Institute for Research on Poverty
Discussion Paper no. 1038-94

Wages, Racial Composition, and Quality Sorting in Labor Markets

Barry T. Hirsch
Department of Economics
Florida State University

David A. Macpherson
Department of Economics
Florida State University

July 1994

The authors acknowledge financial assistance from the Small Grants Program of the Institute for Research on Poverty, University of Wisconsin. Helpful suggestions were received from Maria Cancian, Darren Grant, Robert Hauser, Maurice MacDonald, Edward Schumacher, Eric Solberg, Andrew Weiss, and workshop participants during presentations at the Institute for Research on Poverty and the Jerome Levy Economics Institute.

Abstract

This paper examines the relationship between wage rates and the racial composition of jobs, using large cross-sectional and longitudinal samples constructed from monthly Current Population Surveys for 1983–92. Support is found for a "quality sorting" model that posits an equilibrium in which the racial composition of jobs serves as a skill index of unmeasured labor quality. Estimation of standard wage-level equations shows that wages of both black and white workers are substantially lower in occupations with a high density of blacks. Consistent with the quality sorting hypothesis, the magnitude of the relationship is reduced sharply after accounting for occupational skill measures. Longitudinal wage-change estimates controlling for person-specific quality indicate little if any causal effect of racial composition on wages. Estimates of racial discrimination are reduced only moderately after accounting for racial composition; unexplained differentials occur within occupations or reflect inter-occupational differences uncorrelated with racial composition and occupational skill measures.

Wages, Racial Composition, and Quality Sorting in Labor Markets

I. INTRODUCTION

In contrast to the considerable effort given to the study of racial wage differentials and labor market discrimination, scant attention has been paid to how wages vary with the racial composition of jobs. A recent paper by Hirsch and Schumacher (1992) concludes that wages for both white and black workers are significantly lower in industry-occupation-region groups with a high proportion of black workers.[1] They argue that their results are not easily explained by standard statistical discrimination models, the crowding hypothesis, or taste theories of discrimination. They propose what they refer to as a "quality sorting" explanation, but provide no direct evidence. A quality sorting equilibrium implies that the proportion of black workers in a job is correlated with measured and unmeasured labor productivity differences. Lower-quality white and black workers are sorted into jobs with a relatively large proportion of black workers, while higher quality white and black workers sort into jobs with a small proportion of black workers. Wages vary with the racial density of jobs, independent of the race or measured characteristics of individual workers. Such an equilibrium is likely to have arisen as a result of past and present employment discrimination.

This paper outlines alternative variants of a quality sorting model. The model is evaluated through the estimation of wage-level and wage-change equations examining the relationship between wages and racial composition. We conclude that the racial composition of occupations is an important and typically neglected correlate of wages, but that it serves primarily as an index of otherwise unmeasured occupational and worker skills and not as a causal determinant of wages.

Our study is noteworthy for the size and quality of the assembled data sets. We develop an unusually large longitudinal data base from the January 1983 through December 1992 monthly Outgoing Rotation Group files of the Current Population Surveys (CPS ORG). Additional data sources used in the paper include the 1973–78 May CPS files, the CPS ORG files for 1979–82, March

CPS files for 1983–92, the five January CPS Displaced Worker Surveys (DWS) between 1984 and 1992, the *Dictionary of Occupational Titles*, and various supplements to the CPS (from the *DOT* and the CPS supplements we derive variables pertaining to job characteristics).

Section II of the paper develops alternative variants of the simple quality sorting model and outlines the empirical strategy designed to evaluate it. Section III follows with a description of data sources and a presentation of descriptive evidence on the racial wage gap and occupational segregation during the 1973–92 period. In Section IV, we present empirical results from the estimation of wage-level equations and examine issues of specification, functional form, and differences in racial composition effects across different worker groups. Section V presents evidence from wage-change equations using three alternative panel data sets. A concluding section evaluates the study's findings.

II. QUALITY SORTING, WAGES, AND RACIAL COMPOSITION

A. A Simple Model

In the model of quality sorting that follows, we distinguish between quality that is measured (unmeasured) and quality that is observed (unobserved). Measured quality is that reflected in productivity-related characteristics in the researcher's information set (years of schooling, age, etc.), while unmeasured quality cannot be directly measured. Observed quality refers to that known to employers, whereas unobserved quality is not known to employers at the time of hire, but may become known over time.

Employers and workers are sorted on the basis of observed and unobserved quality. Racial composition, measured by the proportion of black workers in a job (*B*), will be negatively correlated with white and black wage rates under several scenarios. A simple model illustrates the key points. Workers are either high or low quality, and their marginal revenue products are represented by H or L. Let P_b and P_w represent the proportions of black and white workers with productivity H, and $(1\text{-}P_b)$

and $(1-P_w)$ the proportions with productivity L. Assume initially that individual productivities are independent of the place of work and that workers are paid their marginal products. The economy-wide average productivities or wages of black and white workers, V_b and V_w respectively, are

(1) $V_b = P_bH + (1-P_b)L$,

(2) $V_w = P_wH + (1-P_w)L$.

The productivity or wage differential between white and black workers is

(3) $V_w - V_b = (P_w-P_b)(H-L)$.

The sign of the racial wage gap is determined by the sign of (P_w-P_b), and the magnitude of the gap is a function of differences in the proportions of high-quality workers and in productivity between high- and low-quality workers.

The average productivity and, by assumption, the average wage in an integrated workplace is

(4) $V = BV_b + (1-B)V_w$

$= B[P_bH + (1-P_b)L] + (1-B)[P_wH + (1-P_w)L]$,

where B represents the proportion of workers who are black. Following expansion, and the cancellation and rearrangement of terms, we obtain

(5) $V = (P_bB-P_wB)(H-L) + P_w(H-L) + L$.

If the proportion of white and black workers of high and low quality are equal $(P_w=P_b)$, the first term cancels out and average wages in a workplace are independent of B. If $P_w-P_b>0$ then the average wage will decrease with respect to B.

We sketch out four variants of this model. In each case assume that $P_w-P_b>0$. Variants one and two assume that P_b and P_w do not vary with B across jobs, although jobs with a low B will have more H workers owing to $P_w-P_b>0$. We conclude below that variants one and two are too simplistic and do not explain extant evidence. In the first variant, assume that all workers are paid according to their individual productivities and that there is no unmeasured quality. Variant one predicts that average wages across jobs decrease with respect to *B*, absent controls. No relationship between individual wages and *B* (or between individual wages and race) would exist, however, after controlling for measurable worker characteristics.

Variant two assumes that there is unmeasured quality, but no unobserved quality. Employers (but not researchers) know each worker's productivity and pay a wage H or L accordingly. In this case, not only are average wages across jobs negatively related to *B*, but also individual wages are negatively related to *B* after accounting for measurable characteristics other than race. Inclusion of a race dummy (or estimation of separate regressions by race) would eliminate this relationship, however, since information on individual race would capture (on average but with individual error) both (H-L) and differences in the probabilities of being H or L for white and black workers. The finding that wages decrease significantly with respect to *B* for both white and black workers, following inclusion of detailed controls, permits rejection of variants one and two.

Variant three of the model provides what we believe is a more realistic approximation of the labor market. It retains the assumption that individual productivities are observed by employers, but allows P_b and P_w to increase across jobs as *B* decreases. That is, the proportion of black and white workers who are of high quality increases as the proportion of black workers in an occupation decreases. An employer who can observe worker productivity and chooses to hire a high-skill labor force (due, say, to the firm's technology) selects both a relatively lower proportion of black workers than in the average workplace (since P_w-P_b>0), and a higher than average proportion of H quality black and white workers. Both black and white wages will decrease with respect to *B*, even following control for measurable characteristics, since *B* captures unmeasured quality or selective hiring (i.e., the higher proportion of H quality workers) in a labor market.

Variant three of the quality sorting model would account for the negative correlation between wages and B in a wage-level equation (Hirsch and Schumacher, 1992), since B serves as a quality index for white and black workers. But it would not be consistent with such a correlation in a longitudinal framework. If workers are paid based on observed individual productivity, changes in an individual's wage following a job change should be uncorrelated with changes in B. A zero coefficient on ΔB in a wage-change equation would provide evidence consistent with variant three of the quality sorting model, while a negative coefficient on ΔB would allow rejection of variant three of the model.

Variant four of the quality sorting model leads to the prediction of a negative relationship between white and black wages and B both in wage-level and wage-change equations. Rather than assuming that a worker's productivity H or L is fixed across jobs, let individual black and white worker productivity vary such that productivity is higher in jobs with a lower B. That is, worker productivity is determined by the worker *and* the job. In terms of the model, some black and white workers who are L quality in occupations with a high B will be H quality in occupations with a lower B. Worker quality may vary with the job because of optimal sorting on the basis of skill (this conclusion follows strictly only if comparative advantage can be measured by a one-dimensional skill index) or greater effort in jobs with higher average productivity. Note that a similar equilibrium might arise if some portion of (but not all) worker quality remains unobservable to employers, so that worker pay is a function of individual and group productivity. Variant four of the model leads to a prediction of a negative relationship between the wage and B in *both* a wage-level and wage-change equation, for white and black workers. That is, white and black workers switching to occupations with a lower B can expect a wage increase, and vice versa. Estimation of levels and wage-change equations, therefore, can distinguish between alternative variants of our model and permit inferences about the type of quality sorting that is typical in labor markets.[2]

Existing models in the literature can be interpreted as special cases of the quality sorting hypothesis. For example, Lang's (1986) "language" model of discrimination may help explain the relationship between racial composition and wages. In Lang's model, black and white workers display different language and communication patterns, and workers who do not acquire majority traits face a labor market penalty necessary to compensate employers for higher communication costs. The relatively large number of black and small number of white workers not acquiring majority language patterns are crowded into job markets with a high *B* and wage penalties owing to high communication costs between white workers and employers. Blacks who acquire majority language patterns receive a return on these skills by being able to acquire jobs in predominantly white labor markets with lower communication costs. The language model thus illustrates an example of quality sorting, with the majority language pattern representing a quality attribute by which workers are sorted.

An extension of the language model yields an additional insight. Labor markets with a relatively large number of black workers provide greater contact between blacks and whites and a larger accumulated stock of communication capital (for related arguments in a historical context, see Whatley, 1990). In this case, there exists a smaller penalty associated with minority language patterns in labor markets with a relatively large number of minority workers (and employers). This implication is consistent with the finding of smaller penalties associated with the racial density of an occupation among workers with less schooling or who work in the South, and of a nonlinear relationship in which the marginal wage effect of *B* declines with *B*.

This study's emphasis on a quality sorting explanation for the relationship between wages and racial composition does not rule out other explanations. As discussed by Hirsch and Schumacher (1992), however, alternative explanations are not convincing. Discrimination resulting from employer, employee, and consumer preferences may well have led to a quality sorting equilibrium in which unmeasured (and perhaps unobserved) skills vary inversely with the proportion of blacks in an occupation.[3] Preference models of discrimination (Becker, 1971; Arrow, 1973), however, predict either a positive relationship between white wages and *B*, or a widening racial gap as *B* increases. Neither pattern is evident in the data. For example, employer discrimination implies either no wage effects in markets with sufficient numbers of nondiscriminatory employers, or wage penalties for blacks relative to whites in labor markets with large numbers of black workers. Widespread employee discrimination implies a wage premium for white workers who work alongside black workers. And consumer discrimination implies either employment segregation within and across occupations, or wage differentials between white and black workers in similar jobs. The preference models do not help account for systematically lower wage rates for white workers in jobs with relatively large numbers of black workers.

Statistical discrimination models emphasize the role of imperfect information about individual worker productivities; models differ in their implications about the effects of racial composition. The Aigner-Cain model (see Cain, 1986) assumes an equal distribution of abilities among black and white workers, while "indicator" characteristics (e.g., schooling) are less reliable measures of productivity for blacks than for whites. In the Aigner-Cain model, there is no group discrimination. High-ability whites are paid more than high-ability blacks and low-ability whites are paid less than low-ability blacks, but average wages are equivalent. The model does not predict a relationship between average white and black wages and *B* but, rather, that predominantly black workplaces exhibit lower wage *dispersion* and lower returns to indicator variables than do predominantly white workplaces.[4] Likewise, alternative models of statistical discrimination do not appear to be fully consistent with empirical evidence relating black and white wages to racial composition (Hirsch and Schumacher, 1992).

Occupational crowding, by which blacks are relegated to particular occupations and have limited mobility, leads to lower wages in crowded occupations for blacks and whites alike, while leading to higher wages for both races in noncrowded occupations (Bergmann, 1971). Although broadly consistent with the evidence of a negative relationship between *B* and white and black wages, the crowding explanation is far from compelling, in no small part because the size of the black labor force is relatively small. For crowding to occur, there must be a sizable black workforce and employment discrimination must be the prevailing norm (such a description appears to apply to much of the South prior to the mid-1960s). If crowding were the primary mechanism through which *B* is currently correlated with wages, the effects of *B* should be strongest in occupations with a very high proportion of black workers, in the South, and among workers with lower levels of schooling. If anything, evidence indicates precisely the opposite—weaker effects at lower schooling levels and in the South, and a nonlinear relationship between log wages and *B* whereby wages decrease with respect to *B* at a *decreasing* rate.

B. Wages, Racial Composition, and Quality Sorting: Empirical Specification

The quality sorting hypothesis is examined through the estimation of wage-level and wage-change equations. If racial composition serves as an occupational skill index, then the absolute value of the coefficient on *B* should decrease as measurable skill-related variables are introduced into the wage equation. These variables will measure skill directly and indirectly, since some skills not explicitly measured are likely to be positively correlated with those dimensions of skill that are measured. We will then estimate wage-change equations, in which unmeasured person-specific differences in productivity fixed over time are netted out. If variant three of the quality sorting hypothesis is correct, then the impact of a change in *B* on the log wage change should be substantially smaller than the estimated effect in a levels equation, since the change model controls for unmeasured person-specific skills correlated with *B*.

More concretely, the relationship between wages and racial composition is estimated first in levels by

(6) $\ln W_{itb} = \Sigma\beta_{kb}X_{iktb} + \Theta_b \ln B_{itb} + \Phi_{ib} + \epsilon_{itb}$,

(7) $\ln W_{itw} = \Sigma\beta_{kw}X_{iktw} + \Theta_w \ln B_{itw} + \Phi_{iw} + \epsilon_{itw}$,

where subscripts b and w designate black and white; $\ln W_{it}$ is the natural log of hourly earnings for individual i in year t; X_k consists of X_1=1 and k-1 variables measuring personal and/or job characteristics and region; β_k includes a constant and k-1 coefficients corresponding to variables in X; ln*B* is the natural log of the proportion of black to total employment in the worker's detailed occupation (alternative measures of *B* are discussed subsequently); Θ is the coefficient on ln*B* representing an elasticity of wages with respect to *B*; Φ_i is an unmeasured person-specific fixed effect invariant over time (a one-year period in most of our models); and ϵ is an error term with zero mean and constant variance. A value of $\Theta<0$ implies that wages decrease with respect to *B*, ceteris paribus, while $\Theta>0$ implies the opposite. The quality sorting hypothesis implies that estimates of Θ will become less negative as measures of occupational skill are introduced explicitly into the wage equation.

Estimation of the wage equation in levels will not account directly for unmeasured worker-specific quality differences. If the omitted fixed effect Φ_i is negatively correlated with ln*B*, as the quality sorting model implies, then levels estimates of Θ_w and Θ_b in (6) and (7) will be downward biased (away from zero). Omitted fixed effects associated with quality sorting may be accounted for by estimating wage-change equations. Letting Δ represent changes between adjacent years [t-(t-1)], [(t-1)-(t-2)], etc., the following longitudinal equations are obtained:

(8) $\Delta\ln W_{iyb} = \Sigma\beta_{kb}\Delta X_{ikyb} + \Theta_b\Delta\ln B_{iyb} + \Delta\epsilon'_{iyb}$,

(9) $\Delta\ln W_{iyw} = \Sigma\beta_{kw}\Delta X_{ikyw} + \Theta_w\Delta\ln B_{iyw} + \Delta\epsilon'_{iyw}$,

where y represents one-year time periods. An advantage of the longitudinal analysis (ignoring for the moment econometric problems to be discussed) is that person-specific fixed effects owing to unmeasured quality fall out, allowing unbiased estimates of Θ_b and Θ_w.

III. DATA AND DESCRIPTIVE EVIDENCE

Our study utilizes several sources of data. The primary wage-level analysis is conducted using a large sample (n=1,611,829) of black and white workers from the monthly Current Population Survey Outgoing Rotation Group (CPS ORG) files for the period from January 1983 through December 1992.[5] We match to each individual record in the CPS variables measuring job characteristics. These variables are either obtained directly from the *Dictionary of Occupational Titles*, calculated by us from the CPS ORG files, or calculated from special CPS supplements. Wage-change equations are estimated using three alternative panel or retrospective data sets. Because households are included in the CPS in the same month for two consecutive years, the CPS ORG files permit construction of large panels of individuals in adjacent years for the periods 1983/4 through 1991/2 (n=392,877 pairs of observations). A detailed description of the construction of the CPS ORG panel files is provided in Appendix 1. We also estimate wage-change equations using the 1983–92 March CPS Annual Demographic files (n=133,992), which contain information for outgoing rotation groups on wages and occupation of the current job, as well as for the longest-held job last year. Finally, change equations are estimated based on the January 1984, 1986, 1988, 1990, and 1992 CPS Displaced Worker Surveys (DWS), which provide current and retrospective job information for workers who have had permanent layoffs within the past five years (n=10,634). The advantages and disadvantages of these longitudinal data sets are discussed in Section V. In addition to the data sets used in our regression analysis, descriptive information on wages, the racial wage gap, and racial composition are provided for the

1973–92 period using the 1973–78 May CPS files and the 1979–82 CPS ORG files, in addition to the 1983–92 CPS ORG files described above (a total sample of 2,500,253).

Included in the analysis are all black and white wage and salary workers ages sixteen and over, with complete data provided on usual weekly earnings, usual hours worked per week, occupation, and other needed variables. Excluded are workers whose principal activity is school (about 3½ percent of the potential sample) or who had either occupation or industry allocated by the Census (about 1 percent). Wage rates are measured by usual weekly earnings divided by usual hours worked per week, in December 1992 dollars, with the monthly CPI-U as the deflator. Excluded from the sample are workers with measured real wage rates less than $1.00 (about 0.1 percent). Nominal weekly earnings are top-coded by the Census at $999 for the surveys through 1988 and at $1,923 after 1988. We assign to workers whose weekly earnings are at least $999 in December 1988 dollars ($1,176.42 in December 1992 dollars) the value $1,553.70, representing mean earnings for the post-1988 sample exceeding the $1,176.42 limit.

Racial composition is measured by *B*, the proportion of black workers in the worker's 3-digit Census occupation, calculated from the CPS ORG files. It is measured by the ratio of black to total employment by occupation (*B* is calculated for all wage and salary workers ages sixteen and over, with no other sample restrictions and with nonwhite nonblacks included in the denominator). In Table 1, measures of *B* for 1973–92 are provided based on two-year averages for 1973–74, 1975–76, and 1977–78 using the 1973–78 May CPS files (these include all CPS rotation groups but for May only), and on an annual basis thereafter based on the CPS ORG monthly files for 1979–92 (these include one-quarter of the total CPS samples but for all months in a year). In subsequent regression analyses for the 1983–92 period, *B* is measured by a three-year moving average. This is intended to reduce measurement error resulting from small sample sizes in some occupation cells, most

TABLE 1
Black and White Mean Wages, Racial Wage Gaps, Racial Composition, and Occupational Segregation by Year, 1973–1992

	Blacks			Whites			Wage	Adjusted	Duncan
Year	N	Wage	*B*	N	Wage	*B*	Ratio	Wage Gap	Index
Males:									
1973/4	3,480	12.17	.155	38,858	15.62	.087	.779	-.111	.367
1975/6	3,258	11.77	.150	36,568	15.24	.084	.772	-.097	.363
1977/8	3,459	11.71	.143	40,866	15.30	.088	.765	-.118	.336
1979	7,341	11.55	.141	80,167	14.58	.091	.792	-.081	.317
1980	8,288	11.01	.136	91,805	13.93	.090	.790	-.080	.303
1981	7,606	10.98	.137	85,229	13.78	.089	.797	-.079	.311
1982	6,935	10.74	.137	79,176	13.85	.089	.776	-.090	.321
1983	7,023	10.99	.138	78,067	13.92	.091	.790	-.095	.316
1984	7,119	10.89	.140	78,371	13.90	.094	.784	-.089	.315
1985	7,379	10.91	.142	79,278	14.07	.095	.775	-.102	.318
1986	7,497	11.04	.140	77,673	14.28	.096	.773	-.116	.308
1987	7,643	11.03	.144	77,624	14.32	.098	.770	-.117	.305
1988	7,150	11.03	.143	74,491	14.09	.099	.783	-.104	.297
1989	7,200	10.80	.143	75,951	13.94	.100	.775	-.123	.297
1990	7,430	10.68	.142	79,444	13.91	.099	.768	-.130	.293
1991	7,113	10.67	.144	76,451	13.74	.100	.777	-.134	.292
1992	6,834	10.49	.143	74,686	13.64	.099	.769	-.145	.296
Females:									
1973/4	3,335	9.10	.213	26,370	10.07	.100	.904	.003	.367
1975/6	3,341	9.04	.202	26,180	9.91	.098	.912	.013	.352
1977/8	3,622	9.26	.187	30,971	9.82	.099	.943	.008	.342
1979	7,597	9.00	.173	61,325	9.43	.101	.954	.010	.302
1980	8,761	8.69	.169	72,855	9.06	.102	.959	.020	.296
1981	8,268	8.56	.165	68,990	9.04	.102	.947	.009	.296
1982	7,829	8.62	.162	65,944	9.19	.101	.938	.004	.299
1983	7,872	8.81	.159	65,786	9.35	.102	.942	-.013	.299
1984	8,115	8.79	.164	65,909	9.43	.106	.932	-.012	.300
1985	8,572	8.96	.162	67,556	9.60	.107	.933	-.002	.293
1986	8,747	9.15	.157	67,928	9.85	.107	.929	-.024	.291
1987	9,004	9.22	.157	68,915	9.96	.109	.926	-.028	.289
1988	8,596	9.12	.160	66,247	9.96	.111	.916	-.035	.280
1989	8,617	9.29	.157	68,032	10.15	.111	.915	-.046	.269
1990	9,163	9.32	.155	71,577	10.28	.110	.907	-.056	.269
1991	8,877	9.50	.154	69,849	10.35	.111	.918	-.045	.274
1992	8,852	9.48	.155	69,191	10.46	.111	.906	-.043	.274

Calculations are from the 1973–78 May CPS and the 1979–92 Annual CPS ORG Files (n=2,500,253). Wages are measured by usual weekly earnings divided by usual hours worked, in December 1992 dollars. Adjustments for top-coding are described in the text. *B* measures the proportion of black to total employment in workers' detailed occupation. The wage ratio is the mean of black to white real wages; the adjusted log wage gap is the regression coefficient on a black dummy from a log wage equation with controls for years of schooling completed, potential experience (measured by age minus schooling minus 6) and its square, and dummies for married spouse present, ever-married spouse not present, part-time, public sector, region (8), large metropolitan area, industry (13), and occupation (5). In order to insure a time-consistent specification, union status and separate federal, state, and local dummies are not included. The Duncan segregation index is calculated by year by $\frac{1}{2}\Sigma|w_j - b_j|$, where w and b are the proportions of nonblack and black employment in occupation j.

particularly in the wage-change analysis. In addition, ln*B* rather than *B* is included in the regression analysis owing to a nonlinear relationship between lnW and *B*.

Table 1 presents evidence for the period 1973–92 on black and white wage rates, wage ratios, the log wage gap from a regression model with standard control variables, the racial composition as measured by *B*, and the Duncan index of segregation. The Duncan index, defined as $\frac{1}{2}\Sigma|w_j-b_j|$, where w_j and b_j are the proportions of nonblack and black employment in occupation j, ranges between 0 designating complete integration (i.e., the distributions of black and white employment across occupations are equivalent) and 1 designating complete segregation (occupations are either all black or all white).

Focusing first on wage rates and the racial gap among males, we see that black and white wage rates fell over the 1973–92 period, while there was remarkably little change in the black/white wage ratio. The apparent constancy of the racial wage ratio masks several important changes. In tabulations not shown, as well as in work by others (e.g., Bound and Freeman, 1992), there is evidence of a widening racial gap during the 1980s among younger and more-educated cohorts. Our figures for women indicate substantial real wage growth among both races and smaller racial wage gaps than among men. But the evidence clearly indicates a widening racial gap among women during the 1980s and early 1990s (for related evidence, see Card and Lemieux, 1993). In fact, the adjusted wage gap for women turns from slightly positive in the 1970s and early 1980s to about a 5 percent wage disadvantage by the 1990s.

Descriptive evidence presented in Table 1 helps identify changes over time in the racial composition of occupations. Provided are mean values of *B* for black and white males and females for the period 1973–92, as well as the Duncan index of segregation, calculated separately by gender. Both pieces of evidence indicate declining racial segregation, particularly among women. *B* has increased over the twenty-year period by about 1 percentage point for white men and women (from roughly 9 to 10 percent for men and from 10 to 11 percent for women). *B* has decreased most significantly for black women, from about 21 to 16 percent, while decreasing from 15 to 14 percent for black males.

Similarly, the Duncan segregation index has declined from .367 to .296 among males, and from .367 to .274 among women.[6] Much of the decline in occupational segregation occurred during the 1970s; there has been no apparent progress during the 1990s.

Table 2 presents the means of wages, schooling, and selected occupational variables, classified by *B* category for each race-sex group for the pooled 1983–92 sample. Workers are segmented into four occupational categories based on racial composition, with breakpoints for *B* of .04, .09, and .16 (the pattern of results is invariant to alternative breakpoints). Average wages for black and white workers decrease substantially as *B* increases, although this pattern is clearly stronger among men than among women. Average years of schooling likewise decline with respect to *B*, indicating that little can be said about the effect of racial composition on wages, absent detailed control for skill-related variables. Occupation-level variables that either proxy or measure directly skill level differ systematically with respect to racial composition. Average job tenure is lowest in occupations with a large *B*, the proportion of part-time jobs increases with respect to *B*, the proportion of workers receiving formal job training is lowest in high *B* jobs, computer use falls as *B* increases, and values of the DOT skill measures GED and SVP decline sharply with respect to racial composition.[7] DOT measures of occupational work conditions indicate less pleasant conditions in jobs with relatively high proportions of black workers, with environmental disamenities, hazards, physical requirements, and strength being higher in high *B* than in low *B* occupations. Everything else equal, differences in job amenities should lead to equilibrium wages that rise with *B*. Subsequent regression analyses will examine directly how much of the relationship between wages and *B* can be accounted for by various combinations of occupational characteristics.

TABLE 2
Means of Selected Individual and Occupational Variables by Race and Racial Composition Category, 1983–1992

	Black Means Value of *B*					White Means Value of *B*				
	0–.04	.04–.09	.09–.16	.16+	All	0–.04	.04–09	.09–16	.16+	All
Males:										
Wage (12/92 $)	15.971	13.286	10.429	9.095	10.855	18.546	15.619	11.714	9.794	13.981
Schooling	14.526	13.078	11.854	11.426	12.095	14.721	13.546	12.275	11.565	13.027
DOT-GED	4.240	3.559	2.380	1.945	2.583	4.265	3.696	2.498	1.937	3.158
DOT-SVP	4.170	3.090	0.998	0.523	1.456	4.120	3.373	1.217	0.552	2.422
OCC-Training	0.512	0.460	0.296	0.227	0.320	0.522	0.470	0.323	0.226	0.398
OCC-Computer	0.479	0.348	0.201	0.095	0.210	0.483	0.380	0.222	0.098	0.305
OCC-Job Tenure	7.300	7.793	6.624	6.054	6.742	7.437	7.864	6.901	6.147	7.283
OCC-Part-time	0.088	0.088	0.164	0.221	0.162	0.079	0.078	0.151	0.207	0.118
DOT-Environment	0.105	0.342	0.407	0.595	0.446	0.129	0.301	0.405	0.618	0.357
DOT-Hazards	0.074	0.190	0.193	0.216	0.196	0.084	0.173	0.202	0.220	0.178
DOT-Physical	1.239	1.777	1.956	1.925	1.878	1.204	1.648	1.928	1.891	1.712
DOT-Strength	1.814	2.248	2.559	2.743	2.521	1.836	2.155	2.515	2.733	2.306
N	2,203	18,184	27,030	24,971	72,388	91,632	342,341	227,120	110,943	772,036
Females:										
Wage (12/92 $)	13.340	11.145	9.546	7.541	9.171	13.066	10.742	9.481	7.874	9.946
Schooling	14.620	13.594	12.963	11.502	12.543	14.310	13.472	13.037	11.784	13.071
DOT-GED	4.246	3.770	2.928	2.298	2.896	4.022	3.716	3.017	2.382	3.272
DOT-SVP	3.997	2.548	1.309	0.578	1.352	3.467	2.457	1.384	0.660	1.849
OCC-Training	0.533	0.450	0.394	0.276	0.361	0.505	0.434	0.406	0.294	0.403
OCC-Computer	0.510	0.523	0.388	0.150	0.326	0.466	0.502	0.397	0.185	0.410
OCC-Job Tenure	6.425	6.413	6.390	5.638	6.088	6.169	6.217	6.351	5.526	6.129
OCC-Part-time	0.129	0.201	0.247	0.298	0.255	0.164	0.219	0.253	0.286	0.239
DOT-Environment	0.041	0.068	0.133	0.396	0.223	0.039	0.058	0.120	0.400	0.139
DOT-Hazards	0.027	0.025	0.044	0.156	0.085	0.022	0.020	0.039	0.152	0.050
DOT-Physical	1.053	1.404	1.402	1.954	1.624	1.110	1.390	1.362	1.915	1.463
DOT-Strength	1.698	1.617	1.893	2.454	2.052	1.756	1.617	1.884	2.367	1.844
N	1,153	21,413	28,436	35,413	86,415	34,610	311,882	209,574	124,924	680,990

All means are calculated across individuals in the 1983–92 CPS ORG. Wage is usual weekly earnings divided by usual weekly hours, in December 1992 dollars, with treatment for top-coding discussed in the text. Schooling measures years completed. Variables preceded by "OCC" and "DOT" are means of variables in workers' designated occupations; all except OCC-Computer are fixed over the 1983–92 period. OCC-Training is the proportion of workers receiving formal or informal training on the current job, calculated from the January 1983 and 1991 CPS supplements. OCC-Computer is the proportion using computers at their job, calculated from the October 1984 and 1989 CPS supplements and trended linearly back to 1983 and forward to 1992. OCC-Tenure is years with current employer, calculated from the May 1983 and 1988 CPS Pension Supplements and the January 1983, 1987, and 1991 CPS. OCC-Part-time is the proportion of workers in the occupation who usually work fewer than thirty-five hours a week, calculated from the CPS ORG files. DOT measures are taken from the *Dictionary of Occupational Titles* and matched to workers based on 1980 Census of Population occupation codes: DOT-GED is a 1–6 index of general educational development, DOT-SVP is years required for occupational proficiency or specific vocational preparation, DOT-Environment is the number of work environment disamenities from 0–5, DOT-Hazards is the proportion in hazardous jobs, DOT-Physical is the number of physical demands from 0–4, and DOT-Strength is measured by a 1–5 index from low to high strength required.

IV. WAGES AND RACIAL COMPOSITION: CROSS-SECTIONAL EVIDENCE

Table 3 presents the racial composition coefficients from alternative specifications of log wage equations, estimated in levels for black males, white males, black females, and white females, using the pooled CPS ORG files for 1983–92. We provide coefficients from two specifications, a "standard" model including a typical set of individual and labor market characteristics, and an "expanded" model including a detailed set of job characteristics, most of which measure the skill composition or working conditions of the occupation.

Included in the standard specification are variables measured at the individual level: years of schooling completed and potential experience and its square; and dummies for union coverage, part-time, marital status (2), public sector (3), region (8), large metropolitan area, industry (13), occupation (5), and year (9). The expanded specification adds variables measuring means at the occupation and industry levels. Occupation variables included are DOT-GED, a 1–6 index of general educational development measuring necessary reasoning, writing, and mathematical skills; DOT-SVP, representing specific vocational preparation and measured by years required for proficiency in an occupation; OCC-Training, the proportion of workers receiving formal or informal training on the current job; OCC-Computer, the proportion of workers using a computer on the job; OCC-Job Tenure, years with current employer; OCC-Part-time, the proportion of workers in the occupation who usually work fewer than thirty-five hours a week; DOT-Environment, the number of work environment disamenities from 0–5; DOT-Hazards, the proportion in hazardous jobs; DOT-Physical, the number of physical demands from 0–4; and DOT-Strength, a 1–5 index from low to high strength required. Industry-level variables included are IND-Union, measuring the proportion covered by a union in the worker's industry; and IND-Big Firm, measuring the proportion of workers in firms with at least 1000 employees. The notes to Tables 2 and 3 provide the sources from which these variables are calculated.

TABLE 3

Racial Composition Coefficients with Log, Linear, and Quadratic Composition Measures, Pooled Data Set, 1983–1992

Group and Specification	Model 1 ln*B*	Model 2 *B*	Model 3 *B*	 B^2	Model 1' ln*B**
Black Males:					
Standard	-0.0958 (0.0038)	-0.6861 (0.0299)	-1.5758 (0.0844)	2.5529 (0.2267)	-0.0990 (0.0036)
Expanded	-0.0293 (0.0042)	-0.1519 (0.0326)	-0.6154 (0.0903)	1.2944 (0.2353)	-0.0349 (0.0041)
White Males:					
Standard	-0.0841 (0.0009)	-1.0562 (0.0113)	-2.2735 (0.0301)	4.5426 (0.1041)	-0.0850 (0.0009)
Expanded	-0.0478 (0.0010)	-0.5303 (0.0127)	-1.6223 (0.0325)	3.8951 (0.1066)	-0.0480 (0.0010)
Black Females:					
Standard	-0.0476 (0.0030)	-0.1823 (0.0213)	-1.3907 (0.0645)	3.0729 (0.1548)	-0.0727 (0.0030)
Expanded	-0.0176 (0.0035)	0.0011 (0.0244)	-0.8546 (0.0730)	2.1467 (0.1725)	-0.0378 (0.0035)
White Females:					
Standard	-0.0461 (0.0010)	-0.2096 (0.0090)	-1.6573 (0.0257)	4.5338 (0.0752)	-0.0768 (0.0009)
Expanded	-0.0305 (0.0011)	-0.0637 (0.0103)	-1.3859 (0.0282)	3.9956 (0.0793)	-0.0528 (0.0011)

*Excludes workers whose wages are less than 1.1 times the minimum wage.

Sample sizes for Models 1–3 are as follows: black males, 72,388; white males, 772,036; black females, 86,415; and white females, 680,990. Sample sizes for Model 1' are 65,331; 732,693; 73,887; and 606,702, respectively. The "standard" specifications include variables measured at the individual level: years of schooling completed and potential experience and its square; and dummies for union coverage, part-time, marital status (2), public sector (3), region (8), large metropolitan area, industry (13), occupation (5), and year (9). The "expanded" specifications add variables measuring means at the occupation and industry levels. Occupation variables included are those defined in Table 2: DOT-GED, DOT-SVP, OCC-Training, OCC-Computer, OCC-Job Tenure, OCC-Part-time, DOT-Environment, DOT-Hazards, DOT-Physical, and DOT-Strength. Additional occupational- and industry-level variables are OCC-Female, measuring the proportion female and calculated from the CPS ORG files; IND-Union, measuring the proportion covered by a union in the worker's industry and calculated from the CPS ORG files; and IND-Big Firm, measuring the proportion in the worker's industry in firms with ≥1000 employees, calculated from the May 1983 CPS Pension Supplement and the March 1989–92 CPS files. Standard errors are in parentheses.

Column one presents coefficients from equations including our preferred variable, ln*B*, as the measure of racial composition. For comparison, we present results from alternative equations, one including *B* and another *B* and B^2. As evident from the quadratic specification, the relationship is clearly nonlinear, with lnW decreasing with respect to *B* at a decreasing rate. Coefficients on *B* from the linear specification turn out to be misleading, differences across groups resulting more from being measured at different levels of *B* than from true differences in the wage-composition relationship. For example, the linear specification tends to exaggerate differences in the racial composition effect for white relative to black males, owing to the fact that white workers are observed at relatively low levels of *B*. We prefer the simple ln*B* specification because this facilitates a compact presentation of results, allows easy comparison across groups of workers and for models with alternative explanatory variables, and provides a statistical fit superior to the linear and highly similar to specifications with a quadratic term or (in work not shown) with a set of racial density dummies.[8]

Results from the "standard" model, presented in Table 3, indicate significantly lower wages for black and white males and, to a lesser extent, black and white females, in occupations with high concentrations of black workers. These results are consistent with the far more limited evidence presented previously in Hirsch and Schumacher (1992). Point estimates close to -.10 for males indicate that each 10 percent increase in *B* is associated with wage rates 1 percent lower; point estimates are half as large for women.[9]

We next investigate the quality sorting hypothesis in an "expanded" model that introduces explicit measures of occupational skill and other job characteristics (in a later section we estimate wage-change models controlling for unmeasured person-specific skills). Appendix 2 includes complete regression results for the expanded model. Coefficients on most variables are consistent with expectations, although signs on job characteristics cannot be predicted unambiguously given worker heterogeneity (Hwang, Reed, and Hubbard, 1992). We forgo discussion of these results in order to save space. Inclusion of occupational characteristics again sharply reduces the effect of racial composition for all four demographic groups, reducing coefficients in general by more than half. For males, the elasticity of wages with respect to *B* is about -.03 to -.05, while for women the elasticities are even smaller (-.02 to -.03). The sharp reduction in the coefficient on ln*B* resulting from adding skill-related individual and occupational variables provides strong support for the thesis that *B* serves primarily as an occupational skill index. Even in the expanded model, however, racial composition remains a significant and nontrivial determinant of earnings for both white and black workers.

We next consider the possibility that the pattern of results found to this point is due in part from the presence of a minimum wage floor. Our evidence indicates a nonlinear relationship between the log wage and *B* (hence the use of ln*B* or a quadratic specification), as well as smaller racial composition effects in the South, among women, and for less-educated workers. A possible explanation for these results is that these workers are most likely to be affected by a binding minimum wage floor. Among low-wage workers and at high levels of *B*, further increases in *B* can have relatively little negative effect since wages are already close to their minimum.[10]

In order to explore this issue, we estimate racial density coefficients with a sample excluding workers below, at, or near the minimum wage. We exclude all workers with wage rates below 1.1 times the minimum wage, in December 1992 dollars.[11] These results are shown in the far right column of Table 3 (Model 1'). Omitting low-wage workers has the predicted result. Coefficient estimates for males are little affected, since relatively few are below the wage threshold. By contrast, coefficients for both black and white women increase in absolute value following the exclusion of low-wage workers. In fact, differences in coefficients on ln*B* between women and men are rather minor following this exclusion. Although the minimum wage has an effect on estimates of Θ among women, it is not sufficient to explain the nonlinearity of the lnW-*B* relationship for either women or men. Use of either ln*B* or a quadratic in *B* remains strongly preferred to the linear specification; the nonlinear relationship occurs in the data at wages well above the minimum wage threshold. An implication of our results is that minimum wage laws (at historical levels) have done little to reduce the negative relationship between racial density and the wages for white and black men, but have mitigated these effects among women.

Table 4 provides estimates of Θ from alternative specifications of lnW equations. We present in turn models with only ln*B* and no controls (line 1); a base model with all individual characteristics but no industry or occupation dummies (2); the base model plus thirteen industry dummies (3); the base model plus five occupation dummies (4); the standard model seen previously, consisting of the base model plus industry and occupation dummies (line 5); the standard model plus the separate addition of occupation-level skill measures DOT-GED (6), DOT-SVP (7), OCC-Training (8), OCC-Computer (9), OCC-Job Tenure (10), and OCC-Part-time (11); the standard model plus the joint addition of these six skill-related variables (12); the standard model plus the DOT work condition variables (13); and the expanded model seen previously, containing all individual, occupation, and industry variables (14).

If ln*B* serves as an index for otherwise unmeasured job skills, as implied by the quality sorting hypothesis, its coefficient should move toward zero as one introduces variables that measure directly or are correlated with worker skills. The evidence in Table 4 supports quite clearly this interpretation. The magnitude of Θ is reduced most sharply in response to the inclusion of individual characteristics (line 2), broad occupation dummies (4), DOT-GED (6), and OCC-Computer (9). Comparing lines (13) and (5), DOT working conditions have little effect on estimates of Θ for males, but increase the magnitude of the estimate for women. That is, women in occupations with relatively high densities of black workers tend to have working conditions associated with significantly higher

TABLE 4

Racial Composition Coefficients from Alternative Specifications

Specification	Black Males	White Males	Black Females	White Females
1. No Controls	-0.2732	-0.2844	-0.2820	-0.1892
	(0.0034)	(0.0009)	(0.0028)	(0.0010)
2. Base (individual characteristics only)	-0.1825	-0.1556	-0.1447	-0.0979
	(0.0030)	(0.0008)	(0.0026)	(0.0009)
3. Base + 13 industry dummies	-0.1911	-0.1560	-0.1508	-0.1216
	(0.0029)	(0.0008)	(0.0026)	(0.0009)
4. Base + 5 occupation dummies	-0.0871	-0.0909	-0.0335	-0.0153
	(0.0037)	(0.0009)	(0.0030)	(0.0010)
5. Standard model (Base + 5 occ., 13 ind., dummies)	-0.0958	-0.0841	-0.0476	-0.0461
	(0.0038)	(0.0009)	(0.0030)	(0.0010)
6. Standard + GED	-0.0511	-0.0491	-0.0097	-0.0164
	(0.0040)	(0.0010)	(0.0032)	(0.0010)
7. Standard + SVP	-0.0895	-0.0806	-0.0579	-0.0508
	(0.0037)	(0.0009)	(0.0030)	(0.0010)
8. Standard + Training	-0.0815	-0.0729	-0.0446	-0.0414
	(0.0038)	(0.0009)	(0.0030)	(0.0010)
9. Standard + Computer	-0.0367	-0.0422	-0.0058	0.0003
	(0.0041)	(0.0010)	(0.0033)	(0.0010)
10. Standard + Job Tenure	-0.0665	-0.0667	-0.0473	-0.0438
	(0.0038)	(0.0009)	(0.0030)	(0.0010)
11. Standard + Part-time	-0.0733	-0.0780	-0.0242	-0.0351
	(0.0038)	(0.0009)	(0.0031)	(0.0010)
12. Standard+SVP, GED, Training, Computer, Job Tenure, Part-time	-0.0347	-0.0499	-0.0196	-0.0155
	(0.0041)	(0.0010)	(0.0034)	(0.0011)
13. Standard + DOT measures other than SVP, GED	-0.0855	-0.0795	-0.0555	-0.0641
	(0.0038)	(0.0009)	(0.0031)	(0.0010)
14. Expanded (Standard + all job characteristics)	-0.0293	-0.0478	-0.0176	-0.0305
	(0.0042)	(0.0010)	(0.0035)	(0.0011)
N	72,388	772,036	86,415	680,990

Data set is the CPS ORG files for 1983–92. Shown are coefficients and standard errors on the racial composition variable ln*B*. See Tables 2 and 3 for definitions of individual, occupation, and industry variables. The base model includes years of schooling completed; potential experience and its square; and dummies for union coverage, part-time, marital status (2), public sector (3), region (8), and year (9). Standard errors are in parentheses.

pay. The negative wage-composition relationship exists in spite of these differences. Although the magnitudes of estimated racial composition effects have been reduced sharply by the inclusion of skill-related individual and job characteristics, ln*B* remains a significant determinant of wage rates for all workers, regardless of gender or race (line 14).

Table 5 presents estimates of the racial composition relationship for the four demographic groups based on classification by age, schooling, region, private-public sector, union status, part-time and full-time, production worker status, and Hispanic status. Although there is variability in the pattern of coefficients across the demographic groups, the effects of racial composition generally are largest among prime-age workers, those with higher education, in the non-South, in the private sector, in nonunion jobs, among full-time workers, and in nonproduction jobs. The results with respect to schooling are consistent with the quality sorting explanation, since the black-white ability gap widens as schooling levels increase (O'Neill, 1990). In fact, the broad tendency is for racial density effects to be larger in broad sectors with a substantial number of information-based occupations and substantial heterogeneity in skill levels across occupations. The fact that the marginal effects of racial composition are generally smaller where blacks are overrepresented (e.g., less educated, South, young, union, and production workers) argues against crowding explanations of the racial composition effect.

In results not presented, we examine changes over time in Θ, the coefficient on ln*B*. Coefficients for males are relatively stable over time, apart from evidence of a moderate increase in the magnitude of Θ beginning in 1989. For women, however, we find clear-cut evidence of an increase in the effect of ln*B* through time, with weak effects of ln*B* during the early and mid-1980s, followed by increases during the late 1980s for black women and in 1989 for white women. Note that the unexplained racial wage gap among women increased sharply during this same time period (Table 1). Although not further explored here, the clear suggestion is that demand shifts in the

TABLE 5

Racial Composition Coefficients among Alternative Worker Groups, Wage-Level Equations, Pooled for 1983–1992

	Black Males		White Males		Black Females		White Females	
Group	Standard	Expanded	Standard	Expanded	Standard	Expanded	Standard	Expanded
All Workers	-.0958	-.0293	-.0841	-.0478	-.0476	-.0176	-.0461	-.0305
Ages 16–29	-.0846	-.0301	-.0614	-.0202	-.0458	-.0298	-.0333	-.0271
Ages 30–44	-.1079	-.0277	-.0883	-.0489	-.0653	-.0254	-.0630	-.0394
Ages 45–99	-.0686	-.0158	.0926	-.0632	-.0096	-.0002	-.0251	-.0160
Educ: 0–11	-.0270	.0168	-.0575	-.0221	.0363	.0294	.0500	.0286
Educ: 12	-.0713	-.0118	-.0713	-.0320	-.0343	-.0121	-.0321	-.0248
Educ: 13–15	-.1072	-.0437	-.0817	-.0450	-.0623	-.0350	-.0542	-.0598
Educ: 16	-.1564	-.0872	-.1019	-.0706	-.1443	-.0823	-.1189	-.0701
Educ: >16	-.0968	-.0725	-.0670	-.0673	-.1057	-.0471	-.0663	-.0237
Non-South	-.1091	-.0534	-.0842	-.0482	-.0468	-.0281	-.0440	-.0318
South	-.0796	-.0041	-.0837	-.0473	-.0428	-.0087	-.0509	-.0269
Private	-.0844	-.0356	-.0825	-.0604	-.0448	-.0225	-.0476	-.0402
Public	-.0954	-.0098	-.0637	-.0110	-.0606	-.0234	-.0545	-.0068
Nonunion	-.0946	-.0319	-.0789	-.0517	-.0459	-.0152	-.0478	-.0350
Union	-.0718	-.0135	-.0674	-.0224	-.0473	-.0242	-.0335	-.0083
Full-time	-.1029	-.0363	-.0811	-.0484	-.0655	-.0334	-.0590	-.0360
Part-time	-.0002	.0051	-.1026	-.0660	.0244	-.0080	-.0153	-.0361
Nonproduction	-.1142	-.0525	-.0914	-.0594	-.0450	-.0263	-.0443	-.0336
Production	-.0598	.0031	-.0546	-.0209	-.0393	.0723	-.0668	.0204
Non-Hispanic	-.0961	-.0294	-.0830	-.0488	-.0474	-.0172	-.0468	-.0322
Hispanic	-.0718	-.0246	-.0860	-.0330	-.0622	-.0525	-.0478	-.0272

Standard and expanded specifications are described in note to Table 3. Shown are coefficients on the racial composition variable ln*B*. All models include year dummies. Sample sizes and standard errors are not shown.

economy since the late 1980s that favored information-based jobs and increased the returns to skill help explain the pattern of racial composition effects over time.[12] This interpretation further supports the view that *B* serves largely as an index of occupational skill.

A final probe of our wage-level results concerns potential problems associated with matching aggregate data on occupations to individual worker data. Matching grouped to individual data biases downward standard errors, but need not bias coefficient estimates. To provide a check on our results, we use a two-step estimation strategy suggested by Dickens and Ross (1984) utilizing all the information in the sample but without biased standard errors. In the first-step wage equation, only variables measured at the individual level or that vary within detailed occupations are included. The mean of the equation's error term for each detailed occupation is then calculated (this is equivalent to including detailed occupation dummies). The wage differences by occupation calculated in the first step form the dependent variable in a second-step regression. These are estimated by WLS, with occupation sample sizes as weights (this corresponds to the implicit weighting in our previous single equation results), as well as by OLS, which gives equal weights to occupation rather than individuals. Included in the second-step equations are $\ln B$ and broad occupation dummies for the standard model; the expanded model includes all occupational variables that vary across but not within detailed occupations. Table 6 presents these results. As expected, the WLS point estimates are close to what we obtained previously using single-step estimation, with a tendency to be slightly larger in magnitude. Standard errors are no longer biased downward, however, being substantially larger than those shown previously. Estimated OLS coefficients, based on equal weights to occupation cells, are close to zero and generally insignificant.

TABLE 6

Second-Step Aggregate OLS and WLS Estimates of Racial Composition Coefficients, with Correction for Standard Errors

	Black		White	
	OLS	WLS	OLS	WLS
Males:				
Standard	-.0195	-.0900	-.0245	-.0849
	(.0141)	(.0096)	(.0057)	(.0086)
Expanded	.0294	-.0377	-.0053	-.0570
	(.0146)	(.0091)	(.0057)	(.0090)
N	479		495	
Females:				
Standard	-.0359	-.0356	-.0044	-.0533
	(.0182)	(.0103)	(.0074)	(.0100)
Expanded	.0206	-.0107	.0124	-.0406
	(.0188)	(.0101)	(.0075)	(.0107)
N	438		491	

Standard and expanded specifications are described in note to Table 3. All models include year dummies. Standard errors are in parentheses. Units of observation are 3-digit Census occupations. N equals the number of non-empty occupation cells following sample restrictions, out of a possible 497 CPS occupations (made time-consistent between the 1983–91 files using 1980 Census of Population codes and the 1992 file using 1990 codes). The dependent variable in the two-step model is the mean of the error term, by occupation, from a first-step log wage regression including all variables that vary within occupations. The second-step regression includes ln*B* and all other variables that vary across but not within occupations. Weighted least squares (WLS) estimates use occupation sample sizes as weights.

V. WAGES, RACIAL COMPOSITION, AND QUALITY SORTING: LONGITUDINAL EVIDENCE

In order to test directly the hypothesis that unmeasured worker skill differences correlated with the racial composition of occupations account for much of the observed negative relationship between *B* and wages, longitudinal wage-change equations are estimated. As explained previously, variant three of the quality sorting hypothesis implies that wage changes among occupational switchers should not be systematically correlated with changes in racial composition. We estimate wage-change equations using three alternative panel or retrospective data sets, each with advantages and disadvantages.

Econometric problems with longitudinal analysis are potentially serious and warrant discussion up front. Many wage-change studies are plagued by relatively small overall sample sizes, and in particular small numbers of true switchers for the variable under consideration (e.g., union status). By contrast, the CPS panels developed here provide very large sample sizes with many occupational switchers. Because our matched pairs are for adjacent years, we must assume that wage changes associated with the change in racial composition are realized quickly. This implies that the fixed quality effects that we net out in the longitudinal analysis, while unmeasured by the researcher, are known to the employer and quickly reflected in the wage following a job switch. More problematic is the maintained assumption that changes in racial composition are exogenous. We know that job switching is potentially endogenous—voluntary switchers are more likely to switch because of wage increases, involuntary switchers are more likely to face wage loss, and switchers are likely to differ from stayers. Endogenous job switching need not bias estimates of Θ, however, if occupational switching is uncorrelated with changes in racial composition. The problem of endogenous switching is addressed through the use of the Displaced Worker Surveys (DWS), where occupational changes are largely involuntary and far more likely to be exogenous.

A more serious problem in the longitudinal analysis is possible coefficient bias toward zero resulting from measurement error in $\Delta \ln B$. Bias is most serious in longitudinal models where intertemporal variance owing to measurement error is large relative to true variance of a right-hand-side variable (for an application, see Freeman, 1984). Such bias is potentially important in the CPS ORG matched panels since a substantial number of persons have occupation misclassified in the CPS (Mellow and Sider, 1983) and the time period of the panels is short (i.e., one year), leading to a relatively high noise to signal ratio. Bias associated with mismeasured occupational change is addressed in several ways, including the use of alternative data sets (the March CPS and DWS files) containing retrospective occupation information providing more accurate occupational change measures than do the CPS ORG panels.

Although we expect bias from measurement error in the longitudinal analysis, several points are in order. Our sample excludes all worker-year pairs where occupation (or industry) has been allocated by the Census in either the first or second year. Second, there exists serial correlation in response error (for evidence on earnings, see Bound and Krueger, 1991) so that a stayer who reports the incorrect occupational category in year 1 may report the same category in year 2. In this case, "two wrongs make a right" since the person would be classified correctly as having $\Delta \ln B=0$. Most important, among those workers misclassified by occupation, it is likely that they report a closely related occupation whose B is similar to B in their actual occupation.[13]

The primary way in which we reduce measurement error bias in the CPS ORG panels is by separating occupational switchers who do and do not report a change in industry. The CPS ORG longitudinal results in Table 7 provide coefficient estimates on $\Delta \ln B$ for workers who switch *both* occupation and industry; separate coefficients are estimated for those who change occupation but not industry (these are not shown). The reason for this procedure is that workers who report changes both in industry and occupation are far more likely to be true occupational changers than those

TABLE 7

Panel Data Estimates of ln*B* and Δln*B* Coefficients for Wage-Level and Wage-Change Models Using the CPS ORG, March CPS, and CPS Displaced Worker Surveys

	Black Males		White Males		Black Females		White Females	
	Levels	Change	Levels	Change	Levels	Change	Levels	Change
CPS ORG Panel:								
Standard	-.1011	-.0078	-.0867	-.0326	-.0582	.0020	-.0519	.0040
	(.0078)	(.0128)	(.0018)	(.0029)	(.0066)	(.0114)	(.0020)	(.0032)
Expanded	-.0317	.0089	-.0535	-.0250	-.0248	.0131	-.0314	.0037
	(.0086)	(.0137)	(.0020)	(.0031)	(.0075)	(.0123)	(.0023)	(.0034)
N	16,229		190,274		20,445		165,929	
March CPS Panel:								
Standard	-.0872	-.0254	-.0758	-.0281	-.0451	-.0145	-.0454	.0122
	(.0142)	(.0306)	(.0033)	(.0063)	(.0111)	(.0263)	(.0035)	(.0078)
Expanded	-.0150	.0296	-.0452	-.0182	-.0311	-.0433	-.0314	.0038
	(.0160)	(.0342)	(.0037)	(.0066)	(.0130)	(.0305)	(.0041)	(.0089)
N	5,498		59,747		6,628		62,119	
DWS—Plant Closings and Layoffs Sample:								
Standard	-.1059	-.0132	-.1077	-.0243	-.0393	.0005	-.0769	-.0076
	(.0358)	(.0328)	(.0109)	(.0095)	(.0408)	(.0403)	(.0149)	(.0136)
Expanded	-.0698	-.0039	-.0756	-.0202	-.0393	-.0167	-.0540	-.0153
	(.0358)	(.0332)	(.0124)	(.0108)	(.0414)	(.0407)	(.0171)	(.0161)
N	477		6,389		401		3,367	

Standard and expanded specifications for the CPS ORG samples are described in the note to Table 3. All models include year dummies. Standard errors are in parentheses. Change equations have ΔlnW as the dependent variable; the coefficients on Δln*B* are presented. Change equations from the March CPS files differ in specification from those in the CPS ORG because they include changes in region and public-sector status, but exclude changes in marital status, union status, and in federal, state, and local worker status. The DWS results are based on the January 1984, 1986, 1988, 1990, and 1992 CPS supplemental Displaced Worker Surveys. The sample consists of workers who were age twenty and older and who were displaced from a full-time, private-sector job because of a plant closing, slack work, or a position or shift that was eliminated. The sample was futher restricted to workers who were reemployed at the survey date in a full-time wage and salary job, and excludes those displaced from the construction industry. The dependent variable in the change equation is the difference between the log of current weekly earnings and the log of predisplacement weekly earnings. In addition to the usual change variables, panel estimates include dummies for year of displacement and survey year.

workers reporting a different occupation but the same industry. Thus, estimates of Θ are less affected by measurement error for industry movers than for industry stayers. In a preliminary analysis not shown, we confirmed that coefficients for occupation-only switchers are closer to zero than are coefficients for occupation and industry switchers. And in a separate analysis matching workers in the January 1986 and January 1987 CPS surveys (the latter contains a measure of years in current occupation that we treat as a "true" measure of switching), we confirmed that those recorded as both occupation and industry changers are far more likely to be "true" switchers than those recorded as changing only occupations, and that most "true" occupational switchers also switch industries.

Although our primary analysis is based on the very large CPS ORG panels, we supplement this analysis with evidence from the March CPS files for 1983–92 and the five CPS DWS surveys. Both of these data sets provide measures of $\Delta \ln B$ less affected by error from mismeasured occupational change. The DWS has the added advantage of measuring occupational change over a longer time period and recording primarily involuntary and exogenous changes.

The March CPS files record not only responses on current earnings, occupation, and other characteristics, but also on a respondent's occupation, industry, and class (private, public, self-employed) in the longest job held during the previous year; annual earnings, weeks worked, and hours worked per week the previous year; and state of residence the previous March. Occupation change is far less likely to be measured with error in the March files than in the CPS ORG panels. The CPS ORG panels rely on information from two interviews, one year apart, typically conducted by different individuals and possibly with different household members, and coded by different Census coders. The March files, by contrast, rely on information from a single interview with a single household member by a single interviewer and with a single occupation coder (except in the event of an occupation change).[14] An added advantage of the March files is that they include information on workers who have changed households or locations or who could not be matched from the ORG files from separate years. The disadvantages of the March files are that sample sizes are smaller; the wage rate in the previous year is calculated from earnings, weeks, and hours information that can be determined in part by jobs other than the longest held (random measurement error from this source should not bias coefficients since it is on the left-side of the wage-change equations); and information is not available that would allow construction of variables measuring changes in union and marital status.

A third panel data set is constructed using the January 1984, 1986, 1988, 1990, and 1992 CPS Displaced Workers Surveys (DWS). The DWS provide information on whether workers have lost or left a job during the past five years because of a plant closing, an employer going out of business, a layoff from which a worker was not recalled, or some other similar reason. As with the March surveys, it is relatively unlikely that there will be a false identification of switchers in the DWS. In addition, job and occupational change among displaced workers is more likely to be exogenous. The analysis could be limited further to only those affected by plant closings, since some layoffs may not be completely exogenous (for a similar argument and use of the DWS, see Gibbons and Katz, 1992); however, we do not do this owing to the small sample sizes of black workers affected by plant closings. Although overall sample sizes in the DWS are far smaller than in the other two data sets, the sample includes only job changers; hence the number of occupational switchers is of a magnitude similar to that in the March CPS files.

Table 7 presents estimates of Θ from the standard and expanded specifications of both wage-level and wage-change models based, alternatively, on the CPS ORG panels for 1983/4–1991/2, the March CPS files for 1983–92, and the five January CPS DWS files for 1984 through 1992. Wage-level estimates are based on the second year for each worker-year pair—1984–92 for the CPS ORG files, 1983–92 for the March CPS, and the five survey years for the DWS. Despite differences in samples and specification, the results from all three data sets support the quality sorting hypothesis. In virtually all cases, controlling for unmeasured person-specific effects through the estimation of change equations leads to coefficients of the racial composition variable much closer to zero. The change in estimates of Θ between the levels and change equations is particularly large for black males and white females, indicating that for these groups *B* serves as a particularly strong index of unmeasured skill. And among the large group of white males, estimates of Θ fall by more than half as one moves from the wage-level to the wage-change specifications. Similar evidence is found for black women in most cases, the exception being the expanded change model from the March CPS. The similarity of estimates of Θ between the standard and expanded wage-change models provides particularly strong support to variant three of our quality sorting model. Once one accounts for unmeasured person-specific skills, estimates of Θ are close to zero, regardless of whether occupational skill measures are explicitly measured. These results support the hypothesis that *B* serves primarily as a proxy for unmeasured worker skills, rather than the productivity of jobs independent of workers.

The similarity in estimates between the CPS ORG panels, where mismeasurement of occupational changes is a serious concern, and the retrospective March CPS files, where changes are measured far more accurately, provides reassurance that measurement error in the $\Delta \ln B$ variable does not seriously bias our estimates. The DWS results for black and white males are highly similar to those from the CPS ORG and March CPS panels, indicating that endogenous occupational change in the latter two data sets has not biased substantially our estimates. DWS results for women differ somewhat from those in the other two data sets, but standard errors are large. Indeed, Θ is not statistically significant for any demographic group in the DWS expanded change model.

The clear conclusion from the results presented in Table 7 is that the strong negative relationship between the wages of white and black workers and the proportion of black workers in an occupation can be accounted for entirely or almost entirely by measured and unmeasured worker and job skills. After controlling for job skills and estimation of wage-change models, little evidence of an effect of racial composition on wages remains. Absent controls for occupational skill measures and person-specific fixed effects, *B* is a negative and significant correlate of earnings. On that basis, its inclusion in wage-equation estimates can be justified. It is important, however, that *B* not be regarded as a causal determinant of wages but, rather, as a proxy for otherwise unmeasured worker and job skills.

Our findings also imply that estimates of wage discrimination should include controls for the skill composition of occupations or, absent such measures, a control for racial composition. Wage differentials attributable to differences in *B*, however, must not be included in the unexplained or discriminatory component of the standard wage differential decomposition. In results not shown, we find that *B* accounts for about 27 percent of the explained portion, and 15 percent of the total, racial wage gap among men; corresponding numbers among women are 40 and 23 percent. Addition of *B* to a standard specification (i.e., including occupation dummies), however, reduces unexplained wage differentials (i.e., measures of discrimination) by relatively modest amounts.[15] These unexplained wage differentials (our best measures of current labor market discrimination) reflect within-occupation wage differences as well as inter-occupational differences uncorrelated with racial composition and occupational skill measures. Our analysis provides no direct evidence on the extent to which intra-occupational wage differences not accounted for by measured characteristics are the result of unmeasured differences in worker skills correlated with individual race (as opposed to occupation racial density).

VII. CONCLUSIONS

Our study has examined the relationship between wage rates and the racial composition of jobs. We confirm that wages of black and white workers are significantly lower in occupations with high proportions of black workers, and higher in predominantly white occupations, after accounting for standard measured characteristics. We tested the thesis that quality sorting by race, whereby *B* is negatively correlated with measured and unmeasured labor quality, is an important explanation for the observed relationship between wages and racial composition. We argue that labor market discrimination is likely to lead to a sorting equilibrium in which higher-skilled black and white workers are sorted into higher-productivity jobs with low levels of *B*, and lower-skilled blacks and whites are sorted into jobs with relatively lower productivity and higher *B*.

We tested the quality sorting hypothesis by introducing into wage equations characteristics measuring occupational skill levels (e.g., required years of training, job tenure, computer usage), and by estimating wage-change equations that account for otherwise unmeasured individual-specific differences in productivity fixed over time. Wage-level equations were estimated using the January 1983 through December 1992 Current Population Survey Outgoing Rotation Group (CPS ORG) files. The longitudinal analysis used three data sets: a large panel constructed from the CPS ORG including worker-year pairs for 1983/4 through 1991/2; the March CPS files for 1983–92 containing current and retrospective information on wages and occupation; and the CPS Displaced Worker Surveys (DWS) between 1984 and 1992 containing retrospective information on displaced workers.

The findings in the study are rather clear-cut and provide strong support for the quality sorting hypothesis. Using wage-level analysis, coefficients on ln*B* are reduced sharply when characteristics measuring occupational skills are introduced into the wage equation. Wage-change equations accounting for person-specific skill differences, with or without controls for occupational characteristics, indicate little if any relationship between racial composition and wages. Workers switching into or out of occupations with different racial densities do not realize wage changes associated with the change in racial composition.

This study provides extensive evidence on what largely has been ignored in the literature on racial wage differences—wages for both white and black workers vary systematically with the racial composition of jobs.[16] An implication of our study is that *B* provides an important control for what are typically unmeasured worker quality and occupational skill differences. It is important, however, that racial composition be interpreted as such, rather than as a causal determinant of wages. These results are not without policy implications. If differences in worker skills are a driving force behind current racial employment and wage differentials, our results provide support for strong efforts at enhancing training for African-Americans, both in and out of school.[17] Such a policy emphasis has particular urgency given recent demand shifts in the economy that have favored skill-intensive and information-based occupations and produced a rising return to skills. If our quality sorting explanation is valid, then a narrowing of black-white skill differences will be necessary to weaken what is now a strong negative correlation between wages and the racial composition of jobs.

APPENDIX 1

Construction of the Longitudinal Sample from the CPS Outgoing Rotation Files

Households are included in the CPS for eight months—four months in the survey, followed by eight months out, followed by four months in. Outgoing rotation groups 4 and 8 (a quarter sample) are asked earnings supplement questions (earnings, hours, union status, etc.). Individuals potentially can be identified for the same month in consecutive years; that is, individuals in rotation 4 in year 1 can be matched to records in rotation 8 in year 2. The CPS contains household identification numbers (ID), but not reliable individual identifiers.

The longitudinal file is created in the following manner. Separate data files are created for males and females, and for pairs of years (rotation 4:1983 and rotation 8:1984, rotation 4:1984 and rotation 8:1985, etc.). Within each file, individuals are sorted as appropriate on the basis of ascending and descending household ID, year, and age. To be considered an acceptable matched pair, a rotation 8 individual has to be matched with a rotation 4 individual with identical household ID, identical survey month, and an age difference between 0 and 2 (since surveys can occur on different days of the month, age change need not equal 1). Several passes are necessary because a single household may contain more than one male or female pair. Checks are provided to insure that only unique matches are selected. For each rotation 8 individual, the search is made through all rotation 4 individuals with the same ID to make sure there is only one possible match; the file is resorted in reverse order and each selected rotation 4 individual is checked to insure a unique rotation 8 match. As uniquely matched pairs are identified they are removed from the work file. Incorrect changes in the variables marital status, veteran status, race, and education (e.g., a change in schooling other than 0 or 1, a change from married to never married, etc.) are used to delete "bad" observations in households where there are multiple observations and ages too close to separate matched pairs. Several passes at the data are made. In households where two pairs of individuals could be separated based on a 1 year but not the 0 to 2 year age change, a 1 year criterion is used. If a unique pair cannot be identified based on these criteria, they are not included in the data set (e.g., four observations with two identical pairs, or three individuals with two possible matches using the 0 to 2 age change criterion).

The match rate of earners in the longitudinal analysis is just under two-thirds. That is, two-thirds of rotation 4 workers in year 1 are found employed in year 2, or two-thirds of rotation 8 workers in year 2 are found employed in year 1. The principal reasons that matches cannot be made are if a household moves (thus changing household ID), if an individual moves out of a household, if an individual is not employed or fails to meet other sample selection criteria in one of the two years, or if the Census is unable to reinterview a household and/or receive information on the individual. The match rate here is somewhat lower than the rate of 68.8 percent for the 1987–88 CPS reported by Card (1992), who uses a broader-based sample and a less-stringent probabilistic matching algorithm obtained from the Bureau of Labor Statistics. Peracchi and Welch (1992) analyze attrition rates among matched March CPS files and conclude that age is the most important determinant of a successful match. Factors that lessen match probabilities are poor health, low schooling, and not a household head, while sex and race are unimportant match predictors following control for other factors.

The sample size for the 1983/4–1991/2 panel is 392,877, 24.4 percent of the size of the 1983–92 levels sample of 1,611,829. The difference in sample sizes between the cross-sectional and longitudinal analyses can be approximated as follows. Because the unit of observation in the panel is the pair of observations in adjacent years, the potential sample size is initially cut in half. Since we do not match the half of the 1983 sample that entered the CPS in 1982, or the half of the 1992 sample that exited the CPS in 1993, the potential sample is reduced by a further 10 percent (i.e., 90 percent of the full 1983–92 sample is in for two years). A 100 percent match rate of workers between adjacent years would produce a sample for each period half as large as the corresponding cross-sections. We achieve a match rate of about 63 percent. Finally, sample sizes are reduced further to roughly half the normal size for the 1984/5 panel and to one-quarter for 1985/6, owing to changes in CPS sampling that did not allow matching on the basis of household I.D.* Hence the combined multiplier relating the size of the longitudinal sample to the initial full sample is about .24; i.e., .50 x .90 x .63 x .86 = .24, where .86 is the ratio (9.0-7.75)/9.0, with the numerator representing the "loss" of panels for 1984/5 plus 1985/6.

*The CPS ran a test sample from July–September 1985 in order to implement new population weights. Rotation 4 households interviewed in July 1984 through September 1985 were not reinterviewed a year later in 1985 and 1986.

APPENDIX 2

Regression Coefficients, Expanded Wage Models, 1983–92 CPS ORG

Variables	Black Males		White Males		Black Females		White Females	
ln*B*	-.0293	(.0042)	-.0478	(.0010)	-.0176	(.0035)	-.0305	(.0011)
Individual Characteristics:								
Schooling	.0342	(.0007)	.0442	(.0002)	.0395	(.0007)	.0426	(.0003)
Experience	.0180	(.0004)	.0258	(.0001)	.0140	(.0003)	.0152	(.0001)
Exper/100	-.0248	(.0008)	-.0407	(.0003)	-.0197	(.0007)	-.0252	(.0003)
Union	.1615	(.0037)	.1581	(.0013)	.1371	(.0033)	.1513	(.0015)
Part-time	-.1480	(.0052)	-.1660	(.0019)	-.0968	(.0036)	-.0994	(.0012)
Married w/spouse	.0815	(.0040)	.1271	(.0013)	.0270	(.0033)	.0316	(.0014)
Other ever married	.0496	(.0051)	.0727	(.0019)	.0220	(.0036)	.0293	(.0016)
Hispanic	-.0828	(.0125)	-.1081	(.0019)	-.0203	(.0115)	-.0500	(.0021)
Federal	.0322	(.0076)	.0132	(.0030)	.0397	(.0063)	.0238	(.0034)
State	.0012	(.0084)	-.0651	(.0027)	-.0170	(.0059)	-.0423	(.0026)
Local	-.0203	(.0060)	-.0839	(.0023)	-.0202	(.0049)	-.0717	(.0021)
Large metro area	.0967	(.0035)	.1311	(.0011)	.1086	(.0030)	.1320	(.0011)
Job Characteristics:								
DOT-GED	.0380	(.0050)	.0014	(.0015)	.0382	(.0043)	.0384	(.0016)
DOT-SVP	.0167	(.0024)	.0248	(.0006)	.0125	(.0022)	.0130	(.0007)
OCC-Training	.1708	(.0173)	.1093	(.0050)	.1274	(.0153)	.3471	(.0055)
OCC-Computer	.1633	(.0124)	.2173	(.0037)	.2020	(.0094)	.1075	(.0035)
OCC-Job Tenure	.0014	(.0010)	-.0013	(.0003)	.0168	(.0010)	.0062	(.0004)
OCC-Part-time	-.1295	(.0168)	-.1883	(.0059)	.0419	(.0131)	-.0338	(.0053)
DOT-Enviroment	-.0017	(.0050)	.0235	(.0016)	-.0540	(.0070)	.0016	(.0032)
DOT-Hazards	-.0208	(.0105)	.0238	(.0032)	.1805	(.0132)	.0913	(.0052)
DOT-Physical	.0265	(.0035)	.0031	(.0010)	.0177	(.0031)	.0506	(.0010)
DOT-Strength	-.0192	(.0045)	-.0224	(.0016)	.0197	(.0043)	.0154	(.0016)
OCC-Female	-.0719	(.0094)	-.0705	(.0033)	-.0715	(.0077)	-.1161	(.0030)
IND-Big Firm	.1688	(.0092)	.1842	(.0029)	.1799	(.0074)	.1776	(.0027)
IND-Union	.1622	(.0137)	.0538	(.0045)	.0521	(.0122)	-.0457	(.0048)
Region (8)	yes		yes		yes		yes	
Industry (13)	yes		yes		yes		yes	
Occupation (5)	yes		yes		yes		yes	
Year (9)	yes		yes		yes		yes	
R^2	.4666		.5043		.4979		.4704	
N	72,388		777,036		86,415		680,990	

Standard errors are in parentheses. The omitted reference group is full-time, nonunion, non-Hispanic, never married, private-sector worker in the northeast, professional or managerial occupation, agricultural sector, in 1983. Tables 2 and 3 provide definitions for all variables.

Endnotes

[1]See Hirsch and Schumacher for a discussion of previous literature. The relationship between wages and racial composition is examined empirically by Sorensen (1989) and England (1992), although it is not the focus of their work, and by Ragan and Tremblay (1988) in a study of employee discrimination. Earlier papers by Chiswick (1973) and Reich (1978) have a different emphasis, but provide discussion and evidence related to racial composition and the income distribution.

[2]Other variants of a quality sorting model can be developed. A search model might suggest that *voluntary* job switches will result in increased wages regardless of whether *B* increases or decreases. Such a pattern should be particularly evident among young workers, for whom job shopping is most important. And one should not observe such a pattern where job switching is not voluntary (we examine exogenous job switching using the Displaced Worker Surveys). Because the simple search model cannot account for our basic pattern of results, we do not pursue this approach. A more complex search model (e.g., Jovanovic, 1979), in which the productivity of heterogeneous workers only becomes known to employers over time (i.e., with experience), does not lead to unambiguous implications about the relationship of wages and *B*.

[3]Although not a focus of this study, it is important to note that labor market equilibria evident today have a historical basis. Racial discrimination and imperfect information about individual productivities interacted to lead to equilibria in which jobs with a high *B* have lower measured and unmeasured skills. Black workers brought to the labor market a significantly lower quantity and quality of human capital (Card and Krueger, 1992), faced substantial employment barriers (Higgs, 1977; Margo, 1990; Wright, 1986; Heckman and Payner, 1989), and were relegated to jobs typically providing low levels of job training. White workers in predominantly black labor markets were likely to be low skilled. For example, Wright (1986, p. 103) quotes a white planter from the late 1800s who preferred black labor because "the class of white men that offer for hire out there as a rule are a very sorry class of men." As discrimination lessened, labor market equilibria remained in which there was an inverse relationship between the productivity (or wages) and *B*. Gill (1989) provides more recent

evidence indicating larger differences for blacks than whites between desired and actual occupational attainment. This finding is consistent with some maintenance of racial barriers to occupational attainment.

[4]The Aigner-Cain model treats human capital endowments as predetermined. Lundberg and Startz (1983) show that statistical discrimination leads to a suboptimal level of human capital investment for more able members of low-productivity groups.

[5]The CPS ORG files are not provided as public use tapes by the Bureau of the Census or ICPSR. They are made available through the Data Services Group at the Bureau of Labor Statistics.

[6]Silber (1989) compares the Duncan index to alternative indices of segregation. The Duncan index and *B* can be sensitive to changes in occupational definition. Because the CPS occupation codes changed significantly beginning in 1983 (with the use of the 1980 rather than the 1970 Census of Population codes), numbers for 1973–82 and 1983–92 are not strictly comparable. There appears to be no significant break in the series between 1982 and 1983, however. Beginning in 1992, the CPS adopted the 1990 Census of Population codes. These changes were modest and we were able to construct time-consistent occupational categories for 1983–92 (six small occupational categories were merged into larger categories, reducing the number of potential occupations from 503 in 1983–91 to 497 in 1983–92). Time-consistent codes are used in all the regression analyses.

[7]GED is a 1–6 index of general educational development and SVP measures years required for proficiency in an occupation.

[8]In work not shown, we estimate models with *B* calculated separately for workers in the South and non-South (Hirsch and Schumacher define it within more-aggregated occupation-by-industry-by-region cells). Estimates using regionally disaggregated measures of *B* correspond more closely to the proportion of black workers at an individual's work site. But the explanatory power of such measures is no greater than when *B* is defined on a nationwide basis, and estimates of racial composition coefficients differ more substantially across sectors (e.g., South versus non-South) in what we believe is a misleading fashion. We conclude that the racial composition effect is statistically driven by

nationwide occupational differences in the proportion of employees who are black, consistent with the thesis that *B* serves primarily as an occupational skill index.

[9]Table 1 provides mean values of *B*. Increases of 10 percent in *B* for black and white workers correspond to changes of approximately 1.5 and 1 percentage points, respectively.

[10]We thank Andrew Weiss for suggesting this point.

[11]The nominal minimum wage was $3.35 from January 1983 through March 1990, $3.80 from April 1990 through March 1991, and $4.25 after April 1991. The percentage of workers excluded from each group is as follows: black males, 9.8 percent; white males, 5.1 percent; black females, 14.5 percent; and white females, 10.9 percent.

[12]Bound and Johnson (1992), Juhn, Murphy, and Pierce (1993), and Krueger (1993) provide evidence that changes in the wage structure are consistent with technological changes and demand shifts leading to increased return to observed and unobserved skills.

[13]Workers reporting a change in occupation only, or in occupation and industry, have changes in *B* that are lower than would occur if changes in occupation are randomly assigned. Thus, measurement error may not be too serious even when occupation is misclassified. This is in contrast to the frequently discussed case of mismeasurement of union status, where workers are assigned values of 0 or 1.

[14]We thank Jay Stewart of the BLS for informing us about occupational coding. He also notes that BLS analysts have concluded that retrospective surveys such as the March CPS may understate true occupational change.

[15]Addition of *B* to the wage equation reduces the unexplained differential for males by 12.2 percent and for females by 23.8 percent. Hirsch and Schumacher (1992) provide explicit evidence on *B* and wage-gap estimates.

[16]It is worth comparing the conclusions here on racial composition to those reached in studies of gender composition (e.g., Johnson and Solon, 1986; Sorensen, 1989). As is widely known, both men and women earn less in predominantly female occupations. Macpherson and Hirsch (forthcoming), however, conclude that from two-thirds to three-fourths of the gender composition effect can be

accounted for by occupational characteristics measuring training and job attachment and by unmeasured worker-specific differences in tastes and skills.

[17]Rivera-Batiz (1992) and O'Neill (1990) provide direct evidence that racial differences in ability account for some of the racial gap in employment and wages, respectively.

References

Arrow, Kenneth. 1973. "The Theory of Discrimination." Pp. 3–33 in *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees. Princeton: Princeton University Press.

Becker, Gary S. 1971. *The Economics of Discrimination,* 2nd ed. Chicago: University of Chicago.

Bergmann, Barbara R. 1971. "The Effect on White Incomes of Discrimination in Employment." *Journal of Political Economy* 79(2): 294–313.

Bound, John and Richard B. Freeman. 1992. "What Went Wrong? The Erosion of Relative Earnings and Employment among Young Black Men in the 1980s." *Quarterly Journal of Economics* 107(1): 201–32.

Bound, John and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *American Economic Review* 82(3): 371–92.

Bound, John and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* 9(1): 1–24.

Cain, Glen G. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey." Pp. 693–785 in *Handbook of Labor Economics*, ed. Orley Ashenfelter and Richard Layard. Amsterdam: North-Holland.

Card, David. 1992. "The Effect of Unions on the Distribution of Wages: Redistribution or Relabelling?" National Bureau of Economic Research, Working Paper no. 4195, October.

Card, David and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107(1): 151–200.

Card, David and Thomas Lemieux. 1993. "Wage Dispersion, Returns to Skill, and Black-White Wage Differentials." Working Paper no. 312, Industrial Relations Section, Princeton University, February.

Chiswick, Barry R. 1973. "Racial Discrimination in the Labor Market: A Test of Alternative Hypotheses." *Journal of Political Economy* 81(6): 1330–52.

Dickens, William T. and Brian A. Ross. 1984. "Consistent Estimation Using Data from More than One Sample." National Bureau of Economic Research, Technical Working Paper no. 33.

England, Paula. 1992. *Comparable Worth: Theories and Evidence*. New York: Aldine de Gruyter.

Freeman, Richard B. 1984. "Longitudinal Analyses of the Effects of Trade Unions." *Journal of Labor Economics* 2(1): 1–26.

Gibbons, Robert and Lawrence Katz. 1992. "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?" *Review of Economic Studies* 59(3): 515–35.

Gill, Andrew M. 1989. "The Role of Discrimination in Determining Occupational Structure." *Industrial and Labor Relations Review* 42(4): 610–23.

Heckman, James and Brook Payner. 1989. "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina." *American Economic Review* 79(1): 138–77.

Higgs, Robert. 1977. *Competition and Coercion: Blacks in the American Economy, 1865–1914*. New York: Cambridge University Press.

Hirsch, Barry T. and Edward J. Schumacher. 1992. "Labor Earnings, Discrimination, and the Racial Composition of Jobs." *Journal of Human Resources* 27(4): 602–28.

Hwang, Hae-shin, W. Robert Reed, and Carlton Hubbard. 1992. "Compensating Wage Differentials and Unobserved Productivity." *Journal of Political Economy* 100 (August): 835–58.

Johnson, George and Gary Solon. 1986. "Estimates of the Direct Effects of Comparable Worth Policy." *American Economic Review* 76(6): 1117–25.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87(5, pt. 1): 972–90.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101(3): 410–42.

Krueger, Alan B. 1993. "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989." *Quarterly Journal of Economics* 108(1): 33–60.

Lang, Kevin. 1986. "A Language Theory of Discrimination." *Quarterly Journal of Economics* 101(2): 363–82.

Lundberg, Shelly J. and Richard Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review* 73(3): 340–47.

Macpherson, David A. and Barry T. Hirsch. Forthcoming. "Wages and Gender Composition: Why Do Women's Jobs Pay Less?" *Journal of Labor Economics*.

Margo, Robert. 1990. *Race and Schooling in the South, 1880–1950*. Chicago: University of Chicago Press.

Mellow, Wesley and Hal Sider. 1983. "Accuracy of Response in Labor Market Surveys: Evidence and Implications." *Journal of Labor Economics* 1(4): 331–44.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." *Journal of Economic Perspectives* 4(4): 25–45.

Peracchi, Franco and Finis Welch. 1992. "How Representative Are Matched Cross Sections? Evidence from the Current Population Survey." C.V. Starr Center for Applied Economics, New York University, Economic Research Report no. 92-53, October.

Ragan, James F., Jr. and Carol Horton Tremblay. 1988. "Testing for Employee Discrimination by Race and Sex." *Journal of Human Resources* 23(1): 123–37.

Reich, Michael. 1978. "Who Benefits from Racism? The Distribution among Whites of Gains and Losses from Racial Inequality." *Journal of Human Resources* 13(4): 524–44.

Rivera-Batiz, Francisco L. 1992. "Quantitative Literacy and the Likelihood of Employment among Young Adults." *Journal of Human Resources* 27(2): 313–28.

Silber, Jacques G. 1989. "On the Measurement of Employment Segregation." *Economics Letters* 30(3): 237–43.

Sorensen, Elaine. 1989. "Measuring the Effect of Occupational Sex and Race Composition on Earnings." Pp. 49–69 in *Pay Equity: Empirical Inquiries*, eds. R. T. Michael, H. I. Hartmann, and B. O'Farrell. Washington, D.C.: National Academy Press.

Whatley, Warren C. 1990. "Getting a Foot in the Door: 'Learning,' State Dependence, and the Racial Integration of Firms." *Journal of Economic History* 50(1): 43–66.

Wright, Gavin. 1986. *Old South, New South: Revolutions in the Southern Economy Since the Civil War*. New York: Basic Books.