Changes in the Black-White Wage Gap — the Role of Selection Revisited

Yuanyuan Chen *

May 8, 2005

Abstract

This paper examines the extent to which non-random selection in labor market affects assessing the black-white wage difference by applying a longitudinal method of imputing wages for nonworkers. Since fixed effects are crucial in determining the wages of nonworkers, the cross-sectional methods of controlling only on observables understate the impact of selection. Using the Panel Study of Income Dynamic (PSID) 1970 - 2000 data, I find that the selective bias can explain 40 percent of the observed change in the black-white wage gap over 1970-2000.

JEL Classification: J22, J21, C2

*Author's Email Address: ychen@imf.org
I am grateful to Peter Gottschalk, Author Lewbel, Donald Cox, Shang-jin Wei and all participants in the seminar at Boston College Economics Department for helpful guidance and comments. All errors are mine.

1 Introduction

It is well documented that during the 1960s and 1970s, the black-white wage difference in the United States narrowed rapidly. According to Juhn (1991), in 1963, the average black male earned about 63 percent as much as a white worker; and this ratio increased to 75 percent in the late 1970s. The rapid convergence, however, did not continue in the early 1980s when the relative wage for black men declined to 70 percent in 1984. During the 1990s, the black-white wage ratio once again began to increase and reached 79 percent in 1998, the highest value in the history.[1]

Whether the observed change in the mean wages reflects a change in opportunity or a selection bias has been much debated in the literature. Heckman and Butler (1977) argue that a significant portion of the black-white wage convergence is due to the changing composition of working blacks. While the wages for black men were rising faster compared to their white counterparts, their relative employment rates were falling. This decline in labor market participation was especially pronounced among less-educated and low-wage workers. Therefore, the observed black-white wage gap might overstate the gains in the opportunities open to blacks.

Using the Panel Study of Income Dynamic (PSID) for 1970 - 2000 data, this paper examines the selection bias and its impact on measures of the black-white wage gap. This paper differs from the previous studies by exploiting the longitudinal nature of the data. The merit of longitudinal data with multiple observations per individual is that it captures

[1] Couch and Daly (2000).

the individual specific time-invariant "ability", such as personal working habit and family background, which is crucial to determining wages but has not been emphasized in the previous cross-sectional analysis. Here, I apply the Wooldridge (1995) panel estimation method to correct for the selection bias and analyze the selection based on observed variables, such as age and education level, unobserved individual effects, and transitory factors, respectively. Wages of nonworkers are predicted from the panel estimation to examine the extent to which the selection bias has affected the observed black-white wage gap.

I find that nonrandom selection can explain 65 percent of the observed wage convergence between 1970 and 1980 in the PSID data and 40 percent of the change in the racial wage gap between 1970 and 2000. The impact of the selection is most pronounced for the least educated group who experienced the most severe decline in employment rate. Moreover, my estimates confirm that the selection on individual specific effect is quite large. Given the same observed characteristics, those who are more often employed have much higher average over time wages than those who are less often employed. In particular, the average white nonworker would have earned 35 percent less than the average worker, conditional on observed characteristics; the average black nonworker earned 41 percent less. Therefore, previous studies based on cross-sectional analysis without accounting for individual effects could result in a serious bias in imputing wages of nonworkers.

The paper is organized as follows: The next section reviews what has been done in previous studies. Section 3 proposes an econometric model and a technique used to predict the wages for nonworkers, followed by a discussion of measurement issues in Section 4.

Section 5 presents the main results in panel data analysis. Section 6 compares the results estimated by the panel method with those by cross-sectional methods.

2 Literature Review

Since Heckman and Butler(1977), numerous studies have tried to examine the extent to which selection bias affects the black-white wage gap. All these studies are based on cross-sectional data and impute wages of nonworkers either by making strong assumptions about where they would have been in the wage distribution had they worked or by matching on observables. In one of the earliest studies, Brown (1984) imputes the median of "uncensored" distribution for aggregate Current Population Survey (CPS) data, by assuming that nonworkers earned less than the median agent in the underlining wage distribution. He finds that two-thirds of the observed black-white earning convergence from 1953 to 1978 can be attributed to selection. Under this fairly strong assumption, Brown (1984) provides an upper bound for the impact of selection on the black-white wage convergence.

Motivated by the magnitude of Brown's results, many researchers have attempted to explore the impact of selection bias in more details. However, there is little consensus on whether selection bias has significant large impact on black-white wage gaps. On one side, some studies show that selective bias is trivial to the trend of racial wage gaps. Vroman (1986) assumes that the median earnings of dropouts are 85 percent of the median workers of the same age and race and finds that over the period of 1964-1985, the selective withdrawal of blacks overstates the racial wage convergence by only 14 percent, which is much smaller than the estimate of 66 percent in Brown (1984). Vroman rejects the selective withdrawal hypothesis by demonstrating that the labor market dropouts who are transfer recipients have higher earnings than workers. Smith and Welch (1986) and Welch (1990) match respondents to the March CPS in adjacent years and compare the earnings of workers who worked one year and not the next, or vice-versa, and they do not find that labor market entrants and exiters earn less than workers. They conclude that selection bias accounts for only a negligible portion of the racial wage convergence.

Several studies, however, show a significantly large impact of selective withdrawal bias. Blau and Beller (1992) impute wages of nonworkers using a regression-matching estimator combined with the assumption that nonworkers receive wage offers only 60 to 80 percent as high as workers, after conditioning on education and experience. They find that wage gains especially for young black men and women in the 1970s may have been overstated. Neal and Johnson (1996) estimate the racial wage gaps in median regression by assigning nonworkers a wage rate below the median (zero) and show that inclusion of nonworkers significantly changes the estimate of the racial wage gap. Juhn (2002) imputes the wages for nonworkers by matching on education, experience, and weeks worked. She concludes that selection bias reduces the racial wage convergence by about 30 percent between 1968 and 1988. Chandra (2000), Johnson, Kitamura and Neal (2000), and Heckman, Lyons and Todd (2000) also provide evidence that is consistent with the selective withdrawal hypothesis. More recently, Chandra (2003) argues that making inference on CPS data understates selection bias since it does not contain information on the incarcerated population. The selectivity corrected estimate in census data indicates complete stagnation of the racial wage gap between 1970-1990 with a divergence of 3.5 to 6 percentage points between 1980-1990.

3 Model

Although the low educated population are more likely to be unemployed in labor market, there is no necessary implication that an unemployed person would have lower potential wages than an employed person. The reason why one person is unemployed could be that he/she has relatively high reservation wage. In the beginning of this section, I define the labor market selection problem in a statistical model that addresses the relationship between reservation wages, offered wages and labor market participation. Then I briefly discuss the approaches under different types of selection, and emphasize on the Wooldridge (1995) method which is used to impute wages of nonworkers in the empirical work in section 5.

3.1 Reservation wage and offered wage

To compute the wages for nonworkers, I first pin down the basic question: what makes an individual decide to work? Search theory suggests that everyone has a reservation wage (W^r) which is the lowest wage offer (W^o) that the person would accept. Hence, an individual works only if $W^0 > W^r$. Wage offers are determined by personal productivity characteristics (x_{it}^0), such as education and age, and an individual effect (α_i^o). Reservation wages depend on age, education and other factors that affect the reservation wages (x_{it}^r), such as family background, and personal permanent preference to work (α_i^r).

$$
\begin{aligned}
W_{it}^r &= x_{it}^r\beta^r + \alpha_i^r + \varepsilon_{it}^r \\
W_{it}^0 &= x_{it}^0\beta^0 + \alpha_i^0 + \varepsilon_{it}^o
\end{aligned} \tag{1}
$$

Hence, the propensity to work (y_{it}^*) can be defined as the difference of these two,

$$
y_{it}^* = (x_{it}^0\beta^0 - x_{it}^r\beta^r) + (\alpha_i^0 - \alpha_i^r) + (\varepsilon_{it}^o - \varepsilon_{it}^r) \tag{2}
$$

A person only works ($y_{1it} = 1$) if his/her propensity to work is positive ($y_{1it}^* >= 0$). The potential wage offers (W_{it}^*) are observed only for workers. To simplify, the model can be written as follows:

$$
y_{1it}^* = x_{1it}\beta_1 + \alpha_{1i} + \varepsilon_{1it} \tag{3}
$$

$$
W_{it}^* = x_{2it}\beta_1 + \alpha_{2i} + \varepsilon_{2it} \tag{4}
$$

$$
\begin{aligned}
y_{1it} &= 1 \text{ if } y_{1it}^* >= 0 \\
W_{it} &= W_{it}^* * y_{1it}
\end{aligned}
$$

where $x_{1it} = x_{it}^o - x_{it}^r$,$\alpha_{1i} = \alpha_i^0 - \alpha_i^r$, $\varepsilon_{1it} = \varepsilon_{it}^0 - \varepsilon_{it}^r$, $x_{2it} = x_{it}^o$, $\alpha_{2i} = \alpha_i^0$,and $\varepsilon_{2it} = \varepsilon_{it}^o$.

3.2 Method to correct for selection bias

In the following section, I explore different types of estimates for the objective equation (4), from the simplest case where there is no selection bias to the most complicated where selection operates through all components of the equation.

In the simplest case where selection operates only through observables (x_{it}) that appear in the wage equation, an ordinary least square (OLS) regression on (4) generates consistent estimates. In the case of selection on unobservables ($cov(\alpha_{2i}, \alpha_{1i}) \neq 0$ or $cov(\varepsilon_{1it}, \varepsilon_{2it}) \neq 0$), the estimates from OLS are biased.

There are several econometric models for correcting selection bias. Chart 1 shows how these models are related to each other. When there is no variation among the individual effects ($var(\alpha_{2i}) = 0$), cross-sectional estimates are consistent and the typical method to correct selection problem is the Heckman (1979) model, when the transitory factors in structural and participation equations ($\varepsilon_{1it,}\varepsilon_{2it}$) are assumed to be joint normal. By relaxing the restriction of homogeneity of personal effects across individuals ($var(\alpha_{2i}) \neq 0$), panel estimates are necessary. When there is no selection on unobservables, ($cov(\alpha_{1i}, \alpha_{2i}) = 0$ & $cov(\varepsilon_{1it}, \varepsilon_{2it}) = 0$), fixed effect or random effect estimates are unbiased.[2] If further relaxing the assumption and allowing the selection on personal effects, ($cov(\alpha_{1i}, \alpha_{2i}) \neq 0$ & $cov(\varepsilon_{1it}, \varepsilon_{2it}) = 0$), fixed effect estimates are still consistent, as suggested by Verbeek and Nijman (1992a,b). In the case where selection operates through both individual effects and idiosyncratic errors, it is impossible to separate selection between individual effect and idiosyncratic errors in the cross-sectional selection model (Heckman, 1979). Wooldridge (1995), Kyriazidou (1997a) and Lewbel (2002) propose selectivity correction methods on the panel tobit model that solves this problem.

In this paper, I use Wooldridge (1995) model which assumes that the transitory factor in the selection equation (ε_{1it}) is normal. The spirit of correcting for selection bias using

[2] If $E(\alpha_{2i}/x_{2it}) = 0$, random effects are more efficient. Otherwise only fixed effect estimates are consistent. The Hausman test on our data suggests $E(\alpha_{2i}/x_{2it}) \neq 0$, so random effect is not considered in this paper.

Wooldridge (1995) is similar to Heckman (1979), except that, to explore the nature of longitudinal data, it captures fixed effects by a linear projection of a vector of individuals' characteristics.

Wooldridge (1995) assumes that ε_{1it} and ε_{2it} are linearly dependent

$$E[\varepsilon_{2it}|\varepsilon_{1it}, \alpha_{1i}, \alpha_{2i}, x_{2it}, y_{1it}] = \rho \cdot \varepsilon_{1it}$$

with ε_{2it} independent of all lags and leads of ε_{1it} conditional on the contemporaneous error ε_{1it}. Under this assumption, arbitrary serial dependence and heterogeneity in idiosyncratic errors of both equations (3) and (4) are allowed. Moreover, the normality assumption is only necessary for ε_{1it}, but not for ε_{2it}. The expectation of wage equation (4) can be written as:

$$E(W_{it}|\varepsilon_{1it}, \alpha_{1i}, \alpha_{2i}, x_{1it}, y_{1it}) = \alpha_{2i} + x_{2it}\beta_2 + \rho \cdot \varepsilon_{1it} \tag{5}$$

The distribution of ε_{1it} cannot be estimated as it depends on α_{1i}, Wooldridge (1995) applies the Chamberlain (1980) method and assumes that individual effect is a linear projection of a person's vector of characteristics throughout the panel:[3]

$$\alpha_{1it} = \Psi_0 + \Psi_1 X_{1i1} + \Psi_2 X_{1i2} + ... + \Psi_T X_{1iT} + c_i \tag{6}$$

substituting (6) into (3):

$$y^*_{1it} = x_{1i,t}\beta_1 + \Psi_0 + \Psi_1 X_{1i1} + \Psi_2 X_{1i2} + ... + \Psi_T X_{1iT} + \xi_{it} \tag{7}$$

[3] In practice, when a panel is long (T is large) and vector x_{1it} contains a large set of variables, fixed effect (α_{1i}) is characterized by a large set of variables, which significantly reduces the degree of freedom of the estimation. One way to solve this problem is to replace the vector of individual characteristics x_{1i} by the mean $\overline{x}_{1i}$, as suggested by Mundlak (1978). The trade off is that it imposes that personal characteristics (x_{1it}) in each period have a constant influence on the fixed effect.

where $\xi_{it} = \varepsilon_{1it} + c_i$ and ξ_{it} are independent of x_{1it}. Equation (5) can be written as

$$\begin{aligned} E(W_{it}|\varepsilon_{1i}, \alpha_{1i}, \alpha_{2i}, x_{1i}, y_{1i}) &= \alpha_{2i} + x_{2it}\beta_2 + \rho \cdot \varepsilon_{1it} \\ &= \alpha_{2i} + x_{2it}\beta_2 + \rho \cdot (\xi_{i,t} - c_i) \\ &= (\alpha_{2i} - \rho \cdot c_i) + x_{2it}\beta_2 + \rho \cdot \xi_{i,t} \\ &= \varsigma_i + x_{2it}\beta_2 + \rho \cdot \xi_{it} \qquad (8) \end{aligned}$$

The estimation strategy is as follows:[4]

1. Estimate the participation equation (7) using a pooled probit, and compute the value of Mills ratio $\widehat{\lambda}_{it}$.

2. Estimate wage equation below using fixed effects regression over the sample of $y_{1it} = 1$.

$$y_{2it} = \varsigma_i + x_{2it}\beta_2 + \rho \cdot \widehat{\lambda}_{it} + e_{it} \qquad (9)$$

Finally, the wages that nonworkers would have earned can be predicted as:

$$\begin{aligned} E(W_{it}|y^*_{1it} &< 0) = \widehat{\alpha}_{2i} + x_{2it}\widehat{\beta}_2 + \widehat{\rho} \cdot E(\varepsilon_{1it}|\varepsilon_{1it} < -x_{1it}\widehat{\beta}_1 - \widehat{\alpha}_{1i}) \\ &= \widehat{\alpha}_{2i} + x_{2it}\widehat{\beta}_2 - \widehat{\rho} \cdot \frac{\phi(-x_{1it}\widehat{\beta}_1 - \widehat{\alpha}_{1i})}{\Phi(-x_{1it}\widehat{\beta}_1 - \widehat{\alpha}_{1i})} \qquad (10) \end{aligned}$$

4 Measurement

I use the data from Panel Study of Income Dynamics (PSID) 1970-2001 to track the trend of black-white wage gaps and employment, and compare the wage gaps before and after taking

[4] Wooldridge (1995) also discusses the case when $y_{1it} = \max(0, y^*_{1it})$, the Mills ratio is replaced with the predicted residual $\widehat{e}_{1it}$ obtained from the pooled tobit.

account of nonworkers.[5] For examining the impact of selection, I concentrate on the decade changes 1970-1980, 1980-90, 1990-2000. These years are comparable, since they are all at business cycle peaks (or close to business cycle peaks). The sample focuses on black and white male heads in the families who are 22-62 years old.[6] To obtain a larger dataset, both the SEO and SRC samples are included.[7] Individual weights are applied both in measuring wage trends, since the PSID individual weights are designed to combined the SRC and SEO sample to be statistical representative of the general population. Weights are not applied in the regressions.

Individuals are classified as workers if they were ever employed in the last year. Hourly wages are the reported wage rates or the annual labor income divided by hours worked if the wages are missing. Hourly wages are in 1999 dollars, using the CPI-U-RS price index.

For the self-employed, since their wage rates and total working hours are not properly reported and their total earnings depend on both labor and capital inputs, I take them as unemployed persons. Their wage rates are excluded in measuring the observed wage trend, but included in the wage trend of accounting nonworkers.[8] In this context, this paper corrects both selection between workers and nonworkers and also selection between workers

[5] Since 1968, the PSID data has been collected annually through 1997 and only biennially thereafter. The survey on hourly wage rate only started since 1970.

[6] The resulting sample size are 2,581 observations for 1970, 3,666 observations for 1980, 4,881 observations for 1990, and 4,100 observations for 2000. Between 8 and 10 percent of the weighted sample is black.

[7] At the beginning of the PSID survey, about three-fifth of the families were drawn from a representative sampling frame of the U.S, known as the "SRC" sample, and two-fifth were drawn from a set of individuals in low-income families called "SEO" sample.

[8] Juhn (2000) also imputes wages for the self-employed as for nonworkers.

and self-employed persons. The self-employed are taken as unemployed in the participation equation (7) and wage equation (9). Wages for the self-employed are predicted by equation (10), as wages of nonworkers.

I estimate wages of black and white nonworkers separately, applying the Wooldridge (1995) model described above. The first step is a probit estimate of the participation equation (7), all individuals in the sample are included. The dependent variable is whether individuals worked. The independent variables include workers' characteristics (age, age square, and a set of education dummies) which affect both offered wages and reservation wages, and the identifying exclusive variables (number of children in the family, a set of dummy variables indicating marital status, and whether parents were poor or rich), year dummies, and individual fixed effects which are characterized by a linear function of the mean of persons' own characteristics throughout the panel.

The second step is a fixed effect regression of the wage function (9). The dependent variables are the log of hourly wage rates. Besides the variables which determine offered wages and year dummies mentioned above, I also include a dummy variable which equals one if the wage rate is imputed by earnings divided by hours.

5 Panel Data Analysis

I first describe the recent trends in observed wages and employment for black and white males in the PSID data, which is the starting point for analysis. Then, I focus on analysis of selection on personal specific effects and selection on transitory factors. Finally, I compare the black-white wage gaps before and after accounting for nonworkers and calculate the impact of selection bias on the changes of racial wage gap, and conduct this analysis for different education groups.

5.1 Trends in the racial wage gap and employment

This section describes the recent trends in observed wages and employment for black and white males in the PSID data. Figure 1 shows the log differentials of hourly wage rates between black and white male workers. The racial wage gap declined in the early 1970s, from .31 in 1970 and to .27 in 1976. The gains of blacks, however, was followed by a slight setback in the late 1970s and a deterioration in the late 1980s when the relative wages for blacks declined at the rate of 1.1 percent per year. In 1989, the wage gap hit its peak of .44. The loss of wages for blacks during the 1980s has been noted by a number of papers, (Juhn, Murphy and Pierce, 1991; Smith, 1993; Card and Lemieux, 1996) with the suggestion that at least part of the slowdown can be explained by the increases in education returns and skill prices during the 1980s. In the 1990s, the black-white wage gaps fluctuated, with a decline in the early 1990s, a bounce back between 1993 and 1995, and a decline again after that. In 1998, the wage gap reached .29, the lowest point in the decade but still higher than the gaps in mid-1970s.

The trend of racial wage gaps in the PSID in the 1980s is somewhat different from other studies using CPS or census data. By using CPS data, Juhn, Murphy, and Pierce (1991) and Couch and Daly (2000) show that from 1975 to the 1980s, the black-white wage gap remained essentially unchanged until 1990 when the racial wage gap once again declined at the rate approaching one percent per year throughout the whole decade of the 1990s, with its lowest level of the history in 1998. Chandra (2003), on the other hand, also finds a divergence of black-white wage gap in census data during the 1980s, but with a lower rate at 3.5 percentage points. This difference pattern shown in the PSID and the CPS data can be partially explained by that the above wage trends based on the CPS data includes all male workers whereas in the PSID only male family heads are included. However, after including comparable samples in the two datasets, both for male heads, the disparity remains significant as shown in Figure 1. While both datasets show much less progress for blacks in the 1970's than in the 1960's and a slightly setback in the early 1980's, the CPS shows a continued slow progress in the late 1980's whereas the PSID shows a deterioration over the same period. Other studies (Fitzgerald, et al. 1998a; Card and Lemieux, 1994) compare the two datasets by including only male heads also show the similar pattern. My companion paper explores the reason of the difference in the two datasets.

Figure 2 shows the employment trends for both blacks and whites. The employment among black males, 22 to 62, has been declining since the 1960s. In 1970, the employment rate of blacks was 95 percent and very close to 98 percent employment rate of whites. Since then, both blacks and whites experienced a decline in the employment rates, but it was much more pronounced for blacks. By the period of recession in the early 1980s, the employment rate of blacks had fallen to 86 percent compared to 93 percent for whites. After the recession, the employment of both races increased slightly, but the employment of blacks once again fell in the 1990s. In 1995, the employment rate of blacks reached its lowest point of 82 percent, compared to 92 percent for whites. Freeman and Holzer (1986), Welch (1990), Juhn (1992) and Jaynes (1992), Juhn (2002) have documented a very similar trend in relative employment rates of black males, based on the CPS data.

Figure 3 presents the male employment rate by education groups. The decline of employment was the most remarkable among high school dropouts, from a high employment rate of 92 percent in 1970 to 81 percent in 1996, while the employment rates of college graduates remained above 95 percent.

Figure 4 shows the employment rate of high school dropouts for both blacks and whites. The employment rate of low educated blacks decreased from 91 percent in 1970 to 68 percent in 1993, its lowest point in the panel. Then their employment started to increase, and reached 78 percent in 1999. Compared to the CPS data, the decline in employment rates of low educated blacks in the PSID is much slower. Juhn (2000) reports employment rates of 93 percent in 1970 and 60 percent in 1993 based on the sample of working age males. When I restricted the CPS sample to only the male heads, the results are very similar as in the PSID.

5.2 Selection on fixed effects

The merit of using longitudinal data is that it allows controlling for both measured differences in characteristics between workers and nonworkers, such as education and age, and also unobserved time-invariant characteristics that may differ between the two groups. To measure the difference in time-invariant effects, I implement my analysis in three steps. First, I compare the average historical wages (all other periods in the panel, including past and future periods) of workers and nonworkers throughout the panel. In a given year, the average historical (including past and future) wage of an individual, no matter whether the person works in the year or not, is calculated as the mean of wages in all other years when he/she is working. Table 1 shows the mean of the individual historical wages in logarithm. Among workers, it is possible to compare their wages in a given year with the historical wages as in Column (1). The mean current wage is 2.67 in log value for whites and 2.33 for blacks. Their average historical wages are very close to current wage values, 2.68 for whites and 2.34 for blacks. Column (2) shows only the historical wages of nonworkers, as they don't have wages in the year when they are not working. Column (3) compares the historical wages between workers and nonworkers. The wage of an average white worker is 32 percent higher than a nonworker's wage in years when he/she is working; whereas wage of a black worker is 30 percent higher. This implies that those who work more often earn more. However, it is important to be kept in mind that these differences can not all be interpreted as differences in unobserved fixed effects, since part of the difference is the average over time difference on the measured characteristics between workers and nonworkers.

To separate these two factors, my second step is to implement Wooldridge (1995) model as discussed in section 3 and to estimate the correlation between fixed effects that determine individuals' propensity to work and that determine his/her wages, $(cov(\alpha_{2i}, \alpha_{1i}))$. Fixed effects determining individuals' propensity to work are obtained by a probit estimation of the participation equation (7), with a linear projection of a vector of personal characteristics (x_{1i}) that captures fixed effects. Fixed effects determining individual's wages are predicted from fixed effect regression of the wage equation (9), with the value of Mills ratio ($\widehat{\lambda}_{it}$) predicted from the probit estimation. The correlation between the two fixed effects are positive and significant, with .140 for whites and .005 for blacks. Hence, those who have higher wages tend to have, average over time, higher propensity to work, conditioning on observables. Moreover, the predicted fixed effects of nonworkers is 35 percent lower for whites, and 41 percent lower for blacks.

The third step is to investigate the differences in fixed effects between workers and nonworkers by different reasons of not working. The differences are obtained by regressing the predicted fixed effects from wage equation (9) on a set of dummy variables indicating the reasons of unemployment, such as being looking for a job, being disabled, retired, housing keeping, being students and self-employed; the omitted group are workers.[9] Table 2 shows that the disabled group has the lowest average potential wages after controlling for observables: a white disabled earns 36 percent less, and a black earns 28 percent less. The retired and the students have relatively higher fixed effects. Conditional on observables, a white student would have earned 13 percent less than a white worker, compared to 20 percent less for the black.

To sum, the fixed effect can be a key factor in determining wages. Since the fixed effects

[9]This requires an assumption that the frequency of being unemployed for a particular reason does not affect the difference in fixed effects. Moreover, fixed effect can be only obtained for those who ever worked in at least 2 periods.

of nonworkers are much lower than workers, cross-sectional estimates controlling only for observables therefore overstate wages of nonworkers.

5.3 Selection on transitory factors

Table 3 shows the estimated coefficients for the Wooldridge (1995) model. The coefficients on age and education are consistent with other estimates of wage equation. The coefficient on λ is significantly negative, -.19 for whites, and -.52 for blacks.[10] This is consistent with findings of Hoffman and Link (1984) and Baldwin and Johnson (1992), which apply the Heckman (1979) selection model to the CPS and SIPP (the Survey of Income and Program Participation) respectively. These results, however, are counter to conventional wisdom which presumes that transitory increases in wages are associated with increases in propensity to work, and as a result, the correlation of two transitory factors in the wage and participation equations should be positive ($\rho > 0$).

I now trace back to how people decide to work and investigate what is being implied by the negative correlation ($\rho < 0$). Individuals' propensity to work are determined by two factors, offered wages and reservation wages, as indicated in equation (2). Low propensity to work can be due to either low wage offers or high reservation wages. Mathematically, negative selection implies that reservation wage is highly and positively correlated with offered wage.[11] Hence, those who have transitory increases in offered wages also have increases in reservation

[10] The negative selection appears no matter the self-employed are included in the wage sample or not.

[11] $\rho < 0 \Longrightarrow cov(\varepsilon^o_{it} - \varepsilon^r_{it}, \varepsilon^o_{it}) < 0 \Longrightarrow var(\varepsilon^o_{it}) - cov(\varepsilon^o_{it}, \varepsilon^r_{it}) < 0$. Thus, $\rho < 0 \Longrightarrow var(\varepsilon^o_{it})$ is small or $cov(\varepsilon^r_{it}, \varepsilon^o_{it})$ is large.

wages. Furthermore, negative selection is consistent with the statement that, among those who work, persons with higher-employment-probability earn more, given that other measured characteristics are the same.[12]

Therefore, transitory changes in wage rates can be negatively correlated with changes in propensity to work. People who choose not to work may get higher wage offers than those who choose to work, with the same measured characteristics and individual effects, because they have even higher reservation wages.

5.4 Magnitude of selection bias in the racial wage gap

Figure 5 reports the hourly wage trends before and after accounting for nonworkers, for blacks and whites respectively. The average wages of white workers remain almost unchanged over the 1970-2000 period, but with a slight increase of an average rate of .5 percent after correcting selectivity bias. This is caused by the fact that the negative selection on idiosyncratic errors offset positive selection on observables and fixed effects. The average wage of blacks declines after correction, with a magnitude of 2.4 percent average over the years, since a large proportion of blacks were unemployed and most of them were the least educated. In the late 1970s when the employment rate of blacks fell sharply, the decline in black wages

[12] For two workers, a PH.D graduate (P) and a high school dropout (H), the propensity to work of the former is higher $((x_{1it}\widehat{\beta}_1)_P > (x_{1it}\widehat{\beta}_1)_H)$. Under normality assumption, as long as both of them are positive (most individuals are), $(-x_{1it}\widehat{\beta}_1)_P < (-x_{1it}\widehat{\beta}_1)_H$. Hence, $\phi(-x_{1it}\widehat{\beta}_1)_P < \phi(x_{1it}\widehat{\beta}_1)_H$ and $[1-\Phi(x_{1it}\widehat{\beta}_1)]_P > [1-\Phi(x_{1it}\widehat{\beta}_1)]_H$.So $\lambda_P(\frac{(-x_{1it}\widehat{\beta}_1)_P}{[1-\Phi(x_{1it}\widehat{\beta}_1)]_P}) < \lambda_H(\frac{\phi(x_{1it}\widehat{\beta}_1)_H}{[1-\Phi(x_{1it}\widehat{\beta}_1)]_H})$. Therefore the negative coefficient on λ $(cov(\lambda, W) < 0)$, implies $W_P > W_H$.

after including nonworkers was more pronounced.

Table 4 presents how correcting for selectivity bias affects the trend of average black-white wage gaps and calculates the impact of nonrandom selection. Column (1) shows the observed wage differentials between blacks and whites. Column (2) reports the wage differences after including wages of nonworkers predicted from Wooldridge (1995) model.[13] Panel A shows the log wage differences in year 1970, 1980, 1990 and 2000. All the wage differentials are moving averages of three years centered in the indicated year. Panel B calculates the percentage point changes of black-white wage gap in each decades and over the whole period of 1970 to 2000. The effect of selection is computed in Panel C.[14]

In the decade of 1970's, the black-white wage gaps in the PSID dropped from .306 in 1970 to .286 in 1980, suggesting a slow convergence of 2 percentage points. After taking nonworkers into account, the convergence became much slower, decreasing from 2 to .7 percentage points. This indicates that 64 percent of the racial wage convergence in the 1970s was due to non-random selection in the labor market. Between 1980 and 1990, the convergence did not continue and the black-white wages actually diverged by 14.3 percentage points (.43-.286). Correcting for selection increases the divergence from 14.3 to 16.6 percentage points, implying that the loss of blacks in the 1980s would increase by 16 percent if selection bias

[13] Wages of nonworkers are calculated as in equation (1.10).

[14] The effect of selection is calculated as follows: Given the change of black-white wage difference between year t_1 and t_2 is dd_{12}^0, the change after correcting selectivity bias is $dd_{12}^{\prime 1}$, the effect of selection is $(\pm)\left|(dd_{12}^{\prime 1} - dd_{12}^0)/dd_{12}^1\right|$ .Selection is positive when it decreases the black-white wage convergence or increases the divergence, and negative otherwise.

was corrected. In the 1990's, selection played a "negative" role in assessing the *changes* of racial wage gaps. The observed wage differences was .430 in 1990 and .346 in 2000. After accounting nonworkers, the wage differences increased to .480 in 1990 and .377 in 2000. As a result, selection impact actually decreased, from 12 percent in 1990 (.480-.430/.430) to 10 percent in 2000 (.377-.346/.346). The black-white wage convergence from 1990 to 2000 *increases* from 8.4 to 10.3 percentage points after accounting for nonworkers. In contrast to the results in the previous two decades, selection bias actually *understated* the gains of blacks. This is not surprising. While in the pre-1990 decades, selection out of the least-able black in the labor market usually became more and more severe as the decade ended; in the 1990's, this selection effect actually declined. As shown in Figure 4, the employment rate of black high school dropouts showed a continuos decline from 91 percent in 1970 to 68 percent in 1993. After that, the decline of their employment did not continue and started to rise to a value of 78 percent in 1998 (except for some small setbacks in 1995, 1996 and 2000).

Over the whole period of 1970 to 2000, the black-white wages showed a small divergence of 4 percentage points, by increasing from .306 in 1970 to .346 in 2000. This divergence was due to the large drop of the relative black wages in the 1980s. When wages of nonworkers are taken into account, the racial wage gap shows a large divergence of 5.5 percentage points. In other words, by ignoring nonworkers the racial wage divergence between 1970 and 2000 may be understated by as much as 39 percent.

Table 5 presents the changes in average black-white wage gaps and the impacts of selection bias disaggregated by education groups. Among the high school dropouts who have experienced the most severe declines in employment, the selection effect was most pronounced during 1970-2000. Over the whole period, the observed log wage differential increased about 11 percentage points. Once the whole sample including non-workers are taken into account, the racial wage diverged by 20 percentage points. This suggests that roughly half of the underline divergence was hidden by selection bias.

For comparison, Table 5 also shows results for high school graduates and college graduates. The magnitude of selection bias was not large for either of these groups over the whole period from 1970 to 2000. Among high school graduates the rate of racial wage divergence increased by 18 percent, from 12.5 percentage points to 14.8 percentage points. For college graduates, selection played a similar role.

6 Comparison with Cross-sectional Methods

In the following section, I compare the results from panel analysis with cross-sectional methods, by applying matching technique and Brown (1984) method on the same PSID sample. Matching is a commonly used method in correcting selection bias in black-white wage gaps (Blau and Beller, 1992; Chandra, 2000; Juhn, 2002). The studies differs by using different sets of characteristics for matching. Here, I use the method in Chandra (2000) by dividing the data to three education groups (high school dropouts, high school graduates and college graduates) and eight age groups (the whole sample age range is 22-62, so each five years belong to a group). Then I assign the cell-mean value of the observed wages in each (age $\times$ education) cell to all nonworkers in that cell. Brown (1984) assumes that wages of nonworkers are below the median of wage distribution of the whole sample. Therefore assigning any values below the median to the wages of nonworkers does not affect assessing median of the underlying distribution. Here, I assign the lowest wage among workers to all nonworkers.[15]

The first three columns in Table 6 shows the average log racial wage differentials before and after accounting for nonworkers. Compared to the estimates by panel analysis, the results from matching show a smaller impact of selection bias in general. Over the whole period from 1970 to 2000, the observed racial wages diverged by 4 percentage points. Correcting for selection bias by Wooldridge (1995) panel method increases the racial wages divergence to 5.5 percentage points; whereas matching method increase the divergence to 5 percentage points. In other words, using matching method increases the black-white wage divergence by 26 percent whereas using panel method increases by 39 percent. This is because the matching only controls for observables and does not take into account the fact that there might be large differences in the unobserved personal fixed effects between workers and nonworkers.

The last three columns in Table 6 presents the median log racial wage differentials before and after accounting for nonworkers. The estimates of selection bias using Brown's method are higher than the panel estimates. This is consistent with the statement that Brown (1984) provide an upper bound for estimates of selection bias because the assumption that all nonworkers earn less than the median agent is quite strong.[16] Over the whole period from 1970 to 2000, the median of black-white wage gaps among workers increased by 26.5

[15] Neal and Johnson (1996) assign a zero wage rate to all nonworkers when implementing median regressions. Here, since I use log wage rates, I cannot assign a zero wage rate.

[16] Chandra (2000).

percentage points. By including wages of nonworkers predicted from panel estimation, as shown in Column (5), the racial wage divergence over the same period rose to 33 percentage points, suggesting an estimate of selection bias at about 26 percent. By the method of Brown (1984) shown in Column (6), the divergence increased to 37.9 percentage points, leaving an estimate of the bias at about 43 percent.

7 Conclusion

This paper examines the impact of non-random selection in labor market on assessing black-white wage gaps by applying a longitudinal method of imputing wages for nonworkers. These are a few important findings. First, fixed effects are crucial in determining the wages of nonworkers. The fixed effects of white nonworkers are 35 percent lower than for the workers with the same measured characteristics, and 41 percent less for blacks. Controlling only on observables therefore understate the impact of selection. Second, selection on transitory factors, contrary to the conventional wisdom, can be negative. Nonworkers may get higher wages than the workers with the same observed characteristics and individual effects. They choose not to work because they have even higher reservation wages. Third, using the PSID data, I find that the selective bias can explain 40 percent of the observed change in the black-white wage gap over 1970-2000 and 64 percent of the convergence in the 1970s. In contrast to the 1970s and 1990s, the black and white wages diverged in the 1980s. The divergence increased by 16 percent after accounting for nonworkers. Fourth, the selection effect in the 1990s is negative: The selection out of the least-able group in the labor market became less severe in the late 1990s as the employment rate of high school dropouts increased.

References

[1] Baldwin, M. and Johnson, W. G. (1992). Estimating the employment effects of wage discrimination. *Review of Economics and Statistics,* 74(3): 446-55.

[2] Blau, F. D. and Beller, A.H. (1992). Black-white earnings over the 1970s and 1980s: gender differences in trends. *Review of Economics and Statistics,* 74(2): 276-86.

[3] Brown, C. (1984). Black-white earnings ratios Since the civil rights act of 1964: the importance of labor market dropouts. *The Quarterly Journal of Economics,* 99: 31-44.

[4] Butler, R. and Heckman, J. (1977). The government's impact on the labor market status of black Americans: a critical review. In L. J. Hausman et al (Ed.). *Equal Rights and Industrial Relations.* Madison, WI: Industrial Relations Research Association.

[5] Card, D. and Lemieux, T. (1994). Changing wage structure and black-white wage differentials among men and women: a longitudinal analysis. *NBER working paper* No. 4755.

[6] Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47: 225-38.

[7] Chandra, A. (2000). Labor market dropouts and the racial wage gap: 1940-1990. *American Economic Review,* 90(2): 333-38.

[8] Chandra, A. (2003). Is the convergence in the racial wage gap illusory?. *NBER working paper* No. 9476.

[9] Coach, K. and Daily, M. (2000). Black-white wage inequality in the 1990s: a decade of progress. Working papers in applied economic theory: 2000-07. Federal Reserve Bank of San Francisco.

[10] Freeman, R. B. and Holzer, H. (1986). *The black youth employment crisis.* Chicago: University of Chicago Press.

[11] Jaynes, G. D. (1992). The labor market status of black Americans: 1939-85. *Journal of Economic Perspectives*, 4(1): 9-24.

[12] Juhn, C. (1992). Labor market dropouts, selection bias, and trends in black and white wages. *Industrial and Labor Relations Review* (forthcoming).

[13] Juhn, C. (2002). Labor market dropouts, selection bias, and trends in black and white wages. *Industrial and Labor Relations Review* (forthcoming).

[14] Juhn, C., Murphy, K.M. and Pierce,B. (1991). Accounting for the slow-down in black-white wage convergence. In *Workers and Their Wages: Changing Patterns in the United States*. Washington, DC: American Enterprise Institute Press.

[15] Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47: 153-62.

[16] Heckman, J., Lyons, T. and Todd, P. (2000). Understanding black-white wage differentials 1960-90. *American Economic Review*, 90(2): 344-49.

[17] Hoffman, S. D. and Link, C. R. (1984). Selectivity bias in male wage equations: Black-white comparisons. *The Review of Economics and Statistics*, 66(2): 320-24.

[18] Kyriazidou, E. (1997a). Estimation of a panel data sample selection model. *Econometrica,* 65(6): 1335-64.

[19] Lewbel, A. (2002). Endogenous selection or treatment model estimation. *Working paper.*

[20] Mundlak, Y. (1978). On the pooling of time series and cross sectional data. *Econometrica,* 46(1): 69-85.

[21] Neal, D. and Johnson, W. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869-95.

[22] Smith, J. P. (1993). Affirmative action and the racial wage gap. In AEA papers and proceedings: 79-84.

[23] Smith, J. P. and Welch, F. (1989). Black economic progress after myrdal. *Journal of Economic Literature*, 27(2): 519-64.

[24] Verbeek, M. and Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 33: 681–703.

[25] Vroman, W. (1986). *The relative earnings of black men: an analysis of the sample selection hypothesis*. Washington DC: The Urban Institute.

[26] Welch, F. (1990). The employment of black men. Journal of Labor Economics 8(1:2): S26-74.

[27] Wooldridge, J. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68:115–32.