DISCUSSION PAPER SERIES

IZA DP No. 1804

"Soft" Skills, "Hard" Skills, and the Black/White Earnings Gap

C. Simon Fan
Xiangdong Wei
Junsen Zhang

October 2005

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

“Soft” Skills, “Hard” Skills, and the Black/White Earnings Gap

C. Simon Fan
Lingnan University, Hong Kong

Xiangdong Wei
Lingnan University, Hong Kong

Junsen Zhang
Chinese University of Hong Kong and IZA Bonn

Discussion Paper No. 1804
October 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1804
October 2005

ABSTRACT

"Soft" Skills, "Hard" Skills, and the Black/White Earnings Gap*

This paper provides both a theoretical and an empirical investigation into the impact of job skill types on the black/white pay differentials. The theoretical analysis derives that the more intensively "soft"/"hard" skills are used in an occupation, the greater/smaller the black/white pay differential is there in that occupation. Moreover, in response to the differential pay gaps across jobs requiring different levels of "soft"/"hard" skills, blacks are more likely to self-select themselves into the jobs that use "hard" skills more intensively, *ceteris paribus*. Using NLSY data, we find consistent empirical evidence to our theoretical predictions. Hence, the paper bridges the existing literature on racial pay gaps and cognitive vs. non-cognitive skills by explicitly testing the impact of job skill types on racial pay gaps.

JEL Classification: J24, J31, J71

Keywords: soft skills, hard skills, discrimination, pay differentials

Corresponding author:

Junsen Zhang
Department of Economics
Chinese University of Hong Kong
Shatin, NT
Hong Kong
Email: jszhang@cuhk.edu.hk

* The authors wish to thank Jeff Grogger, Dan Hamermesh, James Heckman, Costas Meghir and other participants of the Pacific Rim Conference jointly organized by the Hong Kong Economic Association and the Western Economic Association International, and the 2005 Econometric Society World Congress in London for valuable comments.

1 Introduction

The recent development in the economics of discrimination points out three new directions of research. First, discrimination against blacks varies across occupations in which there are different degrees of direct contact with white customers (e.g. Holzer and Ihlanfeldt, 1998). Second, pre-market factors, as reflected in the disparity of test scores between blacks and whites, are important in explaining racial inequality (e.g. Neal and Johnson, 1996; Carneiro, Heckman and Masterov, 2003). Third, both non-cognitive (or "soft") and cognitive (or "hard") skills are important components of human capital,[1] and the stereotype of being disadvantaged in the society, namely the feeling of being discriminated, reduces the (pre-market) non-cognitive skills for the individuals from disadvantaged groups (e.g. Persico, Postlewaite and Silverman, 2004).[2] Furthermore, some related sociologists' research indicates that black men's lack of "soft" skills is an important reason for their low earnings.[3]

Based on this literature, the current paper examines the impacts of job skill types (cognitive or "hard" skills and non-cognitive or "soft" skills) on the black/white pay differentials and occupational choices. Our theoretical analysis derives two main hypotheses. First, the more intensively "soft" skills are used in an occupation, the greater the racial income gap is there in that occupation. In other words, the racial income gap of an occupation depends on its relative requirement of "soft" skills versus "hard" skills. Second, in response to differential discrimination across occupations and the negative impacts of the stereotype of being disadvantaged on the formation of "soft" skills, *ceteris paribus*, blacks are more likely to self-select themselves into the jobs that use "hard" skills more intensively.

---

[1] For example, see Cameron and Heckman (1993), Green, Machin and Wilkinson (1998), and Heckman and Rubinstein (2001).

[2] For example, Persico, Postlewaite and Silverman (2004) show that "height discrimination" reduces short adolescents' participation in social activities associated with the accumulation of interpersonal skills and other non-cognitive skills and attributes. Similarly, in the literature of psychology, it is shown that discrimination also have negative impacts on black children and adolescents' personal development (e.g. Aronson, Wilson, and Akert, 2001).

[3] For example, based on their face-to-face interviews of managers at 56 firms in the US, Moss and Tilly (1996, p.260) conclude: "*The emphasis employers place on soft skills disadvantages Black male job applicants. This is because many employers see Black men as lacking in precisely the skills they consider increasingly important. Indeed, in our sample the employers placing the greatest emphasis on soft skills are those most*

Generally speaking, occupations that use "soft" skills more intensively are also those with more necessary social interactions. Thus, the first hypothesis can be regarded as an extension and generalization of Holzer and Ihlanfeldt (1998), who investigates the effects of customer discrimination and the degree of direct customer contact on blacks' earnings in the United States. Moreover, the greater racial wage gap in the occupations in which "soft" skills are more intensively used results from two possible sources. First, there may be differential "taste discrimination" across occupations. Second and perhaps more importantly, ethnic minorities, such as blacks, may face comparative disadvantage in accumulating "soft" skills. Carneiro, Heckman and Masterov (2003) have shown that family background is responsible for most of the non-cognitive skill gaps between blacks and whites in the US. The recent study by Persico, Postlewaite and Silverman (2004) and some sociologists point out another possibility for the skill gaps, i.e. the stereotype of being disadvantaged in the society may reduce the accumulation of (pre-market) human capital, particularly non-cognitive skills, for the individuals from disadvantaged groups.

Our theoretical hypotheses are tested using the National Longitudinal Survey of Youth (NLSY) data. Our sample is the pooled male white-collar workers from the 1982 to 2000 waves of NLSY. We first classify jobs into "hard" skill vs. "non-hard" skill ones and "soft" skill vs. "non-soft" skill based on the information provided by the Occupational Information Network (O*NET). Then, the regression analysis yields the following main findings. First, black/white pay differentials tend to be smaller for "hard" skill jobs than for "non-hard" skill jobs, and also smaller for "non-soft" skill jobs than for "soft" skill jobs. Second, based on the Heckman selectivity model, we find that the pattern of self selection implied by our theory indeed existed. Furthermore, the estimated results demonstrate that black white-collar workers tend to self-select themselves into "hard" jobs. Thus, the empirical results provide clear support to our theoretical claims.

In what follows, Section 2 provides the theoretical analysis and discusses the hypotheses; Section 3 describes the data used in the empirical analysis and discusses the empirical methodology employed. Section 4 reports and analyzes the empirical findings. Section 5

---

likely to have negative views of black men as workers."

presents further theoretical implications for our main findings. Section 6 offers conclusions.

2. Theoretical Background and Hypotheses

This section discusses the theoretical background with the help of a simple model, from which we will derive the hypotheses for our empirical analysis. Consider an economy in which labor is the only factor of production. A worker possesses two types of skills: hard and soft. A worker belongs to one of the two types: whites and blacks. Moreover, we consider that there are two jobs in the economy: "salespersons" and "technicians". "Soft" skills are more important for "salespersons" and "hard" skills are more important for "technicians". Formally, consider an individual whose endowment of hard skills is *H*. If the individual chooses to be a "salesperson", his wage will be

$$w_s^i \equiv Z_s^i \left(S^i\right)^a H^{1-a} \tag{1}$$

If the individual chooses to be a "technician", his wage will be

$$w_t^i \equiv Z_t^i \left(S^i\right)^c H^{1-c} \tag{2}$$

where the superscript, "*i*" is either "w" (whites) or "*b*" (blacks); "Z_s^i, Z_t^i, a, c" are all positive coefficients; $0 < a, c < 1$. S^i denotes the individual's "soft" skills. To formulate the idea that "soft" skills are more important for "salespersons" and "hard" skills are more important for "technicians", we assume

$$a > c \tag{3}$$

From (1), we know that in logarithm, the racial wage gap for "salespersons" is

$$\ln(\frac{w_s^w}{w_s^b}) \equiv \ln[\frac{Z_s^w \left(S^i\right)^a H^{1-a}}{Z_s^b \left(S^i\right)^a H^{1-a}}] = \ln(\frac{Z_s^w}{Z_s^b}) + a[\ln\left(S^w\right) - \ln\left(S^b\right)] \tag{4}$$

Similarly, from (2), we know that the racial wage gap for "technicians" is

$$\ln(\frac{w_t^w}{w_t^b}) \equiv \ln[\frac{Z_t^w\left(S^i\right)^c H^{1-c}}{Z_t^b\left(S^i\right)^c H^{1-c}}] = \ln(\frac{Z_t^w}{Z_t^b}) + c[\ln\left(S^w\right) - \ln\left(S^b\right)] \tag{5}$$

From (4)-(5), we get

$$\ln(\frac{w_s^w}{w_s^b}) - \ln(\frac{w_t^w}{w_t^b}) = [\ln(\frac{Z_s^w}{Z_s^b}) - \ln(\frac{Z_t^w}{Z_t^b})] + (a-c)[\ln\left(S^w\right) - \ln\left(S^b\right)] \tag{6}$$

We hypothesize that at any given level of cognitive skills, a black worker possesses a lower level of non-cognitive skills than a white worker, namely,

$$S^w > S^b \tag{7}$$

Carneiro, Heckman and Masterov (2003) have shown that based on the NLSY data there is sizable non-cognitive skill gaps between blacks and whites, and most of these gaps can be explained by family background. Furthermore, as demonstrated by Persico, Postlewaite and Silverman (2001), an individual from disadvantaged groups may acquire less "soft" skills due to his lower self-esteem and less participation in "main-stream" social activities in the periods of personal development. Also, Lang (1986) develops "a language theory of discrimination", which is, in fact, similar to Becker's idea and offers similar conclusions. Lang argues that people can only work together if they "speak" the same language; it is costly to learn a second language or culture; whites and non-whites "speak different languages." Clearly, the disparity in "culture and language" strongly affects a black individual's accumulation of non-cognitive skills in a white-dominated culture. Moreover, according to sociologists' research (e.g. Van Deburg, 1992; Corwin, 2000), black students are discouraged by their peers to behave like whites, namely learn the "culture and language" of whites, which is closed related to their soft skills in the labor market of the United States, by the stereotype "acting white".

Furthermore, occupations that use "soft" skills more intensively are generally those with more necessary social interactions. As shown in Moss and Tilly (1996) and Holzer and Ihlanfeldt (1998), a black individual may suffer from discrimination if he has frequent contact with white customers. This kind of discrimination can be explained in the framework of "taste discrimination" of Becker (1957). Thus, if "taste discrimination" exists more in the occupations in which there are more necessary social interactions, then we have[4]

$$\frac{Z_s^w}{Z_s^b} > \frac{Z_t^w}{Z_t^b} \tag{8}$$

From (5), we know that either (7) or (8) or both are satisfied, we will have

$$\frac{w_s^w}{w_s^b} > \frac{w_t^w}{w_t^b} \tag{9}$$

Thus, we have derived the following hypothesis.

(**H1**) **The wage ratio between whites and blacks is larger for "salespersons" than for "technicians".**

Next, we examine individuals' occupational choices. A worker's utility function is assumed to take the following form:

$$u_j^{ik} \equiv \ln(w_j^{ik}) + \varepsilon_j^{ik} \tag{10}$$

where the superscript "*k*" denotes an individual worker, the subscript "*j*" is either "*s*" (salesperson) or "*t*" (technician); ε_j^{ik}, which is a random variable, measures an individual's idiosyncratic taste for working as a "technician" relative to working as a "salesperson".

Clearly, an individual will choose to work as a "salesperson" if and only if

$$u_s^{ik} > u_t^{ik}$$

namely

$$\ln(w_s^{ik}) + \varepsilon_s^{ik} > \ln(w_t^{ik}) + \varepsilon_t^{ik}$$

that is

[4] Social interactions on the job may include the interactions with employers, co-workers, or customers.

$$\begin{aligned}\Delta\varepsilon^{ik} &\equiv \varepsilon_t^{ik} - \varepsilon_s^{ik} < \ln(w_s^{ik}) - \ln(w_t^{ik}) \\ &= \ln(Z_s^i) + a\ln\left(S^i\right) + (1-a)\ln H - [\ln(Z_t^i) + c\ln\left(S^i\right) + (1-c)\ln H] \\ &= \ln(\frac{Z_s^i}{Z_t^i}) + (a-c)\ln\left(S^i\right) - (a-c)\ln H\end{aligned}$$

We assume that the cumulative distribution function of the random variable, $\Delta\varepsilon^{ik}$, is denoted by "*F(.)*" with the support ($-\infty$, $+\infty$) and is independent of "*i*". Then, we know that the probability of an individual's working as a "salesperson" is

$$\begin{aligned}P(\Delta\varepsilon^{ik} < 0) &= P[\ln(\frac{Z_s^i}{Z_t^i}) + (a-c)\ln\left(S^i\right) - (a-c)\ln H < 0] \\ &= F[\ln(\frac{Z_s^i}{Z_t^i}) + (a-c)\ln\left(S^i\right) - (a-c)\ln H]\end{aligned} \qquad (11)$$

From (3), (7), (8), and (11), we know

$$P(\Delta\varepsilon^{wk} < 0) > P(\Delta\varepsilon^{bk} < 0)$$

Thus, we have the following hypothesis:

(**H2**) **Holding other things constant, white individuals are more likely to be "salespersons" and black individuals are more likely to be "technicians".**

This hypothesis means that individuals from disadvantaged groups respond to discrimination and the stereotype associated with discrimination by self-selecting themselves into the occupations in which "soft" skills are less intensively used.

3. Data & Estimation Methods

Following the theoretical model presented in the early sections, we would like to show if there are any empirical evidence to back up our claims. Specifically, we try to revisit the case of black/white pay differentials in the US and to see if the pay differentials are different for "hard" vs. "soft" skill jobs. More importantly, we want to show these differences if they exist must also influence occupational choice for the black people in the

US.

We use the National Longitudinal Survey of Youth (NLSY) as our main data source for the empirical study. The NLSY has detailed information not only on an individual's own job and other characteristics but also on his/her family background. Besides, the NLSY also has an enlarged sub-sample of blacks, which makes it an ideal dataset for the analyses of black/white pay differentials in the US[5].

Our sample is the pooled male white-collar workers from the 1982 to 2000 waves of NLSY. We use the male sample only as it is more reliable to estimate the male wage equation. We use the white-collar sample to test our theory as it makes more sense to distinguish "hard" vs. "soft" skills for white-collar workers. Due to the relative low skill requirement for blue-collar workers, the distinction between "hard" and "soft" skills for this group of workers may be ambiguous. The 1979 to 1981 waves were excluded from our study as the occupational classification system used in these three waves is the 1970 Census occupational classification, whilst the occupational classification system from the 1982 waves onwards is the 1980 Census occupational classification. Since occupational information is essential for our study and the two classification systems are not directly convertible we decide to use the later waves for consistency. This only reduces our sample a little as many individuals had not started to work in the early waves. Following Cawley, Heckman and Vytlacil (1999), we also restrict our sample to those whose hourly wage is no less than 50 cents per hour (in 1990 dollars).

For the purpose of our empirical study, we need to know whether people hold "hard" skill jobs or "soft" skill jobs. So we must first develop ways to classify jobs into "hard" skill vs. "non-hard" skill ones and "soft" skill vs. "non-soft" skill ones. To do so, we use the information provided by the Occupational Information Network (O*NET)[6]. O*NET is a comprehensive database of worker attributes and job characteristics that replaced Dictionary of Occupational Titles (DOT) to be the primary source of occupational

[5] Several influential previous studies on racial discrimination also utilized this dataset, e.g. Herrnstein and Murray (1994), Neal and Johnson (1996) and Cawley, Heckman and Vytlacil (1999). Cameron and Heckman (2001) also point out the advantage that the NLSY has over CPS on information of family background.

[6] Autor, Levy and Murnane (2003) also used DOT, the predecessor of O*NET to drive measures of routine vs.

information in the US. It assigns standardized scores ranging from 0 to 100 to 275 descriptors for each of the 974 occupations currently. These descriptors cover tasks, knowledge, skills, abilities, work activities, work context, education, training and experience, interest, work style, work value and work needs. We concentrate on 10 descriptors in skills and work activities categories to classify our “hard” skill and “soft” skill jobs. For distinguishing “hard” vs. “non-hard” skill jobs, we use descriptors: (1) science; (2) mathematics; (3) technology design in skills category and (4) analyzing data or information in the work activities category. For distinguishing “soft” vs. “non-soft” skill jobs, we use descriptors: (1) coordination; (2) persuasion; (3) negotiation in skills category, and (4) communication with supervisors, peers, or subordinates; (5) communication with persons outside organization; (6) establishing and maintaining interpersonal relationships in the work activities category.

To define our “hard” vs. “non-hard” skill jobs, we first match a Census 1980 Occupation Code to one or a group of Census 2000 Occupational codes, using various crosswalk tables provided by the National Crosswalk Service Center and the Bureau of Census. Then we find the corresponding 2000 Standard Occupational Classification (SOC) equivalent codes for the resulting Census 2000 Occupational codes following the table provided by the Bureau of Census. Next, we search O*NET using the equivalent SOC codes to find the standardized scores for the four descriptors mentioned above. Finally, we run a bivariate k-mean cluster analysis based on the scores for the four descriptors to classify jobs into “hard” skill vs. “non-hard” skill ones. Once we have separate occupations into “hard” skill vs. “non-hard” skill ones, we assign the “hard” skill dummy to all respondents in our NLSY sample based on their occupations. The same procedures are used to derive the “soft” skill job dummy to all the respondents in the sample. Table 1 and 2 summarize the main results of our cluster analyses and the characteristics for the “hard” and “soft” skill jobs[7]. We can see that for both the “hard” and “soft” skill cases, the cluster analyses have done a quite good job to separate two groups with statistically significant differences in all descriptor scores. It should be noted that our cluster analyses do not separate jobs into two

non-routine tasks of a occupation.

[7] Since we define jobs using Census Occupation codes, we use these two terms inter-changeably in this paper.

types only: “hard” vs. “soft” jobs, mainly because we believe that some jobs may belong to both categories, such as lawyers. Table 3 shows some examples of different types of jobs. So as one would largely expect that statisticians belong to the pure “hard” skill job category while sales workers belong to the pure “soft” skill job category. Table 3 also shows the cross-tabulations of workers in our sample on different types of jobs. Here, there appears to be far more white-collar workers who are in pure “soft” skill than in pure “hard” skill jobs, which is consistent with the fact that there are far more whites than blacks in the sample and whites are more likely to work for the “soft” skill jobs other things being equal.

In empirical studies, we use both of these two dummies to distinguish “hard” vs. “soft” skill jobs and to test our theory. We believe that distinguish skills in both angles could increase the robustness of empirical results.

Our main dependent variable is the log hourly wage rate. We convert wages of all years into the 1990 US dollars. The independent variables include the following human capital variables: (1) potential working experience (age minus age finishing the highest grade of education and minus six) and its square term; (2) tenure for the current job (or job number 1); (3) the highest grade completed. They also include three race dummies (Black, Hispanic and Asian), three region dummies (North East, North Central and West) and one regional unemployment rate measure.

We also want to look at if differences in pay-offs for different skills result in self-selection for blacks. So we need variables to influence people’s choices of job skill types but not necessarily pay directly. We use a variable indicating the highest educational grade completed by either the respondent’s father or mother, and three dummies capturing whether a respondent’s father or mother had a white-collar job, whether a respondent’s father or mother had a “hard” or “soft” skill job, and whether a respondent’s father or mother was born in foreign countries. We believe that the first three variables capture an individual’s cost of accumulating “hard” or “soft” skills based on family background. The last one picks up potential impact of cultural or language barriers on accumulation of “soft” skills. To further satisfy the identification conditions of our selectivity model, we follow

Cawley, Heckman and Vytlacil (1999) and argue that regional dummies and regional unemployment rate influence pay but not choices of skills directly.

The full set of our dependent and independent variables and their corresponding means and standard deviations conditional on race and job types are shown in Table 4. As shown in Table 5, the mean white/black pay differentials for different job types already reveal some interesting patterns. First, the average black/white pay differential for "hard" skill jobs is 2.9% whereas the average black/white pay differential for "non-hard" skill jobs is 8.1%. Secondly, the average black/white pay differential for "soft" skill jobs is 8.2% whereas the average black/white pay differential for "non-hard" skill jobs is 4.4%. These figures show a consistent picture that black people tend to face less discrimination on "hard" skill or "non-soft" skill jobs.

As mentioned before, our empirical study aims to look at the impact of job skills on black/white pay differentials. So we need to run the following standard wage equations:

$$\log wage_{ijt} = X_{ijt}\beta_1 + \varepsilon_{ijt} \tag{7}$$

where the sub-indices *i*, *j* and *t* stand for individual, job skill type and year, respectively. ε_{ijt} is a normally distributed error term with zero mean. X_{ijt} is a vector of independent variables that include the variables described above plus a constant and a time indictor, *t*, which captures the year that an observation was made. Our specification of the wage equation follows closely to that of Cawley, Heckman and Vytlacil (1999)[8]. We also follow their foot-step and estimate our wage equation using the Huber/White/Sandwich robust estimator of variance to allow the error terms to be correlated across years for individuals in the panel data.

Next, if the differences of black/white pay differentials for job skill types are observable to individuals there are reasons to believe that blacks will react to such differences and

[8] We only add one more human capital variable, tenure, to their specification.

self-select themselves into a “hard” skill job if it is not too costly for them to acquire such skills. Hence, observed pay differentials may be truncated or censored by such a selection. We need to run a Heckman selectivity model to correct such a bias and also to test if indeed there is self-selection going on for blacks in the labor market. The selectivity model for “hard” vs. “non-hard” skills is presented as below:

$$\log wage^{h}{}_{ijt} = X_{ijt}\beta^{h} + \varepsilon^{h}{}_{ijt}$$

$$\log wage^{nh}{}_{ijt} = X_{ijt}\beta^{nh} + \varepsilon^{nh}{}_{ijt}$$

$$I^{h} = \alpha(\log wage^{h}{}_{ijt} - \log wage^{nh}{}_{ijt}) + Z_{ijt}\phi + \eta_{ijt} \qquad (8)$$

where the superscript *h* and *nh* stand for “hard” skill and “non-hard” skill, respectively. I^h is the dummy variable for “hard” skill job. $wage^{h}{}_{ijt}$ is observed when *I*=1*, and $wage^{nh}{}_{ijt}$ is observed when *I*=0*. $\varepsilon^{h}{}_{ijt} \sim N(0,\sigma^{h}{}_{ij}), \varepsilon^{nh}{}_{ijt} \sim N(0,\sigma^{nh}{}_{ij}), \eta_{ijt} \sim N(0,1)$ and $Cov(\varepsilon^{k}{}_{ijt}, \eta_{ijt}) = \rho^{k}, k = h, nh$. For the selectivity model for “soft” vs. “non-soft” skills, we simply use wages for “soft” vs. “non-soft” skill jobs and the dummy variable for “soft” skill job.

Z_{ijt} contains the four family background variables plus all variables in X_{ijt} except for the regional dummies and the regional unemployment variable. This selectivity model can be estimated simultaneously by the full information maximum-likelihood method. A test for selectivity bias is a test for $\rho^{k} = 0$. Furthermore, if $\rho^{k} < 0$, it means that those who chose *k* type of job has the comparative advantage of doing that job.

4. Empirical Results

Following the estimation approaches outlined in the last section, we first run OLS estimation of wage equations conditional on various job skill types. The results are presented in Table 6.

In Table 6, we presented four wage regression results. They are for the “hard” skill job sample, the “non-hard” skill job sample, the “soft” skill job sample and the “non-soft” skill job sample, respectively. To focus on our main results first, we look at the black/white pay differentials. As predicted by our theory, we find that pay differentials are indeed higher for “non-hard” skill jobs than for “hard” skill jobs and also higher for “soft” skill jobs than for “non-soft” skill jobs. More interestingly, there appears to be no black/white pay differentials on “hard” skill and “non-soft” skill jobs. Furthermore, we find the difference in black/white pay differentials for “hard” skill jobs and “non-hard” skill jobs are statistically significant as shown by the t-test for difference listed in the third row from the bottom[9]. The same also holds for the difference in black/white pay differentials for “soft” vs. “non-soft” skill jobs.

The estimated coefficients for the other human capital and regional dummy variables are largely as expected. Pay is positively related to job tenure and education levels, and concave in potential working experience. People living in North East and West tend to receive higher pay than people living in North Central and South. Pay is negatively associated with regional unemployment rate. Finally, we see no evidence of racial discriminations in pay against Hispanic and Asian people. These results are largely maintained in all the wage equations below.

Next, we turn to our selectivity model. Table 7 contains the results for “hard” skill jobs. We can see that the estimated correlation coefficient, ρ , is negative and statistically significant, indicating that there is indeed a selection process going on and people with comparative advantage in doing the “hard” skill jobs were found on those jobs. More importantly, in our structural model for selectivity, the coefficient on Black is positive and significant. This shows that blacks do tend to self-select themselves to “hard” skill jobs given there is more discrimination against them on “non-hard” skill jobs. This further proves our hypothesis 2.

Table 8 shows the results of selectivity model for “soft” skill jobs. Similar to the findings in

[9] The t-test here is the test for the zero coefficient of the cross product of the “hard” skill job dummy and the Black dummy or the cross product of the “soft” skill job dummy and the black dummy in a fully stacked wage

Table 7, we again find that the estimated correlation coefficient, ρ, is negative and statistically significant. This once again supports the view that people self-select themselves to jobs that they have comparative advantage. Furthermore, we see that the coefficient on Black is now negative and significant, indicating that blacks self-select themselves away from the "soft" skill jobs.

In sum, based on our two ways of classifying "hard" vs. "soft" skill jobs, we always find that skilled black workers face less discrimination on "hard" skill jobs and more discrimination on "soft" skill jobs. Furthermore, such a differential experience of discrimination induces blacks to self-select themselves into "hard" skill jobs. Hence, we find overwhelming support to both hypotheses derived from our theoretical analysis.

5. Conclusion

The recent development in the research on discrimination has led to a semi-consensus that racial gap in skills is the main source of racial gap in earnings. Meanwhile, there is a growing literature emphasizing that both "soft" and "hard" skills determine earnings. This paper provides a theoretical and empirical investigation by combining these two strands of literature. It demonstrates that understanding the differential impacts of discrimination on the acquisition of cognitive and non-cognitive skills is crucial for the better understanding of the economic impacts of racial discrimination.

Our theoretical analysis derives two main hypotheses. First, the more intensively "soft"/"hard" skills are used in an occupation, the greater/smaller the racial income gap is there in that occupation. In other words, the racial income gap of an occupation depends on its relative requirement of soft skills versus hard skills. Second, in response to differential pay gaps across jobs with different skill requirement, blacks are more likely to self-select themselves into the jobs that require more "hard" skills.

---

regression by the "hard" skill job dummy or the "soft" skill job dummy.

The main innovation of our empirical study lies in developing two measures to distinguish "hard" vs. "non-hard" skill jobs and "soft" vs. "non-soft" skill jobs, using the information provided by the Occupational Information Network (O*NET). After identifying the skill types of individuals' jobs, the regression analyses based on the National Longitudinal Survey of Youth data revealed the following interesting findings. First, the black/white pay differentials tend to be smaller for "hard" skill jobs than for "non-hard" skill jobs, and also smaller for "non-soft" skill jobs than for "soft" skill jobs. Second, based on the Heckman selectivity model, we find that the pattern of self selection implied by our theory indeed existed, i.e. black white-collar workers do tend to self-select themselves into "hard" skill jobs or "non-soft" skill jobs. Thus, the empirical results provide clear support to our theoretical claims.

Thus, our paper provides more theoretical analysis and empirical evidence on the sources of discrimination. It also has implications on the anti-discrimination policy. In particular, it implies that the government policy should be tuned towards those jobs that require more intensive use of "soft" skills and targeted more on the elimination of the potential disadvantages that blacks may face in accumulating "soft" skills.

References

Aronson, Elliot, Timothy D. Wilson, and Robin M. Akert (2001), Social Psychology, Upper Saddle River, NJ: Prentice Hall.

Autor, David H., Frank Levy and Richard J. Murnane (2003), “The Skill Content of Recent Technological Change: An Empirical Exploration,” *Quarterly Journal of Economics*, 118(4): 1279-1333.

Becker, Gary (1957) *The Economics of Discrimination*, Chicago: University of Chicago Press.

Biddle, Jeff E. and Daniel S. Hamermesh (1998) “Beauty, Productivity and Discrimination: Lawyers' Looks and Lucre,” *Journal of Labor Economics*, 16(1): 172-201.

Cameron, Stephen V. and James J. Heckman (1993), “The Nonequivalence of High School Equivalents,” *Journal of Labor Economics*, 11(1): 1-47.

Cameron, Stephen V. and James J. Heckman (2001) “The Dynamics of Educational Attainment for Black, Hispanic, and White Males,” *Journal of Political Economy*, 109(3): 455-99.

Cawley, John, James Heckman and Edward Vytlacil (1999) “Meritocracy in America: Wages within and across Occupations,” *Industrial Relations*, 38(3): 250-96.

Carneiro, Pedro, James J. Heckman and Dimitry V. Masterov (2003) “Labor Market Discrimination and Racial Differences in Pre-Market Factors,” NBER Working Paper No. w10068.

Chiswick, Barry (1988) “Differences in Education and Earnings Across Racial and Ethnic Groups: Tastes, Discrimination, and Investments in Child Quality,” *Quarterly Journal of Economics*, 103(3): 571-597.

Coate, Stephen and Glenn C. Loury (1993) “Will Affirmative Action Eliminate Negative Stereotypes?” *American Economic Review*, 83(5): 792-810.

Corwin, Miles (2000), *And Still We Rise: the Trials and Triumphs of Twelve Gifted Inner-city High School Students*, New York: William Morrow.

Green, Francis, Stephen Machin, and David Wilkinson (1998), “The Meaning and Determinants of Skills Shortages,” *Oxford Bulletin of Economics and Statistics*, 60(2): 165-87.

Heckman, James J. and Yona Rubinstein (2001), “The Importance of Noncognitive Skills: Lessons from the GED Testing Program,” *American Economic Review*, 91(2): 145-149.

Holzer, Harry J. and Keith R. Ihlanfeldt (1998) “Customer Discrimination and Employment Outcomes for Minority Workers,” *Quarterly Journal of Economics*, 113(3): 835-867.

Lang, Kevin (1986) “A Language Theory of Discrimination,” *Quarterly Journal of Economics*, 101(2): 363-382.

Lundberg, Shelly and Richard Startz (1983), “Private Discrimination and Social Intervention in Competitive Labor markets,” *American Economic Review*, 73(3): 340-347.

Lundberg, Shelly and Richard Startz (1998), “On the Persistence of Racial Inequality,” *Journal of Labor Economics*, 16(2): 292-323.

Moss, Philip and Chris Tilly (1996), ““Soft” Skills and Race: An Investigation of Black Men’s Employment Problems,” *Work and Occupations*, 23(3): 252-276.

Neal, Derek A. and William R. Johnson (1996), "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104(5): 869-95.

Neal, Derek A. (2004), "Why Has Black-White Skill Convergence Stopped?" memio, University of Chicago.

Persico, Nicola, Andrew Postlewaite and Dan Silverman (2004) "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height," *Journal of Political Economy*, 112(5): 1019-53.

Sawhill, Isabel (1988), "Poverty in the U.S.: Why Is It So Persistent?" *Journal of Economic Literature*, 26(3): 1073-1119.

Van Deburg, William L. (1992), *New Day in Babylon: the Black Power Movement and American Culture*, *1965-1975*, Chicago: University of Chicago Press.

Table 1. Bivariate K-mean Cluster Analysis to Derive the “Hard” Skill Job Dummy

	“Hard” Skill Job	**“Non-Hard” Skill Job**	**Col. 2/Col. 3**
Science	66.196	12.588	5.259***
Mathematics	68.072	49.777	1.368***
Technology Design	33.072	12.821	2.580***
Analyzing Data or Information	67.232	51.055	1.319***
Percent of Occupations	39.0%	61.0%	

The numbers should be interpreted as follows: for example, the number in the first cell, 66.196 is the average score of Science for “hard” skill jobs. The number in the last row of the same column, 39.0%, gives the percentage of “hard” skill jobs in white-collar occupations.
***Indicates a significant difference in proportions between the “hard” skill and “non-hard” skill clusters at the 1 percent level.

Table 2. Bivariate K-mean Cluster Analysis to Derive the “Soft” Skill Job Dummy

	“Soft” Skill Job	**“Non-Soft” Skill Job**	**Col. 2/Col. 3**
Coordination	55.775	32.002	1.743***
Persuasion	39.418	14.370	2.743***
Negotiation	33.419	12.880	2.595***
Communication with Supervisors, Peers, or Subordinates	68.380	48.696	1.404***
Communication with Persons Outside Organization	64.255	41.506	1.548***
Establishing and Maintaining Interpersonal Relationships	61.256	34.399	1.781***
Percent of Occupations	69.3%	30.7%	

The numbers should be interpreted as follows: for example, the number in the first cell, 55.775 is the average score of Coordination for “soft” skill jobs. The number in the last row of the same column, 69.3%, gives the percentage of “soft” skill jobs in white-collar occupations.
***Indicates a significant difference in proportions between the “soft” skill and “non-soft” skill clusters at the 1 percent level.

Table 3. Some Examples and Cross-tabulations of “Hard” and “Soft” Skill Job workers

	“soft” skill job		
“hard” skill job	**0**	**1**	**Total**
0	3,484 (athletics, typists, computer library clerks)	13,789 (managers & administrators, school teachers, sales workers)	17,273
1	1,253 (actuaries, statisticians, dentists)	3,053 (architects, physicians, Economists)	4,306
Total	4,737	16,842	21,579

Some examples of each type of jobs are shown in the parentheses.

Table 4. Variable Descriptive Statistics Conditional on Race and Job Type

Variables	Race			
	Black		White	
	“Hard” Skill Jobs	**“Non-hard” Skill Jobs**	**“Hard” Skill Jobs**	**“Non-hard” Skill Jobs**
Log hourly wage	2.511 (.520)	2.155 (.603)	2.586 (.573)	2.344 (.644)
Working experience (years)	8.137 (5.108)	8.895 (5.296)	7.234 (4.914)	8.064 (5.252)
Job tenure (weeks)	162.754 (162.859)	145.422 (167.505)	165.760 (177.532)	180.301 (193.923)
Highest educational grade	14.989 (2.156)	13.659 (2.172)	15.447 (2.457)	14.280 (2.359)
Armed Force Qualification Test	51.309 (28.866)	34.679 (24.963)	78.474 (21.807)	65.672 (24.977)
Regional unemployment rate	2.522 (0.811)	2.700 (.922)	2.728 (.978)	2.787 (1.025)
Regional dummies:				
North East	.111 (.315)	.208 (.406)	.201 (.400)	.225 (.418)
North Central	.174 (.379)	.175 (.380)	.304 (.460)	.295 (.456)
West	.154 (.361)	.118 (.323)	.197 (.397)	.184 (.388)
The highest parental educational grade	12.947 (2.861)	12.120 (2.663)	14.422 (2.966)	13.713 (2.829)
Dummy=1 if father or mother had a white-collar job	.453 (.498)	.357 (.479)	.715 (.451)	.674 (.469)
Dummy=1 if father or mother had a “hard” skill job	.314 (.464)	.283 (.451)	.408 (.492)	.348 (.476)
Dummy=1 if father or mother had a “soft” skill job	.572 (.495)	.436 (.496)	.689 (.463)	.698 (.459)
Dummy=1 if father or mother born in foreign countries	.052 (.222)	.022 (.148)	.095 (.294)	.073 (.260)
Sample size	656	3,519	3,093	10,902

Data source: pooled male white-collar workers in wave 1982-2000 of NLSY

Table 4. Continued

Variables	Race			
	Black		White	
	"Soft" Skill Jobs	"Non-Soft" Skill Jobs	"Soft" Skill Jobs	"Non-soft" Skill Jobs
Log hourly wage	2.236 (.602)	2.147 (.606)	2.436 (.638)	2.246 (.608)
Working experience	8.939 (5.245)	8.338 (5.327)	8.048 (5.135)	7.210 (5.355)
Job tenure	152.091 (169.314)	137.617 (159.762)	180.836 (191.197)	162.240 (187.132)
Highest educational grade	14.048 (2.289)	13.386 (1.953)	14.633 (2.435)	14.157 (2.368)
Armed Force Qualification Test	38.387 (27.047)	34.539 (24.139)	68.534 (24.800)	68.426 (25.216)
Regional unemployment rate	2.649 (.885)	2.731 (.963)	2.746 (1.005)	2.885 (1.047)
Regional dummies:				
North East	.176 (.381)	.237 (.426)	.217 (.412)	.233 (.423)
North Central	.185 (.388)	.147 (.355)	.301 (.459)	.281 (.450)
West	.125 (.331)	.122 (.327)	.186 (.389)	.192 (.394)
The highest parental educational grade	12.336 (2.743)	12.025 (2.618)	13.905 (2.886)	13.730 (2.829)
Dummy=1 if father or mother had a white-collar job	.364 (.481)	.392 (.488)	.684 (.465)	.681 (.466)
Dummy=1 if father or mother had a "hard" skill job	.293 (.455)	.274 (.446)	.356 (.479)	.381 (.486)
Dummy=1 if father or mother had a "soft" skill job	.456 (.498)	.461 (.499)	.704 (.457)	.665 (.472)
Dummy=1 if father or mother born in foreign countries	.029 (.167)	.023 (.150)	.079 (.270)	.072 (.259)
Sample size	3.041	1,134	11,119	2,796

Data source: pooled male white-collar workers in wave 1982-2000 of NLSY

Table 5. Mean White/Black Pay Differentials by Job Types

	"Hard" Skill Jobs (I)	"Non-hard" Skill Jobs (II)	"Soft" Skill Jobs (III)	"Non-soft" Skill Jobs (IV)
Pay Differentials	2.9%	8.1%	8.2%	4.4%
Differences	(I) – (II): 5.2%***		(III)-(IV): -3.8%***	

The percentage mean pay differentials are calculated using the mean pay figures in Table 3.

So for example, 2.9%=(2.586-2.511)*100/2.586.

*** significant at 1% level.

Table 6. OLS Wage Equation

Independent Variables	"Hard" Skill Jobs	"Non-hard" Skill Jobs	"Soft" Skill Jobs	"Non-soft" Skill Jobs
Constant	-9.50 (.83)	-9.681 (1.27)	-7.370 (.97)	-2.303 (.19)
Dummy for Black	.039 (1.08)	-.068*** (3.23)	-.063*** (2.80)	.019 (.59)
Dummy for Hispanic	.064 (1.61)	-.020 (.79)	-.027 (1.08)	.070* (1.71)
Dummy for Asian	-.062 (.61)	.037 (.49)	.015 (.22)	.031 (.16)
Working experience (years)	.083*** (9.36)	.068*** (14.44)	.067*** (13.53)	.079*** (12.13)
Working experience square	-.003*** (9.00)	-.002*** (13.11)	-.002*** (11.62)	-.003*** (11.64)
Job tenure (weeks)	.0004*** (6.03)	.0005*** (13.69)	.0004*** (11.95)	.0007*** (9.77)
Highest educational grade	.068*** (7.66)	.072*** (12.86)	.078*** (13.94)	.058*** (6.15)
Armed Force Qualification Test	.004*** (6.32)	.002*** (6.37)	.003*** (7.99)	.003*** (5.24)
Regional unemployment rate	-.031*** (2.97)	-.059*** (10.33)	-.057*** (9.36)	-.046*** (4.43)
Regional dummies:				
North East	.125*** (3.70)	.130*** (6.19)	.111*** (5.24)	.168*** (5.18)
North Central	.022 (.69)	.006 (.28)	.010 (.46)	-.006 (.19)
West	.079** (2.29)	.100*** (4.52)	.102*** (4.63)	.098*** (2.71)
Year	.005 (.90)	.005 (1.37)	.004 (1.06)	.002 (.26)
Differences in coefficients for Black	.107*** (2.74)		-.083** (2.34)	
Adjusted R^2	.285	.277	.295	.259
Sample size	3,768	14,907	14,519	4,156

The wage equation is estimated using the Huber/White/Sandwich robust estimator of variance to allow the error terms to be correlated across years for individuals in the panel data.
Figures in brackets are t-statistics.
***significant at 1%; **significant at 5%; *significant at 10%.

Table 7. Heckman Selectivity Model for "Hard" Skill Jobs

Independent Variables	Wage Equation	Selectivity Equation
Constant	-10.787 (.89)	7.965 (.37)
Dummy for Black	.009 (.22)	.180** (2.00)
Dummy for Hispanic	.040 (.95)	.032 (.32)
Dummy for Asian	.035 (.27)	-.426 (1.53)
Working experience (years)	.080*** (8.15)	.020 (1.36)
Working experience square	-.003*** (7.80)	-.0008 (1.52)
Job tenure (weeks)	.0004*** (5.61)	-.0003** (2.00)
Highest educational grade	.049*** (4.56)	.061*** (3.55)
Armed Force Qualification Test	.0009 (.79)	.009*** (6.93)
Regional unemployment rate	-.023** (2.02)	
Regional dummies:		
North East	.155*** (4.22)	
North Central	.020 (.56)	
West	.054 (1.45)	
Year	.006 (1.03)	-.005 (.44)
Predicted wage on "hard" skill job – predicted wage on "non-hard" skill job		-1.000* (1.83)
The highest parental educational grade		-.00015 (.02)
Dummy=1 if father or mother had a white-collar job		-.072 (1.32)
Dummy=1 if father or mother had a "hard" skill job		..082 (1.60)
Dummy=1 if father or mother born in foreign countries		.022 (.148)
ρ ($\chi^2(1)$)	-.692*** (8.91)	
Wald $\chi^2(13)$	416.65***	
Pseudo R^2		.064
Sample size	3,727	19,214

Figures in brackets are t-statistics. ***significant at 1%; **significant at 5%; *significant at 10% level.

Table 8. Heckman Selectivity Model for "soft" Skill Jobs

Independent Variables	**Wage Equation**	**Selectivity Equation**
Constant	-5.795*** (7.636)	-47.709*** (2.81)
Dummy for Black	-.056*** (2.62)	-.162** (2.28)
Dummy for Hispanic	-.034 (1.33)	.020 (.22)
Dummy for Asian	.041 (.63)	-.032 (.17)
Working experience (years)	.068*** (13.39)	.052*** (4.20)
Working experience square	-.002*** (11.26)	-.003*** (4.16)
Job tenure (weeks)	.0004*** (11.66)	.0003 (1.43)
Highest educational grade	.077*** (13.51)	.040** (2.33)
Armed Force Qualification Test	.003*** (8.05)	-.001 (1.54)
Regional unemployment rate	-.057** (9.21)	
Regional dummies:		
North East	.113*** (5.25)	
North Central	.007 (.34)	
West	.100*** (4.49)	
Year	.003 (.85)	.024*** (2.78)
Predicted wage on "hard" skill job – predicted wage on "non-hard" skill job		1.075* (1.73)
The highest parental educational grade		.007 (1.00)
Dummy=1 if father or mother had a white-collar job		-.130** (2.51)
Dummy=1 if father or mother had a "soft" skill job		.092* (1.85)
Dummy=1 if father or mother born in foreign countries		.049 (0.80)
ρ ($\chi^2(1)$)	-.090** (4.52)	
Wald $\chi^2(13)$	2701.99***	
Pseudo R^2		.03
Sample size	14,286	19,214

Figures in brackets are t-statistics. ***significant at 1%; **significant at 5%; *significant at 10% level.