selection for women is positive in NLS-72 and zero in NELS-88 due to differing rates of marriage. Since marriage rates, particularly at these ages, have fallen considerably across cohorts, we also investigate whether selection into marriage has changed. Table 6, Panel B reports the results of a regression of an indicator for being married on our measure of ability. We do not find evidence of differential selection (based on test scores) into marriage across cohorts at these ages, and interpret the increasing positive selection of married women as a change at young ages in married women's behavior towards work.

In columns (5) and (6) we examine selection at older ages (29-32) in the Talent and NLS-72 cohorts. For older unmarried women, there is a large rise in labor market selectivity; the coefficient on ability for unmarried women increases from a statistically insignificant 0.0580 in Talent to a statistically significant 0.3420 in NLS-72. The effect of marriage on participation also changes between these cohorts, with marriage having a negative effect on participation for Talent but no impact on participation for NLS-72. Similar to the results for unmarried women, the selection of married women into employment has also become more positive over time (0.0222 in Talent to 0.0801 in NLS-72), though the difference is smaller for married than for unmarried women. In fact, in NLS-72, married women are negatively selected relative to unmarried women, which is consistent with more able married women having high reservation wages due to assortative matching [Neal, 2004, Blundell et al., 2007]. It is also worth noting that we do find some evidence of positive selection into marriage for the NLS-72 cohort at age 32 (Panel B, column 6), suggesting that part of the positive selection among married women may be reflecting differences in characteristics of this group. Yet, the even stronger increase in positive selection among unmarried women points to changes in female work behavior overall at older ages.

Comparing changes in selection over the life-cycle within cohorts, we find no evidence that selectivity changes for either married or unmarried women in the Talent cohort. However, in NLS-72, both married and unmarried women are becoming increasingly positively selected over time, with larger changes for women who are unmarried. We speculate that the changes in selectivity for married women reflect both the increase in assortative matching at older ages and increased disparities between the wages of their husbands at older ages.

6 Discussion

Using test scores as a measure of ability, this paper finds evidence that at young ages, women became less positively selected into work between the late 1960s and late 1990s. At older ages (29-32), there is some evidence that positive selection increased for women between 1972 and 1986, though these results are not statistically significant when we control for experience. The fact that there is no relationship between ability and FTFY participation even at levels of female FTFY participation which are substantially below one suggests that working women and non-participants are not drawn exclusively from either the top or the bottom of the ability distribution [Machado, 2009, Neal, 2004].

Taken together, these results suggest that measured wages may be a decent proxy for the average potential wages of women. This is similar to the findings of Blau and Kahn [2006], who find using imputation techniques that selection accounts for only a modest part of the observed convergence in the wage gap; the uncorrected convergence between 1979 and 1999 is -0.261 log points and the corrected convergence is -0.253 log points.

Our results, however, contrast those of Mulligan and Rubinstein [2008] who use a parametric correction procedure that uses the presence of young children as an instrument for participation. They estimate that selection has switched from negative to positive and that changes in selection explain *all* of the observed convergence of the gender wage gap. There are several reasons that our results could differ: First, our analysis focuses on changes across cohorts rather than in pooled cross-sections. While we find no evidence of negative selection in the cohorts we study, it is possible that selection could have been negative for other cohorts that are included in the cross-section. Second, there is much disagreement [Mroz, 1987, Angrist and Evans, 1998] over whether the presence of young children is a valid instrument for women's decisions to work. If the instrument is not valid, the bias of the estimates depends on the correlation between market wages and fertility. To sign this bias, we use the NELS cohorts to examine the relationship between ability and having children under 6, as well as FTFY participation.

Table 7, Panel A shows the presence of a child under 6 reduces female FTFY participation in all cohorts and at all ages, although the relationship is weaker for more recent cohorts. For instance, 30-year old women with young children are 35.1 percentage points less likely to work in 1972 but only 21.0 percentage points less likely to work in 1986. Panel B examines whether selection into having young children has changed over time. We find that ability is increasingly negatively correlated with the presence of young children at young ages (columns 2 and 3) but becomes positively correlated with the presence of young children at older ages (column 6), probably because high ability women delay child-bearing. Since ability is positively correlated with wages, these results suggest that at older ages, using the presence of young children as an instrument may result in estimates of selection that are too positive.

Finally, in contrast to the conventional wisdom, our analysis suggests that male selection is important for measuring gender wage gaps. We find strong evidence of positive selection of men into work that persist even after work history controls (experience) are added.

Of course, our results are limited to one characteristic that is valued in the labor market - ability - and the consideration of other characteristics on which labor market participants could be selected remains an important topic for future research.

Figure 1: Female Relative Wages and Participation

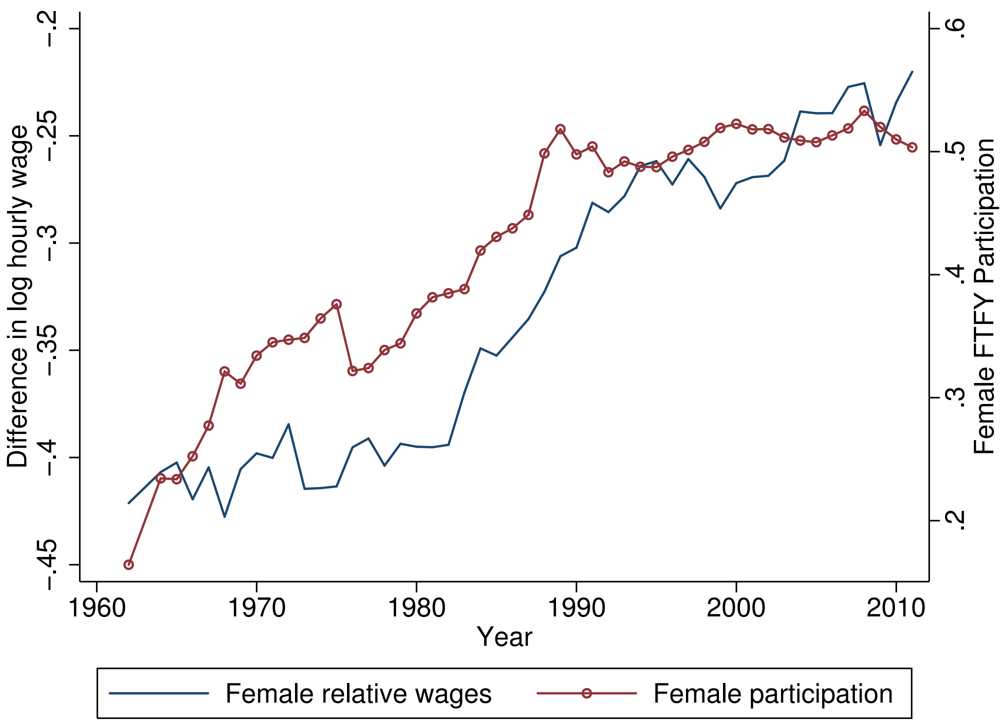
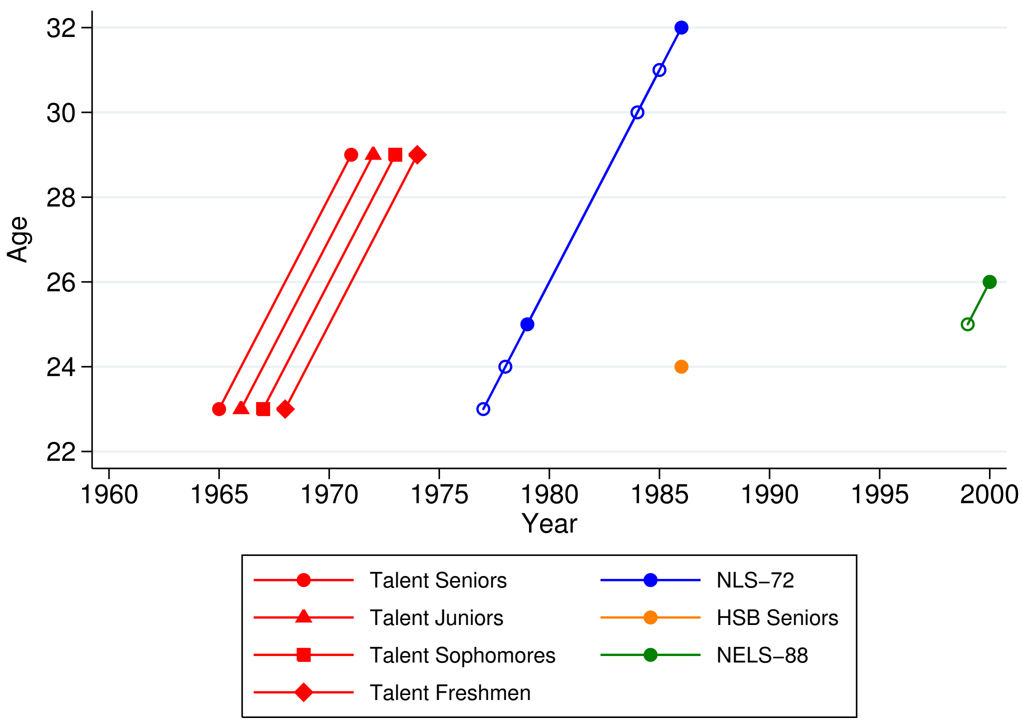


Figure 2: Comparison of Respondents' Ages in Follow-up Survey Data



Souce: Authors' calculations

Note: Closed circles represent follow-up years; open circles represent retrospective reporting.

Figure 3: Female Relative Wages and Participation by Cohort

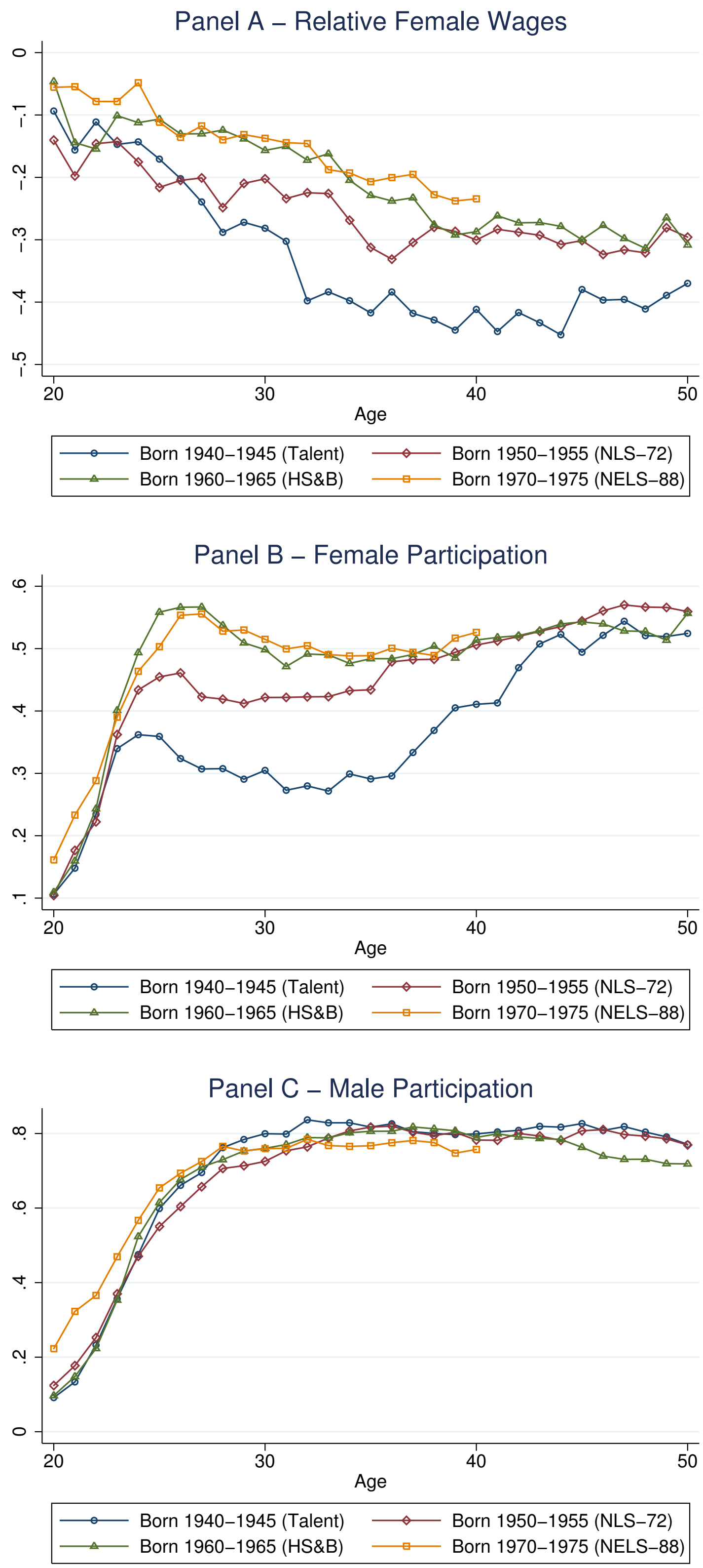


Table 1: CPS versus NELS comparison

Panel A. Percent Full-Time Full-Year

	NELS			CPS		
Dataset, Year (Age)	Male	Female	Difference	Male	Female	Difference
Talent, 1966 (23)	0.451	0.228	0.223	0.519	0.343	0.176
Talent, 1972 (29)	0.833	0.246	0.587	0.786	0.298	0.488
NLS 72, 1979 (25)	0.636	0.420	0.216	0.608	0.445	0.163
NLS 72, 1986 (32)	0.675	0.370	0.305	0.796	0.412	0.384
HSB Senior, 1986 (24)	0.552	0.442	0.109	0.552	0.501	0.051
NELS 88, 1999 (25)	0.682	0.544	0.137	0.712	0.557	0.155

Panel B. Log Hourly Wages of Full-Time Full-Year Participants

	NELS			CPS		
Dataset,Year (Age)	Male	Female	Difference	Male	Female	Difference
Talent, 1966 (23)	2.424	2.360	0.0634	2.579	2.422	0.158
Talent, 1972 (29)	2.991	2.633	0.357	3.000	2.681	0.319
NLS 72, 1979 (25)	2.692	2.482	0.210	2.725	2.504	0.221
NLS 72, 1986 (32)	2.898	2.643	0.255	2.964	2.739	0.224
HSB Senior, 1986 (24)	2.489	2.361	0.127	2.619	2.473	0.145
NELS 88, 1999 (25)	2.612	2.470	0.141	2.653	2.502	0.151

Table 2: Summary Statistics of NELS Datasets

	Talent11 1966	Talent11 1972	NLS72 1979	NLS72 1986	HSB 1986	NELS88 1999
Own Characteristics						
Some College	0.043	0.398	0.245	0.246	0.184	0.158
College Graduate	0.196	0.256	0.287	0.335	0.226	0.433
Math Centile Rank	0.558	0.514	0.534	0.529	0.514	0.505
Married	0.646	0.824	0.568	0.709	0.434	0.429
Child Under 6	0.424	0.593	0.313	0.357	0.201	0.241
FTFY	0.338	0.529	0.528	0.520	0.496	0.612
Full-time Experience (Years)	.	7.156	3.596	7.348	2.965	.
Age	22.97	28.96	24.94	31.29	23.50	25.20
High School Region						
Northeast	0.243	0.249	0.269	0.277	0.249	0.217
West	0.130	0.152	0.161	0.157	0.151	0.148
South	0.267	0.254	0.244	0.250	0.260	0.315
Father's Education						
High School	0.226	0.203	0.324	0.334	0.294	0.283
Some College	0.108	0.110	0.190	0.176	0.240	0.191
College Graduate	0.124	0.120	0.208	0.198	0.257	0.332
Education Missing	0.149	0.166	0.002	0.005	0.023	0.108
Mother's Education						
High School	0.326	0.319	0.463	0.467	0.428	0.351
Some College	0.115	0.116	0.182	0.172	0.260	0.225
College Graduate	0.093	0.096	0.125	0.119	0.152	0.262
Education Missing	0.125	0.128	0.002	0.004	0.018	0.087
N	30415	21574	9906	7353	4303	5780

Notes: Statistics for white high school graduates from the NELS surveys. The omitted category in high school region is the Midwest and the omitted category in Father's and Mother's Education is less than high school graduate.

Table 3: Across Cohort Changes in Selection

	Talent11 1966 Age 23	NLS72 1979 Age 25	HSB 1986 Age 24	NELS88 1999 Age 25
Panel A. Female	(1)	(2)	(3)	(4)
Math Centile Rank	0.1418*	0.1073**	0.2011**	0.0556
	(0.0570)	(0.0294)	(0.0493)	(0.0468)
% FTFY	0.226	0.421	0.443	0.544
N	8075	5022	2229	2849
Panel B. Male	(1)	(2)	(3)	(4)
Math Centile Rank	-0.1082	0.0821**	0.1316*	-0.0184
	(0.0732)	(0.0291)	(0.0535)	(0.0466)
% FTFY	0.419	0.638	0.553	0.682
N	7484	4722	1846	2626

Specifications control for education, parents' education, and high school region and are weighted using sampling weights. Robust standard errors are in parentheses. Significance levels are + p<0.10, * p<0.05, ** p<0.01

Table 4: Within Cohort Changes in Selection

	Talent11		NLS72					
	1966 Age 23	1972 Age 29	1977 Age 23	1978 Age 24	1979 Age 25	1984 Age 30	1985 Age 31	1986 Age 32
Panel A. Female	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math Centile Rank	0.1418*	0.0216	0.0484+	0.0868**	0.1073**	0.1448**	0.1033*	0.1381**
	(0.0570)	(0.0473)	(0.0289)	(0.0293)	(0.0294)	(0.0423)	(0.0409)	(0.0430)
% FTFY	0.226	0.246	0.414	0.435	0.421	0.416	0.423	0.382
N	8075	10993	4996	5012	5022	3564	3568	3575
Panel B. Male	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math Centile Rank	-0.1082	0.0072	-0.0862**	-0.0066	0.0821**	0.1103**	0.1457**	0.2336**
	(0.0732)	(0.0395)	(0.0295)	(0.0293)	(0.0291)	(0.0348)	(0.0323)	(0.0381)
% FTFY	0.419	0.833	0.565	0.633	0.638	0.753	0.772	0.683
N	7484	10230	4707	4707	4722	3329	3321	3321

Specifications control for education, parents' education, and high school region and are weighted using sampling weights. Robust standard errors are in parentheses. Significance levels are + p<0.10, * p<0.05, ** p<0.01

Table 5: Across Cohort Changes in Selection, Controlling for Experience

	NLS72 1979 Age 25	HSB 1986 Age 24	Talent11 1972 Age 29	NLS72 1986 Age 32
Panel A. Female	(1)	(2)	(3)	(4)
Math Centile Rank	0.0420+	0.1786**	0.0377	0.0352
	(0.0240)	(0.0420)	(0.0406)	(0.0446)
Full-time Experience	0.1504**	0.1564**	0.0720**	0.0743**
	(0.0026)	(0.0048)	(0.0031)	(0.0018)
FTFY	0.421	0.443	0.246	0.382
N	5022	2229	10993	3575
Panel B. Male	(1)	(2)	(3)	(4)
Math Centile Rank	0.1123**	0.1934**	0.0475	0.1765**
	(0.0264)	(0.0464)	(0.0395)	(0.0333)
Full-time Experience	0.1500**	0.1420**	0.0345**	0.0752**
	(0.0040)	(0.0057)	(0.0056)	(0.0036)
FTFY	0.638	0.553	0.833	0.683
N	4722	1846	10230	3321

Specifications control for education, parents' education, and high school region and are weighted using sampling weights. Robust standard errors are in parentheses. Significance levels are + p<0.10, * p<0.05, ** p<0.01

Table 6: Selection by Marital Status

	Talent11 1966 Age 23	NLS72 1979 Age 25	HSB 1986 Age 24	NELS88 1999 Age 25	Talent11 1972 Age 29	NLS72 1986 Age 32
Panel A: FTFY by Marital Status						
Math Centile Rank	0.0758	-0.0327	0.1163+	-0.0476	0.0580	0.3420**
	(0.0931)	(0.0450)	(0.0704)	(0.0596)	(0.1174)	(0.0740)
Math Centile Rank X Married	0.0866	0.2242**	0.1411	0.2137**	-0.0358	-0.2619**
	(0.0981)	(0.0530)	(0.0900)	(0.0786)	(0.1201)	(0.0796)
Married	-0.2047**	-0.3500**	-0.1413**	-0.2414**	-0.3793**	-0.0695
	(0.0716)	(0.0324)	(0.0536)	(0.0431)	(0.0734)	(0.0505)
FTFY	0.226	0.421	0.444	0.545	0.246	0.383
N	8075	5019	2222	2844	10993	3559
Panel B: Selection into Marriage						
Math Centile Rank	-0.0051	0.0120	-0.0613	-0.0157	0.0183	0.0791+
	(0.0529)	(0.0285)	(0.0486)	(0.0467)	(0.0424)	(0.0405)
Married	0.753	0.628	0.513	0.489	0.831	0.730
N	8075	5019	2222	2844	10993	3559

Specifications control for education, parents' education, and high school region and are weighted using sampling weights. Robust standard errors are in parentheses. Significance levels are + p<0.10, * p<0.05, ** p<0.01

Table 7: Selection and Fertility

	Talent11 1966 Age 23	NLS72 1979 Age 25	HSB 1986 Age 24	NELS88 1999 Age 25	Talent11 1972 Age 29	NLS72 1986 Age 32
Panel A: FTFY and the Presence of Children Under 6						
Child Under 6	-0.3360**	-0.3971**	-0.2867**	-0.2979**	-0.3509**	-0.2104**
	(0.0223)	(0.0147)	(0.0283)	(0.0263)	(0.0238)	(0.0182)
N	15487	5022	2253	2849	10993	3622
Panel B: Selection into Having a Child Under 6						
Math Centile Rank	-0.0336	-0.0634*	-0.0820+	-0.0603	-0.0176	0.1068**
	(0.0700)	(0.0278)	(0.0427)	(0.0403)	(0.0531)	(0.0375)
% Child Under 6	0.510	0.364	0.253	0.297	0.599	0.359
N	8075	5022	2229	2849	10993	3575

Specifications control for education, parents' education, and high school region and are weighted using sampling weights. Robust standard errors are in parentheses. Significance levels are + p<0.10, * p<0.05, ** p<0.01

Appendix I: Data

We use data from four separate longitudinal surveys of high school students: Project Talent (Talent) for the classes of 1960-1964, the National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HS&B) for the class of 1980, and National Education Longitudinal Study of 1988 (NELS-88) for the class of 1992. These studies first surveyed students in middle school (NELS-88) or high school (Talent, NLS-72, and HS&B) and re-surveyed students into early adulthood. The baseline surveys contain student demographic characteristics and scores on a number of aptitude tests, while the follow-up surveys provide information on educational attainment, labor market experiences, and family formation in the years following high school graduation. For the purposes of comparison, we also use data on the labor market outcomes of these cohorts from the Current Population Survey (CPS).

Project Talent Project Talent is a national longitudinal survey of students who attended grades 9-12 in the U.S. in 1960. The original sample contained over 400,000 students, and the follow-up surveys contain information on about 20,000 students from each initial grade. Follow-up surveys were conducted one-, five-, and 11-years after the expected year of high school graduation, when students were, respectively, approximately 19, 23, and 29 years of age. Data access was provided by the American Institutes for Research.

National Longitudinal Study of the High School Class of 1972 NLS-72 is the first national longitudinal education study conducted by the U.S. Department of Education and provides information on students who were seniors in high school in the spring of 1972. The base-year survey was administered in 1972, and five follow-up surveys were conducted in 1973, 1974, 1976, 1979 and 1986 (one-, two-, four-, seven-, and 14-years after the first interview). We focus primarily on the fourth follow-up in 1979 (n≈19,000) and the fifth follow-up in 1986 (n≈13,000), when students had reached about 25 and 32 years of age, respectively.

High School and Beyond HS&B is the second national longitudinal education study conducted by the U.S. Department of Education and included two cohorts: the 1980 senior class and the 1980 sophomore class. We focus on the approximately 30,000 students in the senior cohort, who were administered follow-up surveys in 1982, 1984, and 1986 (two-, four- and six-years after the baseline survey).[13] We focus primarily on the 1986 survey, when most respondents were 24 years old.

National Education Longitudinal Study of 1988 NELS-88 is the third national longitudinal education study conducted by the U.S. Department of Education and its base-year sample consists of students who were enrolled in the eighth grade in the spring of 1988. The first follow-up occurred in 1990, when most respondents were in 10th grade, and the second follow-up took place in 1992, when most respondents were in their senior year of high school. The second follow-up sample was freshened to be representative of high school seniors in 1992, making trend comparisons to Talent, NLS-72, and HS&B possible. The third and fourth follow-ups occurred in 1994 and 2000 (two- and eight- years after the second follow-up). We focus on the 1999 retrospective measures from the fourth follow-up, when students were approximately 25 years old.

Current Population Survey The CPS is a monthly survey of approximately 50,000 households conducted jointly by the Census Bureau and the Bureau of Labor Statistics. It is the main source of information on the characteristics of the labor force in the U.S. We use public use microdata from the CPS March Supplement, or the Annual Demographic File; we accessed these data through IPUMS-CPS, which harmonizes variables in order to facilitate cross-year comparisons.

Using CPS data from 1962 to 2011, we construct labor market histories of the 1940 to 1975 birth cohorts. The synthetic birth cohorts are generated using information on age in the survey year. Following Mulligan and Rubinstein (2008), we restrict

[13] The sophomore cohort was also surveyed in 1992, 12 years from the first interview, when students had reached the age of 28. While 1992 follow-up contains information on annual earnings, information on the number of hours worked is not available.

the sample to non-Hispanic whites and drop self-employed individuals. We further restrict the analysis to high school graduates (high school degree or more) for purposes of comparison to the longitudinal surveys of high school students. Again following Mulligan and Rubinstein (2008), we impute annual weeks for the 1962-1975 data using the 1976-1978 averages within weeks (in categories), gender and race bins and impute usual weekly hours for the same period using the 1976-1978 averages within weeks, gender, race and full-time/part-time bins. Full-time full-year work is defined as working at least 50 weeks in a year and usually working 35 or more hours per week. We drop observations with imputed values for earnings, and avoid outliers by trimming hourly wage rates below the 1st or above the 99th percentile of their distributions within each survey year. Wages rates are deflated to year 2000 using the CPI index. Gender wage gaps are estimated through a hourly wage regression on a gender dummy, two education dummies (some college and college of more relative to high school graduate) and a quartic polynomial on potential experience (age - years of education - 7). Following the years in which wage and employment information are available in the education surveys, we compare the Talent 5-year follow up to the 1966 CPS, the Talent 11-year follow up to the 1972 CPS, the NLS-72 fourth follow up to the 1980 CPS, the NLS-72 fifth follow up to the 1987 CPS, the HSB third followup to the 1987 CPS, and the NELS-88 follow-up to 2000 CPS data. In each comparison, we restrict age ranges in CPS data so as to match the corresponding ages in each education survey.

* Explain how CPS is matched to the NELS cohorts

Appendix II: Variables

Demographics We construct sex and race (white, black, hispanic, other) from composite variables in the surveys and generate age from information on students' birthdays.

Family Background The educational attainment of students' mothers and fathers is coded into five categories: less than high school, high school graduate, some college, college graduate or more, and missing.

Region We classify students into four geographic regions (Northeast, Midwest, South, and West) based on the locations of their high schools.

Test Score As a measure of cognitive ability, we use the students' centile ranking within the distribution of white high school students, based on the average of their scores on math and verbal cognitive exams.[14] Our measures of math and verbal cognitive ability are described in detail below:

- Talent: In the base-year, students completed tests over two school days. For math, we use the math composite, which is composed of scores on arithmetic reasoning, introductory mathematics, and advanced mathematics. For verbal, we use the verbal composite, which is composed of scores in literature, vocabulary, and English. Following the other surveys, we standardize each of these scores to have a mean of 50 and a standard deviation of 10 in each grade before we construct the composite score from the average of math and verbal.
- NLS-72: In the base-year, students completed a 69-minute battery of six tests. For math, we use the scaled score (mean=50, sd=10) from the mathematics test, a 15-minute, 25-item section on quantitative comparisons. For verbal, we use the scaled scores (mean=50, sd=10) from the vocabulary test, a 5-minute, 15-item section on synonyms, and the reading test, a 15-minute, 20-item section on analysis, interpretation, and comprehension.[15]
- HS&B: The test battery for seniors consisted of 9 parts. For math, we use the scaled score (mean=50, sd=10) from the math exam, a 19-minute, 33-item section

[14] If either the math or verbal score is missing, we use the non-missing score. This follows the National Center for Education Statistics procedure for constructing composite scores, as described in the technical appendix to the High School and Beyond survey.

[15] To construct a verbal score, we take the average of the non-missing scores from vocabulary and reading, weighting by the number of items in each section (15 items in vocabulary, 20 items in reading), or take the non-missing score.

on quantitative comparisons.[16] For verbal, we use the scaled scores (mean=50, sd=10) from the vocabulary test, a 9-min, 27-item section on synonyms, and the reading test, a 15-minute, 20-item section on analysis, interpretation, and comprehension.[17]

- NELS-88: In the base-year though second follow-up, students completed four subject tests in 85-minutes. To minimize ceiling and floor effects, students in the first and second follow-ups received different versions of the tests based on their prior scores; NCES constructed comparable follow-up scores under the assumptions of item response theory. For math, we use the second follow-up item response theory scaled score (mean=50, sd=10) from a 30-minute, 40-item math section, which included word problems, graphs, equations, quantitative comparisons, and geometric figures. For verbal, we use the second follow-up item response theory scaled score (mean=50, sd=10) from a 21-minute, 21-item reading section, which tested students' ability to understand words in context, identify figures of speech, interpret the authors perspective, and evaluate the passage as a whole.

Education We classify educational attainment into three categories: high school, some college, and college or more. Since surveys were selected to be nationally representative at different grades, we restrict our attention to high school graduates, a sample we believe to be relatively stable over this time period. Ideally, we would be able to analyze selection into the labor market for all individuals, including high school drop outs. However, we believe that our analysis is still informative since estimates of gender wage gaps typically compare individuals with similar levels of educational attainment.

Wage and Employment We use data on individuals' most recent employment and their retrospective reporting of previous employment to generate measures of wages, hours,

[16]The math test was in two parts; we take the average of the non-missing scores, weighting by the number of items in each section (25 items in Part I, 8 items in Part II), or take the non-missing one.

[17]To construct a verbal score, we take the average of the non-missing scores from vocabulary and reading, weighting by the number of items in each section (15 items in Vocabulary Part I, 12 items in Vocabulary Part II, 20 items in Reading), or take the non-missing score.

and weeks worked in the last year. Wage questions differ slightly between surveys but are coded as follows to generate individuals' approximate hourly wage:

- Talent 5-year Follow up: For seniors, wage on October 1, 1965; for juniors, sophomores, and freshmen, wage between October 1965 (1966) (1967) and September 1966 (1967) and (1968), respectively[18]
- Talent 11-year Follow up: Wage on September 30, 1971 for seniors, and wage on September 1 of 1972, 1973, and 1974 for juniors, sophomores, and freshmen, respectively[19]
- NLS-72 Fourth Follow up: Wage between October 1976 (1977) (1978) and October 1977 (1978) (1979)[20]
- NLS-72 Fifth Follow up: Wage at the most recent job within the 12 month period prior to the 1986 survey date (e.g., between April 1985 and March 1986), and wages between January and December 1984 (1985)[21]
- HS&B Third Follow up: Wage at most recent job within the 12 month period prior to the 1986 survey date (e.g., between April 1985 and March 1986)[22]
- NELS-88 Fourth Follow up: Most recent wage in 2000 (measured between January 1999 and the survey date) and wage between January and December 1999[23]

[18]For seniors, wages were calculated using the reported pay basis (monthly, weekly, hourly). For workers who were paid weekly, hourly wage was calculated by dividing weekly earnings by the usual hours per week. For workers who were paid monthly, it was calculated by dividing the monthly earnings by the product of usual hours per week and 52/12. For all other grades, wages were calculated by dividing the total earnings by the product of the number of weeks worked and the number of hours worked per week.

[19]Wages were calculated using the reported pay basis as for seniors in the Talent 5-year follow up.

[20]Wages were calculated by dividing weekly earnings by the usual hours per week.

[21]Wage in 1986 are calculated using the reported pay basis (hourly, weekly, biweekly, monthly, and annually); the calculations are the same as before, except that we divide by 2 times the usual hours for biweekly payment and 52 times the usual hours for annual payment. For 1985 (1984), we calculate the wage by dividing the annual earnings in 1985 (1984) by the usual hours times the number of weeks worked.

[22]Wages in 1986 are calculated using the reported pay basis (hourly, weekly, biweekly, monthly, and annually) as in the NLS-72 Fifth Follow up.

[23]The survey date is not reported in the data but is between January and September 2000. Wages in 2000 are calculated using the reported pay basis (hourly, weekly, biweekly, monthly, and annually) as in 1986 for the NLS-72 and HS&B. Wages in 1999 are calculated by dividing annual earnings in 1999 by the usual hours times the number of weeks worked.

Hours is coded as follows:

- Talent 5-year Follow up: Average hours worked per week on the job in October 1965 (1966) (1967) 1968)
- Talent 11-year Follow up: Average hours worked per week on the job in September 1971 (1972) (1973) (1974)
- NLS-72 Fourth Follow up: Average hours worked per week on the most recent job held between November 1976 (1978) (1979) and October 1977 (1978) (1979)
- NLS-72 Fifth Follow up: Average hours worked per week at most recent job in 1986; for 1984 and 1985, we use information on the start and end dates and current/last hours worked for the past 4 jobs (with a maximum of 8 spells each) to construct a labor market history with mutually exclusive spells of employment.[24] Hours is defined as the total number of hours worked in these jobs between January and December divided by the number of weeks worked.
- HS&B Third Follow up: Average hours worked per week at the most recent job in 1986
- NELS-88 Fourth Follow up: Average hours at most recent job in 2000, average hours per typical week in 1999 (all jobs, not just primary).[25]

The reference period for weeks is as follows:

- Talent 5-year Follow up: For juniors, sophomores, and freshmen, weeks worked in the last 12 months, from the October 1965 (1966) (1967) to October 1966 (1967) (1968). Weeks are not available for seniors.
- Talent 11-year Follow up: Weeks worked in the last 12 months (October 1970-September 1971 for seniors, September 1971 (1972) (1973) to August 1972 (1973) (1974) for juniors, sophomores, and freshmen)

[24]When spells of employment overlap, we choose the job with the greatest number of hours.

[25]There was an interview error where some individuals whose jobs ended in 1999 were not asked this reference question, so these observations were flagged.

- NLS-72 Fourth Follow up: Weeks worked in the last 12 months between October 1976 (1977) (1978) and October 1977 (1978) (1979)
- NLS-72 Fourth Follow up: Weeks worked in the last 12 months before the 1986 survey date (February through July); weeks worked between January and December 1984 (1985) are constructed from the labor market history
- HS&B Third Follow up: Weeks worked in the last 12 months before the survey date (February through July) is constructed from the labor market history as in the NLS-72 Fourth Follow up
- NELS-88 Fourth Follow up: No measures of weeks for 2000; number of weeks worked between January and December 1999.[26]

We consider workers as being full-time, full-year (FTFY) if they work at least 35 hours per week and at least 50 weeks per year (or use 11 months because of the reporting by months?). Some surveys also contain information on workers' occupations and industries.

Work Experience Since most follow-up surveys collected retrospective information on job histories, we use the information on hours and weeks worked to generate measures of work experience. Due to differences in questions between surveys, it is not possible to make the variables for work experience completely consistent over time, but the measures still provide evidence on whether there are differences in labor force attachment between cohorts. Our measures of experience are as follows:[27]

- Talent 5-year Follow up: Indicator for being continuously employed since leaving school

[26]There was an interview error where some individuals whose jobs ended in 1999 were not asked this reference question, so these observations were flagged.

[27]For those with two or more jobs in the same period, we calculate experience using the job with the greatest number of hours, so that the sum of part-time and full-time experience does not exceed potential experience. We do not differentiate between civilian and military labor market experience. Because of the questioning in Talent, I thought we might want to construct a similar variable using the full-time/part-time measures in the later surveys. The problem is how to do this; would one weight the number of part-time years by the number of part-time hours/35?

- Talent 11-year Follow up: Years of full-time experience since the June after high school graduation[28]
- NLS-72 Fourth Follow up: Years of full-time and part-time experience since the June after high school graduation
- NLS-72 Fifth Follow up: Years of full-time and part-time experience since the June after high school graduation
- HS&B Third Follow up: Years of full-time and part-time experience since the June after high school graduation
- NELS-88 Fourth follow up: Years of full-time and part-time experience between the June after high school graduation (1992) and June 1994; indicators for being employed full-time or part-time more for than 6-months in 1997 and 1998

Marital History and Fertility Marital status is coded into three categories: married; single; and divorced, widowed or separated. The information asked about respondents' spouses varies across surveys but includes spouses' education and job characteristics (occupation, industry, hours worked, wages, earnings). We use information on the number of children and childrens' birthdays to generate an indicator for the presence of one or more children under the age of 6 years in the household.

Weights We use the provided survey weights to generate estimates that are nationally representative and which account for attrition in the longitudinal data.[29] Weights are as follows:

- Talent 5-year Follow up: Cross-section weight - weight b, which up-weights respondents to account for non-response, Panel weight: NA

[28]Respondents who had done part-time work were asked to include this in their answer by "estimat[ing] how much full-time work [this part-time work] was equivalent to."

[29]In Project Talent, researchers pursued and obtained responses from a sample of individuals who initially did not respond to the follow-up survey. The Talent weights assume that these individuals are representative of other non-respondents with similar characteristics. Weights in NLS72, HS&B, and NEL-88 assume that non-respondents are similar to respondents with the same observable characteristics.

- Talent 11-year Follow up: Cross-section weight - weight b, which up-weights respondents to account for non-response, Panel weight: NA
- NLS-72 Fourth Follow up: Cross-section weight - fourth follow up weight; Panel weight: Base year through fourth follow up
- NLS-72 Fifth Follow up: Cross-section weight - fifth follow up weight; Panel weight: NA
- HS&B Third Follow up: Cross-section weight - testwt3 - weight applied to cases with third follow-up data and enough base year test data to allow construction of test composite, Panel weight: Base year through third follow up
- NELS-88 Fourth follow up: Cross-section weight - 4th follow up weight for 1992 high school diploma graduates, Panel Weight: panel weight for second through fourth follow up

Appendix III: Sample Selection

We restrict the sample to white high school graduates who were 16-19 years-old in their senior year of high school.

References

Joshua Angrist and William Evans. Children and their Parents' Labor Supply: Evidence from Exogenous VAriation in Family Size. *American Economic Review*, 88(3):450–477, Jun 1998.

Francine Blau and Lawrence Kahn. The us gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review*, 60(1):45–66, 2006.

Francine D. Blau and Lawrence M. Kahn. Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics*, 15(1):1–42, Jan 1997.

Richard Blundell, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. Changes in the distribution of female and male wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363, March 2007.

Sean P. Corcoran, William N. Evans, and Robert M. Schwab. Women, the labor market and the declining quality of teachers. *Journal of Policy Analysis and Management*, 23(3): 449–470, 2004.

Lena Edlund and Cecilia Machado. Pill Power: the Prequel. IZA discussion paper no. 5468, 2011.

Nicole Fortin. The gender wage gap among young adults in the united states: The importance of money vs. people. University of British Columbia, 2007.

Nicole Fortin, Philip Oreopoulos, and Shelley Phipps. Leaving Boys Behind: Gender Disparities in High Academic Achievement. 2012.

Claudia Goldin and Lawrence F. Katz. The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 110(4):730–770, August 2002.

Claudia Goldin, Lawrence F. Katz, and Ilyana Kuziemko. The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, 20(4):133–156, Fall 2006.

Reuben Gronau. Wage comparison - a selectivity bias. *Journal of Political Economy*, 82(6): 1119–1143, 1974.

James Heckman. Shadow price, market wages, and labor supply. *Econometrica*, 42(4): 679–694, July 1974.

Chinhui Juhn. Decline of Male Labor Market Participation: The Role of Declining Market Opportunities. *Quarterly Journal of Economics*, 107(1):79–121, Feb 1992.

Chinhui Juhn and Kevin M. Murphy. Wage inequality and family labor supply. *Journal of Labor Economics*, 15(1):72–97, January 1997.

Peter Kuhn and Catherine Weinberger. Leadership skills and wages. *Journal of Labor Economics*, 23(3):395–436, July 2005.

Cecilia Machado. Selection, Heterogeneity and the Gender Wage Gap. 2009.

Thomas A. Mroz. The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica*, 55(4):765–799, Jul 1987.

Casey Mulligan and Yona Rubinstein. Selection, investment and women's relative wage over time. *Quarterly Journal of Economics*, 123(3):1061–1110, August 2008.

Richard J. Murnane, John B. Willet, and Frank Levy. The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2):251–266, May 1995.

Derek Neal. The measured black-white wage gap among women is too small. *Journal of Political Economy*, 112(1):S1–S28, 2004.

Derek A. Neal and William R. Johnson. The role of pre-market factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895, October 1996.

Claudia Olivetti and Barbara Petrongolo. Unequal pay or unequal employment? a cross-country analysis of gender gaps. *Journal of Labor Economics*, 26(4):621–654, October 2008.

June O'Neill. The Gender Gap in Wages, circa 2000. *Journal of Labor Economics*, 93(2): 309–314, May 2003.

June O'Neill and Solomon Polachek. Why the Gender Gap in Wages Narrowed in the 1980. *Journal of Labor Economics*, 11(1):205–228, Jan 1993.

John Pencavel. Mating by Schooling and the Work Behavior of Wives and Husbands. *American Economic Review Papers and Proceedings*, 88(2):326–329, May 1998.

Zhenchao Qian. Changes in Assortative Mating: The Impact of Age and Education. *Demography*, 35(3):279–292, Aug 1998.

James P. Smith and Michael P. Ward. Women in the Labor Market and in the Family. *Journal of Economic Perspectives*, 3:9–23, 1989.

Bruce Weinberg. Computer Use and the Demand for Women Workers. *Industrial and Labor Relations Review*, 53:290–308, 2000.

Catherine Weinberger and Peter Kuhn. Changing levels or changing slopes? the narrowing of the u.s. gender earnings gap, 1959-1999. *Industrial and Labor Relations Review*, 63(3): 384–406, July 2010.