

contributions to economic analysis

**John C. HALTIWANGER, Julia I. LANE, James R. SPLETZER,
Jules J.M. THEEUWES and Kenneth R. TROSKE**
Editors

The Creation and Analysis of Employer-Employee Matched Data

Chapter 7

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database

Kimberly Bayard

University of Maryland

Judith Hellerstein

University of Maryland

David Neumark

Michigan State University

Kenneth Troske

University of Missouri—Columbia

Acknowledgments: We thank an anonymous referee for helpful comments. This research was supported by National Science Foundation grant SBR95-10876 through the NBER. The research in this paper was conducted while the authors were research associates with the Center for Economic Studies, U.S. Bureau of the Census. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Bureau of the Census or the Center for Economic Studies.

1. Introduction

Labor economists have long been occupied with explorations of the sources of wage differences by sex, race, and ethnicity. It is well known that wages earned by minorities and by females fall short of wages earned by white males after accounting for differences in standard human capital proxies and other variables for which measures are readily available in many microlevel data sets (schooling, age or experience, marital status, urban residence, region, etc.).

Aside from this general fact, an additional fact about racial and ethnic wage gaps is that they are considerably larger for men than for women. This is true in the raw data, as well as once we account for numerous determinants of wages or earn-

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

ings. For example, based on 1981 Current Population Survey (CPS) data, Cain (1986, table 13.4) reports that for all workers, black-white earnings ratios are 0.67 for men vs. 0.97 for women, while Hispanic-white earnings ratios are 0.72 for men and 0.90 for women. For full-time, year-round workers, black-white earnings ratios are 0.69 for men vs. 0.90 for women, while Hispanic-white earnings ratios are 0.72 for men and 0.87 for women.¹ As a second example, as we report later in this paper, in log wage regressions including controls for schooling, age, etc., based on the 1990 Census of Population, the estimated black-white (actually, black vs. non-black, non-Hispanic) earnings differential is -0.121 for men vs. -0.022 for women, while the Hispanic-white differential is -0.115 for men vs. -0.045 for women. Finally, in a cross-section of 1990 and 1991 observations from the National Longitudinal Survey of Youth (NLSY), in log wage regressions with no controls, Neal and Johnson (1996) report that black men earn 24.4 percent less than white men vs. an 18.5 percent shortfall for black women, while Hispanic men earn 11.3 percent less than white men vs. a 2.8 percent (and insignificant) shortfall for Hispanic women.

When Neal and Johnson control for AFQT (Armed Forces Qualifying Test) (interpreted as a catchall for premarket factors affecting wages), the black-white difference for men falls to -7.2 percent, while black women are estimated to earn 3.5 percent more than white women (an insignificant difference).² Thus, even if one believes the Neal and Johnson claim that premarket factors account for a sizable fraction of racial and ethnic wage differences, the fact that the difference in the black-white wage gap between men and women persists suggests that this difference is a "labor market" rather than a "premarket" phenomenon.

In our view the larger racial and ethnic wage gaps for men than for women are a rather striking set of stylized facts that have largely been ignored by researchers attempting to understand the sources of racial and ethnic wage differences. In this paper we examine more closely the possible sources of the differences in the wage gap, paying particular attention to whether these differences can be accounted for by differences between men and women in the patterns of racial and ethnic segregation.³ More generally, we believe that research on why racial and ethnic wage

¹ In more recent data for 1995 (reported in Altonji and Blank, forthcoming, table 1) the qualitative pattern of larger racial and ethnic gaps for men than for women is similar, although for women the racial and ethnic gaps have grown, and among men the Hispanic-white earnings ratio has fallen below the black-white earnings ratio; specifically, for full-time, year-round workers, black-white earnings ratios are 0.69 for men and 0.83 for women, while Hispanic-white earnings ratios are 0.58 for men and 0.75 for women.

² In the same specification, they report that the Hispanic-white wage difference for men falls to essentially zero, while Hispanic women earn 14.5 percent more than white women. While these estimates also preserve the large sex difference in the Hispanic-white difference, we are skeptical of the reliability of these estimates for Hispanics.

³ This might be viewed as a specific formulation of the hypothesis that each group of minorities or women suffers from discrimination relative to white males, while the differences in the effects of discrimination among these minorities or women are relatively minor; in particular, we look at the effects of segregation, which might well arise from discrimination.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

gaps differ by sex may ultimately prove useful in helping to understand the sources of these gaps. For example, if one believes that the observed wage differentials are the result of employer or customer discrimination (e.g., Darity and Mason, 1998), then one needs to try to explain why this discrimination is apparently more severe with respect to male employees. In general, if one believes that some other unmeasured characteristic is responsible for these wage differences, then evidence that this characteristic is more important for men than for women would bolster one's case.

This inquiry fits into an extensive literature on the role of segregation in generating racial, ethnic, and sex differences in labor markets, but takes this literature in a new direction. In the literature on sex differences in wages, considerable attention has been focused on the role of occupational segregation, in particular the concentration of women in low-wage occupations (e.g., Johnson and Solon, 1986; Sorensen, 1989a; Macpherson and Hirsch, 1995). However, relatively little attention has been paid to the role of occupational segregation in generating racial and ethnic differences in wages (for an exception, see Sorensen, 1989b), in part because occupational segregation between races and ethnic groups is much less pronounced than occupational segregation between the sexes (King, 1992; Watts, 1995).

Furthermore, even less attention has been paid to the role of segregation along other dimensions such as industry, employer, and job cell (occupation within employer). The main reason for the lack of such attention is that the data sets labor economists typically use to study wage differences are household data sets, which allow one to measure the percentage of female or black workers in an occupation or industry, but not the sex, race, or ethnic composition of firms, establishments, or jobs. Economists interested in studying these other dimensions of segregation have had to turn to other special data sources in which information on the work force is available or can be constructed. For example, Groshen (1991) uses data from the Bureau of Labor Statistics (BLS) Industry Wage Surveys, with which one can measure the percentage of female workers by establishment as well as by job cell. Blau (1977) studies BLS Area Wage Surveys, which cover clerical, professional, and technical occupations, and which allow estimation of the percentage of female workers along the same dimensions. Bayard, Hellerstein, Neumark, and Troske (1998a) construct a data set (called the New Worker-Establishment Characteristics Database, or NWECD) based on a match of employees to their establishments, and carry out an analysis of the roles of sex segregation by occupation, industry, establishment, and job cell, similar to Groshen's. While there are differences in the findings reported in these studies, all find that in addition to being concentrated in low-wage occupations, women are concentrated in low-wage establishments and low-wage job cells.⁴

⁴ The differences concern the relative importance of each of these dimensions of segregation, and the role of the individual's sex after accounting for segregation (effectively, the within-job-cell sex difference in wages). The results in Groshen and in Bayard et al. (1998a) are directly comparable and differ in that Groshen attributes a large portion of the sex gap in wages to occupational segregation and none to within-job-cell sex differences, whereas Bayard et al. find a smaller role for occupational segregation and a larger role for within-job-cell sex differences. Carrington and Troske (1998a) also document the concentration of women in low-wage plants in U.S. manufacturing.

Section 1. Analysis with Employer-Employee Matched Data Wage Differentials in Firms

In this paper, we use the NWECD to study the role of racial and ethnic segregation in generating wage differences among whites, blacks, and Hispanics. The NWECD is uniquely suited to this analysis, as the BLS Industry and Area Wage Surveys contain no information on race or ethnicity. Thus, to the best of our knowledge, this is the first paper that looks at segregation by race and ethnicity at the level of the establishment and job cell.⁵ We consider evidence on the effects of racial and ethnic segregation on wages and the extent to which racial and ethnic wage differences remain after controlling for segregation. Such evidence helps to assess whether equal pay policies are likely to reduce these wage differences (assuming that these remaining differences reflect discrimination).⁶ We are particularly interested in the question posed in the title of this paper, namely whether more severe racial and ethnic segregation among men can explain why racial and ethnic wage gaps are bigger among men than among women.

2. The Data

The NWECD is created from two data sources, the Sample Detail File (SDF), which contains all individual responses to the 1990 Decennial Census One-in-Six Long Form, and the 1990 Standard Statistical Establishment List (SSEL), which is an administrative database containing information for all business establishments operating in the United States in 1990. We construct the NWECD by using detailed location and industry information available in both data sets to match worker records in the SDF to employer records in the SSEL. In this section we briefly discuss the details of the matching process and review results from an analysis assessing the accuracy of the match and the representativeness of these matched data. A much more detailed description of the matching process and our analysis assessing the accuracy of the match and the representativeness of these data is contained in Bayard, Hellerstein, Neumark, and Troske (1998b) (hereafter BHNT). Interested readers are referred there for a more complete description of the construction of these data.⁷

⁵ Carrington and Troske (1998b) use the WECD, a version of the NWECD that covers manufacturing only, to look at the role of racial segregation across establishments. They find that in establishments in which blacks are concentrated, wages of white workers are relatively high, but also that the wage gap between black and white workers is relatively larger in these establishments. Thus, the overall impact of segregation by establishment on the black-white wage gap is unclear.

⁶ See Bayard et al. (1998a) for a thorough discussion of this issue in the context of sex differences in wages.

⁷ Both Bayard et al. (1998a) and Bayard et al. (1998b) are available for downloading from the Web site of the Department of Economics at the University of Missouri-Columbia: www.missouri.edu/~econwww/workpaper.html.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

The Matching Process

Households receiving the 1990 Decennial Census Long Form were asked to report the name and address of the employer in the previous week for each employed member of the household. In addition, respondents were asked for the name and a brief (one- or two-word) description of the type of business or industry of the most recent employer for all members of the household. Based on the responses to these questions, the Census Bureau assigned geographic and industry codes to each record in the data, and it is these codes that are available in the SDF. In addition to this information, the SDF contains the standard set of demographic characteristics collected on the long form of the Decennial Census. To construct the NWECD, we first selected records for the slightly more than 17 million long-form respondents who indicated they were employed in the previous week.

The SSEL is an annually updated list of all business establishments with one or more employees operating in the United States that the Census Bureau uses as a sampling frame for its various Economic Censuses and Surveys. As such, the SSEL contains the name and address of each establishment, geographic codes based on its location, and a four-digit SIC code. In addition, the SSEL contains data on the number of employees and total annual payroll for the establishment, a unique establishment identifier, as well as an identifier that allows the establishment to be linked to other establishments that are part of the same enterprise. To construct the NWECD, we selected the 5.6 million records from the 1990 SSEL.⁸

Matching workers to employers proceeded in four steps. First, we standardized the geographic and industry codes in the two data sets.⁹ Next, we selected all establishments that were unique in an industry-location cell. Third, all workers who indicated they worked in the same industry-location cell as a unique establishment were matched to the establishment. Finally, we eliminated all matches based on imputed data. The resulting data set is what we call the NWECD. The full version of this data set contains 1,056,653 workers matched to 153,291 establishments. Note that these numbers represent only about 6.1 percent of long-form respondents in the SDF and 5.9 percent of establishments in the SSEL, which is a potentially significant limitation of the NWECD. For reasons discussed fully in BHNT, we impose a number of restrictions on this sample prior to our analysis, leading to an analysis sample of 637,718 workers in 32,931 plants.¹⁰

⁸ We select only active establishments with positive payroll and employment.

⁹ Industry codes must also be standardized because the industry code in the SSEL is based on the Standard Industrial Classification (SIC) system, while the Census Bureau assigns three-digit Census Industry Classification (CIC) codes to the SDF data. Since the CIC codes are more aggregated than the SIC codes, we use a concordance table to convert SIC codes to CIC codes.

¹⁰ These restrictions include weekly hours between 30 and 65; at least 30 weeks of work in 1989; ages between 18 and 65; wages between \$2.50 and \$500; work for a private-sector employer; establishment employment of 25 or more; number of matched workers equal to at least 5 percent of SSEL employment; and a few other minor restrictions to flag bad data.

*Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms*

Assessing the Accuracy and Representativeness of the Matched Data

One of the main uses of these data is to construct estimates of characteristics of establishments' work forces (such as the skill of workers within an establishment, or in this particular paper, the percent black, etc.) using the worker data. Therefore, in evaluating these data, we would like to compare estimates of establishment characteristics based on worker data with estimates of the same characteristics based on establishment data. Unfortunately, the only information that is common in the worker and establishment data sets are worker earnings. As a result, part of the analysis in BHNT focuses on comparing estimates of worker earnings from the worker and establishment data. This comparison is made for all workers, and broken out by location in a metropolitan statistical area (MSA), by establishment size, and by one-digit industry. The results in BHNT suggest that with the possible exception of some of the smaller establishments, workers are being matched to the correct establishments. Estimates of average worker earnings based on the SSEL and SDF data are very similar and are positively and significantly correlated across the two data sources. In addition, there appears to be no systematic difference in the quality of the matched data across different industries (with the exception of the construction industry), cells nor by whether or not the establishment is located in an MSA.

In BHNT, to examine whether the NWECD data are representative of the underlying population of workers and establishments, we first compare the average size and cross-industry and cross-location (whether an establishment is or is not located within an MSA) distributions of establishments in the SSEL and establishments in the NWECD. Next, we compare the average earnings and cross-industry and cross-location distributions of workers in the SDF and workers in the NWECD. In addition, we compare the characteristics of workers in the SDF with the characteristics of workers in the NWECD. Finally, to assess whether the NWECD data can replicate well-established relationships between establishment and worker characteristics and wages, we present results from regressions of (log) worker wages on a standard set of worker characteristics as well as results from regressions of (log) establishment-level average annual earnings on a standard set of establishment characteristics for both the matched and full samples of workers and establishments.

The descriptive statistics from this analysis suggest that the NWECD data are not a representative sample of the underlying population of establishments or workers. Workers in the NWECD are less likely to be black and more likely to be married than workers in the SDF. In addition, NWECD workers are more likely to be laborers and to work in manufacturing and services. NWECD workers are also slightly older and more likely to have a high school degree, but less likely to have no high school education or to have a bachelor's or advanced degree. Workers in the NWECD also tend to work more weeks in the previous year, but have slightly lower earnings and hourly wages.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

However, it appears that the non-representativeness does not introduce much bias into estimates of the types of relationships we estimate in subsequent sections of this paper; regression estimates of equations for worker and establishment earnings (except for the intercepts) are very similar for the matched and full samples of workers and establishments. These results, coupled with the fact that the NWECD data set is the largest employer-employee matched data set currently in existence in the United States, suggest that it will be a valuable tool for analyzing a variety of labor market issues. In the remainder of this paper, we turn to evidence on the role of segregation in generating racial and ethnic wage gaps that differ by sex, an empirical application for which these data are uniquely well suited.

3. Methods

Our decompositions of racial and ethnic wage gaps are based on estimates of log wage regressions of the following form,

$$\ln(w) = \alpha + \beta_B B + \beta_H H + \gamma_B \text{OCC\%B} + \delta_B \text{IND\%B} + \lambda_B \text{EST\%B} + \theta_B \text{JOB\%B} + \gamma_H \text{OCC\%H} + \delta_H \text{IND\%H} + \lambda_H \text{EST\%H} + \theta_H \text{JOB\%H} + X\Phi + \varepsilon, \quad (1)$$

where w is the hourly wage, B is a dummy variable equal to 1 if the individual is black, and H is a dummy variable equal to 1 if the individual is Hispanic. The variables OCC\%B and OCC\%H are the percentages black and Hispanic in the individual's occupation (expressed as proportions), and similarly the variables IND\%B , IND\%H , EST\%B , EST\%H , JOB\%B , and JOB\%H are the percentages black and Hispanic in the individual's industry, establishment, and job cell. A vector of control variables is represented by X . These regressions are estimated separately by sex, as are the various percentages black and Hispanic. Note that we allow the effects of segregation to differ by race and Hispanic ethnicity.

With the estimated coefficients of equation (1) in hand, we decompose the difference in average log wages between blacks and whites (denoted w_B' and w_W') as follows:

$$w_B' - w_W' = \beta_B' + \gamma_B'(\text{OCC\%B}_B - \text{OCC\%B}_W) + \delta_B'(\text{IND\%B}_B - \text{IND\%B}_W) + \lambda_B'(\text{EST\%B}_B - \text{EST\%B}_W) + \theta_B'(\text{JOB\%B}_B - \text{JOB\%B}_W) + \gamma_H'(\text{OCC\%H}_B - \text{OCC\%H}_W) + \delta_H'(\text{IND\%H}_B - \text{IND\%H}_W) + \lambda_H'(\text{EST\%H}_B - \text{EST\%H}_W) + \theta_H'(\text{JOB\%H}_B - \text{JOB\%H}_W) + (X_B - X_W)\Phi', \quad (2)$$

where primes on the coefficients indicate estimates, and B and W subscripts on the variables indicate means for blacks and whites, respectively (and similarly for H subscripts below). In this decomposition, β_B' measures the black-white difference that remains after controlling for the variables in X , and for segregation by occupation, industry, establishment, and job cell. Since the inclusion of these segregation measures should account for the relationship between race and any excluded vari-

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

ables related to the job cell, β_B' is often referred to as the "within-job-cell" race difference in wages. The term $\gamma_B'(\text{OCC}\%B_B - \text{OCC}\%B_W)$ measures the extent to which the wages of black and white workers differ because of occupational segregation by race (with segregation of blacks into lower-wage occupations, as it turns out). Similarly, the terms involving δ_B' , λ_B' , and θ_B' capture wage differences attributable to industry, establishment, and job-cell segregation. The second set of terms—beginning with $\gamma_H'(\text{OCC}\%H_B - \text{OCC}\%H_W)$ and ending with $\theta_H'(\text{JOB}\%H_B - \text{JOB}\%H_W)$ —captures black-white wage differences attributable to the differential segregation of blacks and whites into occupations, industries, establishments, and job cells with different percentages Hispanic. To the extent that blacks and whites are in occupations, industries, etc., with similar percentages Hispanic, as turns out to be the case, these effects will be rather small.¹¹

We also construct a similar decomposition for Hispanic-white wage differences. With the estimated coefficients of equation (1) in hand, we decompose the difference in average log wages between Hispanics and whites (denoted w_H' and w_W') as follows:

$$\begin{aligned} w_H' - w_W' = & \beta_H' + \gamma_B'(\text{OCC}\%B_H - \text{OCC}\%B_W) + \delta_B'(\text{IND}\%B_H - \text{IND}\%B_W) & (3) \\ & + \lambda_B'(\text{EST}\%B_H - \text{EST}\%B_W) + \theta_B'(\text{JOB}\%B_H - \text{JOB}\%B_W) \\ & + \gamma_H'(\text{OCC}\%H_H - \text{OCC}\%H_W) + \delta_H'(\text{IND}\%H_H - \text{IND}\%H_W) \\ & + \lambda_H'(\text{EST}\%H_H - \text{EST}\%H_W) + \theta_H'(\text{JOB}\%H_H - \text{JOB}\%H_W) + \Phi'(X_H - X_W). \end{aligned}$$

In this case the second set of terms—beginning with $\gamma_H'(\text{OCC}\%H_H - \text{OCC}\%H_W)$ —captures the effects on the Hispanic-white wage differential of segregation of Hispanics into occupations, industries, establishments, and job cells with other Hispanics, and the first set of terms (involving γ_B' , δ_B' , λ_B' , and θ_B') captures the effects of segregation of Hispanics and whites into occupations, industries, etc., with different percentages black. β_H' measures the within-job-cell wage differential between Hispanics and whites.

The percentages black and Hispanic variables in equation (1) are all estimated directly from the data. The percentages black and Hispanic in the occupation and industry are estimated from the full SDF sample, so measurement error is likely to be minimal.¹² However, the percentages black and Hispanic in the plant and job cell are estimated by necessity from the matched data in the NWECD. On average, 19.37 workers are matched to a plant, so job-cell estimates, in particular, are often based on a small number of observations. Measurement error in these estimates therefore could be sizable, biasing the estimates of λ and θ towards zero (and presumably biasing the other coefficient estimates as well, although a priori the direc-

¹¹ This decomposition can be thought of as the traditional decomposition of Oaxaca (1973), imposing the restriction that the coefficients are the same for racial and ethnic groups.

¹² This also implies that the estimated average proportion black in workers' occupations need not match the sample proportion black. Indeed, the first number is higher because, as already noted, blacks tend to be underrepresented in the NWECD.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

tion of bias is unclear). One motivation for restricting attention to larger establishments (those with 25 or more workers) is to avoid very small cells.

While establishments are well defined, industries and occupations can be defined at a variety of levels of disaggregation. Since a question of primary concern is within- vs. across-job wage differences, we are interested in trying to use relatively narrow occupational classifications. Because we also look at establishment-occupation cells (i.e., job cells), however, and because we are looking at rather narrow racial and ethnic groups, if we use highly disaggregated occupations we can end up with very few observations in some job cells, particularly since we only have a sample of workers in each plant and consequently in each job cell. Therefore, we report evidence from specifications using two alternative levels of occupational disaggregation, beginning first with 13 Census occupations, and then using a considerably greater level of disaggregation involving 72 Census occupations.¹³ Because all workers in an establishment work in the same industry, and because the percent black and percent Hispanic in an industry are estimated using the full SDF, we face no constraint in disaggregating industries finely, and hence we always use the most detailed four-digit SIC codes. To preview the results, we find that the qualitative conclusions are not affected by the level of occupational detail.

We also report results in which we estimate β_B and β_H controlling for fixed occupation, industry, establishment, and job-cell effects, rather than controlling for the percent black and Hispanic in each of these categories; this amounts, of course, to putting in job-cell dummy variables, since these absorb occupation, industry,

¹³ These 13 occupational categories, the corresponding Census codes, and the number of subcategories making up the 72 disaggregated occupations, are as follows:

- (1) Managerial and Professional Specialty Occupations—Executive, Administrative, and Managerial Occupations, codes 3–37, 2 subcategories
- (2) Managerial and Professional Specialty Occupations—Professional Specialty Occupations codes 43–199, 9 subcategories
- (3) Technical, Sales, and Administrative Support Occupations—Technicians and Related Support Occupations, codes 203–235, 3 subcategories
- (4) Technical, Sales, and Administrative Support Occupations—Sales Occupations, codes 243–285, 3 subcategories
- (5) Technical, Sales, and Administrative Support Occupations—Administrative Support Occupations, codes 303–389, 10 subcategories
- (6) Service Occupations—Private Household Occupations, codes 403–407, 1 subcategory
- (7) Service Occupations—Protective Service Occupations, codes 413–424, 1 subcategory
- (8) Service Occupations—Service Occupations, Except Protective and Household, codes 433–469, 7 subcategories
- (9) Farming, Forestry, and Fishing Occupations, codes 473–499, 4 subcategories
- (10) Precision Production, Craft, and Repair Occupations, codes 503–699, 14 subcategories
- (11) Operators, Fabricators, and Laborers—Machine Operators, codes 703–799, 7 subcategories
- (12) Operators, Fabricators, and Laborers—Transportation and Material Moving Occupations, codes 803–859, 5 subcategories
- (13) Operators, Fabricators, and Laborers—Handlers, Equipment Cleaners, Helpers, and Laborers, codes 864–889, 4 subcategories

and establishment effects.¹⁴ In the absence of measurement error, assuming that we have specified the wage regression correctly, we would not expect estimates of β_B and β_H obtained using these fixed effects to differ much from estimates using the percent black and Hispanic variables, since the correlation of B and H with occupation, industry, establishment, and job-cell characteristics should be captured by the percent black and percent Hispanic variables.¹⁵ However, using job-cell dummy variables avoids the measurement error inherent in the percent black and percent Hispanic variables, and therefore should provide more reliable estimates of the within-job-cell racial and ethnic differences in wages (β_B and β_H), even when cell sizes are small; because the sample is one of individuals, job cells with more observations implicitly receive more weight. This specification is also useful because it accounts for the correlation between observations in the same establishment (and job cell). In contrast, when we run ordinary least squares (OLS) for the specifications using the percent black and percent Hispanic variables, the standard errors could be biased downward because of within-establishment or within-job-cell correlations in the error.

4. Results

Descriptive Statistics

Table 1 reports descriptive statistics for black, Hispanic, and white male and female workers in the NWECD.¹⁶ Panel A reports information on wages and on worker and establishment characteristics. The average log hourly earnings data re-

¹⁴ They also absorb the controls for region and MSA.

¹⁵ Because we estimate all of our wage equations separately by sex, to ensure that this is the case we construct the percent black and percent Hispanic variables by sex. When we defined the percent black and percent Hispanic variables over men and women together, the results for Hispanics were very similar to those reported below, while there was less evidence that inclusion of the segregation variables accounted for much of the black-white wage gap among men.

¹⁶ To code race and ethnicity, we began with the question from the Decennial Census asking "Is this person of Spanish/Hispanic origin?" and then asking respondents to indicate specific ethnicity (e.g., Mexican, Cuban, other). We coded the individual as Hispanic if the answer to the "Spanish/Hispanic" question was yes and the person was not black. Additionally, we coded the worker as Hispanic if he or she listed a Latin American race code under the separate "race" question (and also was not black). The "race" question asks respondents to indicate whether they self-identify with one of seven race groups: white, black, Indian, Eskimo, Aleut, Asian or Pacific Islander (several choices), or "other" (in which case they are then asked to indicate the race with which they identify). We coded workers as black if they met one of two conditions: they picked "black" on the race question, or they picked "other" and indicated a race that falls into a "black" category (e.g., African American, Afro-American). Workers could not be coded as both black and Hispanic in our sample. For example, if a worker answered the "race" question as "other, Cuban" then the worker was coded as Hispanic. But if the worker answered the "race" question as "black" but indicated Hispanic-Cuban ethnicity, then the worker was coded as black.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

flect the stylized fact with which this paper began: Racial and ethnic differences are considerably larger for men (-0.23 for blacks and -0.24 for Hispanics) than for women (-0.16 for blacks and -0.13 for Hispanics). Not surprisingly, whites are more likely to be married, to have fewer children, and to have higher educational degrees, all of which are associated with higher wages. Furthermore, the education differences are a bit sharper among men than among women (for education, look at the proportions with no high school degree, a bachelor's degree, or an advanced degree), which may partly explain the larger raw racial and ethnic wage differences among men compared with women. The table next reports information on English language fluency and citizenship. Trejo (1997) finds that English language deficiencies are an important source of lower earnings for Mexican-Americans. We want to include these controls in the wage regressions because they are likely to reflect human capital differences (or more generally to be related to productivity), although we recognize the possibility that there is discrimination based on differences in language or citizenship. The descriptive statistics reveal a rather high proportion of Hispanic women and an even higher proportion Hispanic men who do not report speaking English "very well."¹⁷ Similarly, not surprisingly, high proportions of Hispanic men and women (the latter to a lesser degree) are not U.S. citizens, which may independently affect their labor market outcomes whether or not they are legally permitted to work. Finally, the last rows of this panel of the table indicate that Hispanics work in smaller establishments than do whites, while blacks work in larger establishments.

Panel B of table 1 reports descriptive statistics on segregation by race and ethnicity. This panel reports the mean percentage black and Hispanic by occupation, industry, establishment, and job cell, as in equations (1) and (2), for black, Hispanic, and white workers.¹⁸ The occupation figures are reported for both the more and less aggregated classifications that we use. In addition, to provide more descriptive information on the patterns of segregation, some disaggregation by occupation, industry, and establishment-size categories is reported.

The figures reveal that there is relatively little racial or ethnic segregation along occupation or industry lines. For example, in the first three columns of the top row, the percentage black in the occupation is not much higher for blacks than for Hispanics or whites. This is also true for the three broad categories of occupations and the two broad categories of industries covered in the table. In contrast, segregation

¹⁷ The language disadvantage faced by Hispanics implies that what we identify as the effects of segregation by ethnicity may reflect primarily segregation by English fluency, which might suggest that such segregation is more likely to fade with time, and would also suggest an emphasis on language skills to close the wage gap (see also Trejo, 1997). Unfortunately, we feel that the cell sizes available in the NWECD are too small to analyze segregation by Hispanic ethnicity and language skills. However, in future research we hope to obtain a much more extensive worker-establishment match that will permit such an analysis.

¹⁸ The percentage variables reported in this and subsequent tables, and used in the regressions, are always divided by 100.

Table 1. Sample Descriptive Statistics by Demographic Group for NWECD

	Females			Males		
	Black (1)	Hispanic (2)	White (3)	Black (4)	Hispanic (5)	White (6)
A. Wages and Worker and Establishment Characteristics						
Log hourly wages	2.010 (0.504)	2.038 (0.499)	2.167 (0.489)	2.315 (0.529)	2.306 (0.533)	2.547 (0.509)
Age	38.711 (10.513)	37.446 (10.727)	39.557 (11.029)	38.908 (10.740)	37.183 (11.141)	39.892 (10.963)
Number of children	2.265 (1.943)	2.104 (1.854)	1.793 (1.574)	—	—	—
Ever married	0.794	0.818	0.860	0.789	0.817	0.867
No high school degree	0.218	0.312	0.108	0.248	0.404	0.127
High school degree	0.388	0.294	0.358	0.398	0.273	0.381
Some college	0.202	0.183	0.189	0.196	0.164	0.192
Associate's degree	0.077	0.090	0.126	0.054	0.054	0.073
Bachelor's degree	0.073	0.078	0.136	0.065	0.060	0.134
Advanced degree	0.042	0.043	0.082	0.038	0.046	0.093
Speaks English:						
Very well	0.988	0.726	0.987	0.988	0.646	0.988
Well	0.007	0.153	0.009	0.008	0.196	0.008
Not well	0.004	0.089	0.004	0.004	0.123	0.003
Not at all	0.0004	0.032	0.0003	0.0004	0.034	0.0002

Table 1. Sample Descriptive Statistics by Demographic Group for NWECD (Continued)

	Females			Males		
	Black (1)	Hispanic (2)	White (3)	Black (4)	Hispanic (5)	White (6)
Citizenship status:						
By birth in U.S.	0.969	0.626	0.969	0.968	0.537	0.973
By birth in U.S. Territory	0.001	0.056	0.00002	0.001	0.064	0.00003
By naturalization	0.014	0.139	0.018	0.012	0.137	0.016
Not a citizen	0.016	0.179	0.012	0.019	0.261	0.011
MSA	0.536	0.713	0.544	0.601	0.741	0.591
Establishment size:						
25-49	0.051	0.085	0.067	0.079	0.132	0.097
50-99	0.098	0.118	0.110	0.105	0.143	0.116
100-249	0.195	0.228	0.217	0.183	0.245	0.192
250-499	0.199	0.197	0.197	0.168	0.137	0.165
500-999	0.196	0.165	0.175	0.168	0.119	0.156
1000+	0.260	0.208	0.233	0.298	0.223	0.274
B. Segregation Measures						
% black in occupation (13)	0.110	0.101	0.093	0.083	0.078	0.068
% black in occupation (72)	0.124	0.111	0.097	0.089	0.082	0.069
Blue-collar occupations	0.078	0.075	0.073	0.058	0.054	0.046
White-collar occupations	0.176	0.177	0.172	0.155	0.149	0.151
Service occupations	0.144	0.136	0.133	0.088	0.083	0.077
% black in industry	0.125	0.111	0.110	0.087	0.079	0.075
Goods-producing industries	0.134	0.108	0.105	0.085	0.073	0.070
Non-goods-producing industries	0.118	0.114	0.113	0.093	0.091	0.083

Table 1. Sample Descriptive Statistics by Demographic Group for NWECD (Continued)

	Females			Males		
	Black (1)	Hispanic (2)	White (3)	Black (4)	Hispanic (5)	White (6)
% black in establishment	0.350	0.058	0.058	0.271	0.045	0.047
50+ employees	0.342	0.060	0.060	0.258	0.048	0.049
25-49 employees	0.499	0.034	0.034	0.422	0.028	0.030
% black in job cell (13)	0.487	0.052	0.044	0.462	0.038	0.034
% black in job cell (72)	0.584	0.041	0.036	0.576	0.030	0.027
% Hispanic in occupation (13)	0.068	0.063	0.056	0.083	0.081	0.070
% Hispanic in occupation (72)	0.070	0.066	0.055	0.084	0.083	0.068
Blue-collar occupations	0.043	0.044	0.040	0.050	0.048	0.042
White-collar occupations	0.073	0.081	0.070	0.108	0.113	0.107
Service occupations	0.097	0.098	0.094	0.091	0.093	0.083
% Hispanic in industry	0.056	0.059	0.052	0.061	0.070	0.058
Goods-producing industries	0.071	0.079	0.067	0.063	0.075	0.060
Non-goods-producing industries	0.043	0.045	0.043	0.057	0.059	0.053
% Hispanic in establishment	0.022	0.338	0.022	0.028	0.353	0.025
50+ employees	0.022	0.316	0.022	0.028	0.327	0.025
25-49 employees	0.020	0.568	0.017	0.030	0.521	0.025
% Hispanic in job cell (13)	0.020	0.476	0.017	0.024	0.543	0.017
% Hispanic in job cell (72)	0.016	0.591	0.013	0.019	0.649	0.013
N	24,525	9,105	265,047	19,927	12,380	306,734

Note: Standard deviations are reported in parentheses for continuous variables. Number of children refers to number of children ever born, this is asked of women in the Census, and is set to zero for men. "White" refers to non-black and non-Hispanic. MSA = metropolitan statistical area.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

by establishment and job cell is severe. For example, for women the average percent black in the establishment of employment is 0.350 for blacks, but only 0.058 for Hispanics and whites. In general, racial and ethnic segregation by establishment is lower in establishments with 50 or more employees, compared to those with 25–49 employees, perhaps owing to affirmative action pressures that should be more operative for the former group.

Table 2 reports some baseline OLS regressions describing the multivariate relationships among the variables listed in table 1. Columns (1) and (5) report the raw racial and ethnic differences from regressions with no controls. In columns (2) and (6) we add individual-level controls for age, children, marital status, and education, as well as region of the country and residence in an MSA. For women, the black-white wage differential falls to -0.02 , while the Hispanic-white differential falls by nearly two-thirds. For men, both differentials fall by about one-half. In columns (3) and (7) we add controls for English language fluency and citizenship. The estimated Hispanic-white wage differential falls by about one-half for both women (to -0.023) and men (to -0.068). Still, the gap remains considerably larger for men. In columns (4) and (8) we find that adding controls for establishment size and industry has relatively small effects on the estimated racial and ethnic wage differences, with the exception of the Hispanic-white differential for males, which falls to -0.051 . These establishment-level controls may to some extent be related to unobserved human capital, calling for their inclusion along with the other human capital controls. On the other hand, to the extent that these controls solely reflect establishment-level characteristics, they may “overcontrol” for establishment-level differences, because they may capture dimensions of racial and ethnic segregation. As a consequence, we omit them in the decompositions that follow, although they are of course absorbed in the estimation using establishment fixed effects. Regardless, we see that upon inclusion of either set of control variables, racial and ethnic wage differences remain considerably larger for men than for women.

Next, we turn to estimates incorporating information on racial and ethnic segregation by occupation, industry, establishment, and job cell, both to better understand the sources of racial and ethnic differences in wages, and in particular to see whether greater segregation, or greater effects of segregation, explain the sharper black-white and Hispanic-white wage differentials among men.

Estimates of the Effects of Racial and Ethnic Segregation and Decompositions of Wage Differentials

Table 3 reports results of wage regression estimations using the relatively more aggregated 13 occupation categories. The first five columns report results for women, and the second five report results for men. In column (1), we report estimates from a specification that adds the percent black and percent Hispanic variables to the individual-level controls included in column (3) of table 2. Similarly, in column (6) we report estimates from a specification that adds the percent black and percent

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

Table 2. Log Hourly Wage Regressions

	Females				Males			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	-0.157 (0.003)	-0.022 (0.003)	-0.023 (0.003)	-0.037 (0.003)	-0.232 (0.004)	-0.121 (0.003)	-0.122 (0.003)	-0.114 (0.003)
Hispanic	-0.129 (0.005)	-0.045 (0.004)	-0.023 (0.005)	-0.025 (0.004)	-0.241 (0.005)	-0.115 (0.004)	-0.068 (0.004)	-0.051 (0.004)
Age	—	0.048 (0.001)	0.048 (0.0001)	0.041 (0.005)	—	0.065 (0.0005)	0.065 (0.0005)	0.053 (0.0005)
Age ² /100	—	-0.048 (0.001)	-0.048 (0.001)	-0.039 (0.001)	—	-0.063 (0.001)	-0.063 (0.001)	-0.049 (0.001)
Number of children	—	-0.028 (0.002)	-0.028 (0.002)	-0.015 (0.002)	—	—	—	—
(Age/10) x number of children	—	0.0002 (0.0005)	0.0001 (0.0005)	-0.001 (0.0005)	—	—	—	—
Ever married	—	0.063 (0.002)	0.062 (0.002)	0.050 (0.002)	—	0.196 (0.002)	0.195 (0.002)	0.157 (0.002)
High school degree	—	0.115 (0.002)	0.110 (0.002)	0.088 (0.002)	—	0.180 (0.002)	0.172 (0.002)	0.13 (0.002)1
Some college	—	0.233 (0.003)	0.228 (0.003)	0.193 (0.003)	—	0.254 (0.003)	0.246 (0.002)	0.207 (0.003)
Associate's degree	—	0.451 (0.003)	0.446 (0.003)	0.408 (0.003)	—	0.297 (0.003)	0.289 (0.003)	0.255 (0.003)
Bachelor's degree	—	0.583 (0.003)	0.578 (0.003)	0.552 (0.003)	—	0.481 (0.003)	0.473 (0.003)	0.470 (0.003)

Table 2. Log Hourly Wage Regressions (Continued)

	Females				Males			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advanced degree	—	0.732 (0.003)	0.727 (0.003)	0.740 (0.004)	—	0.593 (0.003)	0.585 (0.003)	0.709 (0.003)
Speaks English very well	—	—	0.189 (0.021)	0.166 (0.020)	—	—	0.302 (0.019)	0.251 (0.018)
Speaks English well	—	—	0.136 (0.022)	0.119 (0.021)	—	—	0.218 (0.020)	0.185 (0.018)
Speaks English not well	—	—	0.088 (0.023)	0.078 (0.021)	—	—	0.129 (0.020)	0.113 (0.019)
Citizen by birth in U.S.	—	—	0.041 (0.006)	0.027 (0.006)	—	—	0.041 (0.006)	0.020 (0.005)
Citizen by birth in U.S. territory	—	—	0.045 (0.017)	0.031 (0.016)	—	—	0.019 (0.015)	-0.009 (0.014)
Citizen by nationalization	—	—	0.084 (0.008)	0.061 (0.007)	—	—	0.113 (0.007)	0.076 (0.007)
MSA	—	0.135 (0.002)	0.135 (0.002)	0.093 (0.002)	—	0.133 (0.002)	0.133 (0.002)	0.087 (0.002)
Nine region controls included:	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Size controls included:	No	No	No	Yes	No	No	No	Yes
4-digit industry controls included:	No	No	No	Yes	No	No	No	Yes
N	298,677	298,677	298,677	298,677	339,041	339,041	339,041	339,041
R ²	0.009	0.352	0.353	0.441	0.018	0.351	0.353	0.458

Note: Standard errors of regression estimates are reported in parentheses. The omitted category for English fluency is "does not speak English," and the omitted category for citizenship is "not a citizen." MSA = metropolitan statistical area.

Table 3. Estimated Log Wage Differentials by Race and Ethnicity, and Percent Black and Hispanic in Occupation, Industry, Establishment, and Job Cell, Based on 13 Occupations

	Females					Males				
	Regression Estimates (1)	Mean Difference Black-White (2)	Contribution to Black-White Wage Gap (3)	Mean Difference, Hispanic-White (4)	Contribution to Hispanic-White Wage Gap (5)	Regression Estimates (6)	Mean Difference Black-White (7)	Contribution to Black-White Wage Gap (8)	Mean Difference Hispanic-White (9)	Contribution to Hispanic-White Wage Gap (10)
Black	-0.012 (0.004)	1.000	-0.012	0.000	0.000	-0.073 (0.004)	1.000	-0.073	0.000	0.000
Hispanic	-0.016 (0.006)	0.000	0.000	1.000	-0.016	-0.029 (0.006)	0.000	0.000	1.000	-0.028
% black in occupation	-1.012 (0.049)	0.018	-0.018	0.009	-0.009	-2.422 (0.041)	0.015	-0.036	0.010	-0.024
% black in industry	0.092 (0.022)	0.015	0.001	0.001	0.000	-0.505 (0.023)	0.013	-0.006	0.004	-0.002
% black in establishment	0.214 (0.009)	0.292	0.063	-0.000	-0.000	0.154 (0.010)	0.224	0.035	-0.002	-0.000
% black in job cell	-0.114 (0.008)	0.443	-0.050	0.008	-0.001	-0.093 (0.008)	0.429	-0.040	0.004	-0.000
Segregation by race	—	—	-0.004	—	-0.010	—	—	-0.047	—	-0.026
% Hispanic in occupation	-0.851 (0.071)	0.012	-0.011	0.007	-0.006	-0.928 (0.045)	0.014	-0.013	0.011	-0.010
% Hispanic in industry	-0.421 (0.033)	0.004	-0.002	0.007	-0.003	-1.887 (0.025)	0.003	-0.006	0.012	-0.023

Table 3. Estimated Log Wage Differentials by Race and Ethnicity, and Percent Black and Hispanic in Occupation, Industry, Establishment, and Job Cell, Based on 13 Occupations (Continued)

	Females					Males				
	Regression Estimates (1)	Mean Difference Black-White (2)	Contribution to Black-White Wage Gap (3)	Mean Difference, Hispanic-White (4)	Contribution to Hispanic-White Wage Gap (5)	Regression Estimates (6)	Mean Difference Black-White (7)	Contribution to Black-White Wage Gap (8)	Mean Difference Hispanic-White (9)	Contribution to Hispanic-White Wage Gap (10)
% Hispanic in establishment	0.169 (0.014)	0.001	0.000	0.316	0.053	0.061 (0.011)	0.003	0.000	0.328	0.020
% Hispanic in job cell	-0.124 (0.012)	0.004	-0.000	0.459	-0.057	-0.091 (0.010)	0.006	-0.001	0.526	-0.048
Segregation by Hispanic ethnicity	—	—	-0.013	—	-0.013	—	—	-0.020	—	-0.061
R ²	0.363	—	—	—	—	0.394	—	—	—	—

Note: Standard errors of regression estimates are reported in parentheses. The control variables included correspond to those in columns (3) and (7) in table 2. See notes to tables 1 and 2. The percentages black and Hispanic are computed for men and women separately.

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

Hispanic variables to the individual-level controls included in column (7) of table 2. We report the estimated coefficients of the black and Hispanic dummy variables as well as each of the percent black and percent Hispanic variables.

Turning first to the within-job-cell racial and ethnic wage gaps, we see that the black-white wage gap for women—which was small (-0.023) to begin with in table 2—becomes slightly smaller (-0.012) once we control for the effects of segregation. The Hispanic-white wage gap for women shrinks by a similar amount, from -0.023 to -0.016 . In contrast, the black-white wage gap for men shrinks from -0.122 to -0.073 , while the Hispanic-white wage gap for men shrinks from -0.068 to -0.029 . Therefore, the sex difference in the Hispanic-white wage gap is largely eliminated once we account for segregation; the difference between the Hispanic-white differential for men and women falls from 0.045 to 0.013 . In contrast, although segregation explains part of the larger black-white wage differential for men, the black-white differential still remains substantially larger for men; the sex gap in this differential is 0.099 ($0.122-0.023$) in table 2, and 0.061 ($0.073-0.012$) in table 3.

Looking at the effects of racial and ethnic segregation for women, we see why the within-job-cell racial and ethnic wage gaps for women shrink a bit once we account for segregation. The estimates in column (1) indicate that working in an occupation, industry (for Hispanics), or job cell with a higher percent black or Hispanic is associated with significantly lower wages. In contrast, working in an establishment with a higher percent black or Hispanic is associated with significantly higher wages (or equivalently, wages are higher in establishments with higher percentages black and Hispanic).

The estimated negative effects for women of a high percentage black or Hispanic appear particularly large for occupation. As noted earlier, however, there is considerably less segregation along these lines than along the lines of establishment or job cell. This is apparent from columns (2) and (4), which report the mean differences in the percentage black or Hispanic between black and white workers (in column (2)), and between Hispanic and white workers (in column (4)). For example, the entries in rows 3 through 6 of column (2) report the mean differences in the percentage black between black and white workers. These differences are small for occupation (0.018) and industry (0.015).¹⁹ However, segregation by establishment (0.292) and job cell (0.443) is much more severe. Thus, the coefficient estimate on the percent black in the occupation (-1.012), for example, seems rather large, but does not contribute that much to the lower wages of black women. In contrast, the smaller coefficient estimate on percent black in the job cell (-0.114), for example, is applied to a much larger difference. The findings for Hispanic women, in rows 8 through 11 of column (4), suggest similar patterns of segregation to those by race, with rather severe segregation by establishment and job cell, but

¹⁹ In contrast, in BHNT, we report much sharper sex segregation, with, for example, a mean difference by occupation of 0.17 .

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

not industry and occupation. In contrast, the numbers at the bottom of column (2) and the top of column (4) indicate that blacks are not particularly concentrated in occupations, industries, establishments, or job cells with high percentages Hispanic, or vice versa.

The mean differences in columns (2) and (4) are used along with the estimates in column (1) to decompose the wage differentials, as reported in columns (3) and (5). In these columns, we also report the combined effects of segregation by race and by ethnicity. The numbers reveal that segregation of black women by race lowers the wages of black women by 0.4 percent, and segregation of Hispanic women by ethnicity lowers the wages of Hispanic women by 1.3 percent. For both groups, the negative segregation effect stems primarily from job-cell segregation, i.e., from the segregation of black or Hispanic women into particular jobs within establishments.

Columns (6) through (10) report results of similar estimations and computations for men. Looking at the effects of segregation shows, correspondingly, that segregation by ethnicity reduces wages of Hispanic males by more than segregation by race reduces wages of black males. Moreover, segregation by race and ethnicity lowers wages of black and Hispanic men by considerably more than black and Hispanic women. Overall, the stronger negative effect of segregation for Hispanic men is summarized in the entry labeled "Segregation by Hispanic ethnicity," which indicates that such segregation lowers wages of Hispanic males by 6.1 percent. In contrast, racial segregation lowers wages of black men by 4.7 percent, while the corresponding numbers for black and Hispanic women are only 0.4 percent and 1.3 percent, respectively.

Note that for both men and women, ethnic segregation appears somewhat more severe than racial segregation, while among the four groups, segregation is most severe among Hispanic men. For example, for men the mean difference between Hispanics and whites in the proportion Hispanic in the job cell is 0.526, vs. a mean difference of 0.429 between blacks and whites in the proportion black. The corresponding numbers for women are 0.459 and 0.443. Turning instead to the estimated coefficients of the segregation variables, the negative effect of industry segregation is stronger for Hispanic males than for the other three groups, while for black men the negative effect of occupational segregation is particularly strong. Also, Hispanic males work in higher paying establishments to a lesser extent than black and Hispanic females and black males. For Hispanic males, ethnic segregation by establishment raises wages by only 2.0 percent, while establishment segregation along race or ethnicity lines raises wages by 6.3, 5.3, and 3.5 percent for these other three groups. Thus, the larger role of ethnic segregation in the lower wages of Hispanic men stems from both more severe segregation and stronger deleterious (or weaker beneficial) effects of segregation, while the larger role of segregation in the lower wages of black men stems mainly from the stronger effects of segregation.

To explore the sensitivity of these results to the level of occupational aggregation, table 4 reports results from a parallel analysis using 72 Census occupations instead of 13. Looking first at the regression estimates of the within-job-cell racial

Table 4. Estimated Log Wage Differentials by Race and Ethnicity, and Percent Black and Hispanic in Occupation, Industry, Establishment, and Job Cell, Based on 72 Occupations

	Females					Males				
	Regression Estimates (1)	Mean Difference Black-White (2)	Contribution to Black-White Wage Gap (3)	Mean Difference Hispanic-White (4)	Contribution to Hispanic-White Wage Gap (5)	Regression Estimates (6)	Mean Difference Black-White (7)	Contribution to Black-White Wage Gap (8)	Mean Difference Hispanic-White (9)	Hispanic-White Wage Gap (10)
Black	-0.009 (0.004)	1.000	-0.009	0.000	0.000	-0.064 (0.004)	1.000	-0.064	0.000	0.000
Hispanic	-0.010 (0.007)	0.000	0.000	1.000	-0.010	-0.014 (0.006)	0.000	0.000	1.000	-0.014
% black in occupation	-1.323 (0.023)	0.027	-0.036	0.014	-0.019	-2.513 (0.029)	0.020	-0.049	0.013	-0.032
% black in industry	0.385 (0.022)	0.015	0.006	0.001	0.000	-0.140 (0.023)	0.013	-0.002	0.004	-0.001
% black in establishment	0.140 (0.008)	0.292	0.041	-0.000	-0.000	0.120 (0.009)	0.224	0.027	-0.002	-0.000
% black in job cell	-0.041 (0.007)	0.548	-0.022	0.005	-0.000	-0.069 (0.007)	0.549	-0.038	0.003	-0.000
Segregation by race	—	—	-0.011	—	-0.019	—	—	-0.062	—	-0.033
% Hispanic in occupation	-0.562 (0.041)	0.015	-0.008	0.011	-0.006	-0.448 (0.033)	0.016	-0.007	0.015	-0.007

Table 4. Estimated Log Wage Differentials by Race and Ethnicity, and Percent Black and Hispanic in Occupation, Industry, Establishment, and Job Cell, Based on 72 Occupations (Continued)

	Females					Males				
	Mean Difference Black-White	Contribution to Black-White Wage Gap	Mean Difference Hispanic-White	Contribution to Hispanic-White Wage Gap	Regression Estimates	Mean Difference Black-White	Contribution to Black-White Wage Gap	Mean Difference Hispanic-White	Hispanic-White Wage Gap	R ²
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
% Hispanic in industry	-0.434 (0.033)	0.004	-0.002	0.007	-0.003	-1.860 (0.025)	0.003	-0.006	0.012	-0.023
% Hispanic in establishment	0.103 (0.011)	0.001	0.000	0.316	0.033	0.037 (0.010)	0.003	0.000	0.328	0.012
% Hispanic in job cell	-0.068 (0.010)	0.003	-0.000	0.578	-0.039	-0.081 (0.009)	0.005	-0.000	0.636	-0.052
Segregation by Hispanic ethnicity	—	—	-0.010	—	-0.015	—	—	-0.013	—	-0.070
R ²	0.371	—	—	—	—	0.398	—	—	—	—

Note: See notes to tables 1-3.

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

and ethnic wage differences, we see again that accounting for segregation makes the black-white wage gap for women very small (-0.009), and similarly for the Hispanic-white wage gap for women (-0.010 , and insignificant). We also see, again, that segregation accounts for a sizable portion of the greater black-white wage gap among men than among women, with the estimated black-white wage gap for males falling from -0.122 in table 2 to -0.064 in table 4. On the other hand, even more so than in table 3, accounting for ethnic segregation lowers the Hispanic-white wage gap among men (-0.014) almost to the same magnitude as for women (-0.010). As we would expect, this is reflected in an even more pronounced difference in the extent to which ethnic segregation lowers wages of Hispanic men. As indicated in the second-to-last row of the table, the combined effect of ethnic segregation on Hispanic women is to lower wages by 1.5 percent, compared with a much larger 7.0 percent figure for men. By way of contrast, the negative effect of racial segregation is only a bit larger for black men than for black women (3.3 vs. 1.9 percent).

Thus, we have a relatively robust finding from two quite different levels of occupational disaggregation indicating that the consequences of segregation are particularly severe for Hispanic men, and account for much or most of the larger Hispanic-white wage gap for men compared to women. In contrast, segregation explains about one-third to nearly one-half of the larger black-white wage gap for men. It appears, therefore, that although segregation is important, some other source aside from segregation accounts for a sizable fraction of the black-white wage gap for men. The evidence in this paper does not speak directly to whether this might be wage discrimination or unobservable productivity differences.

Estimates Using Fixed Effects

Finally, we turn to the fixed-effects analysis mentioned earlier. In part to examine the effects of measurement error in the estimated percentages black and Hispanic by establishment and job cell, we compare results for within-job-cell racial and ethnic wage gaps from the regressions reported in tables 3 and 4 with those that we obtain using fixed occupation, industry, establishment, and job-cell effects. We would think that, in the absence of this measurement error, the two procedures would yield similar results. Although the job-cell dummy variables capture the effects of unobservable variables as well as the effects of the percentages black and Hispanic, the correlation between these unobservables and the black and Hispanic dummy variables should be accounted for by including the segregation variables, so that once these segregation variables are included, unobservable characteristics of the job cell should be uncorrelated with race and ethnicity.

In panels A and B of table 5, the first three rows summarize the earlier results. We first report the estimated coefficients of the black and Hispanic dummy variables from the basic wage regressions without segregation controls (corresponding to columns (3) and (7) of table 2). We then report these estimates once the segregation controls are

Table 5. Estimated Log Wage Differentials by Race and Ethnicity, Summary and Comparison with Fixed-Effects Estimates

	Females		Males		Sex Difference in Wage Gap	
	Black (1)	Hispanic (2)	Black (3)	Hispanic (4)	Black (5)	Hispanic (6)
A. 13 occupations:						
Without segregation controls, coefficients	-0.023 (0.003)	-0.023 (0.005)	-0.122 (0.003)	-0.068 (0.004)	0.099 (0.004)	0.045 (0.006)
With segregation controls, coefficients	-0.012 (0.004)	-0.016 (0.006)	-0.073 (0.004)	-0.029 (0.006)	0.061 (0.006)	0.013 (0.008)
Combined segregation effect:						
Race	-0.004	-0.010	-0.047	-0.026	—	—
Hispanic ethnicity	-0.013	-0.013	-0.020	-0.061	—	—
With job-cell dummy variables, coefficients	-0.022 (0.003)	-0.035 (0.005)	-0.074 (0.003)	-0.038 (0.005)	0.052 (0.004)	0.003 (0.007)
N	298,677		339,041			
B. 72 occupations:						
Without segregation controls, coefficients	-0.023 (0.003)	-0.023 (0.005)	-0.122 (0.003)	-0.068 (0.004)	0.099 (0.004)	0.045 (0.006)
With segregation controls, coefficients	-0.009 (0.004)	-0.010 (0.007)	-0.064 (0.004)	-0.014 (0.006)	0.055 (0.006)	0.004 (0.009)

Table 5. Estimated Log Wage Differentials by Race and Ethnicity, Summary and Comparison with Fixed-Effects Estimates (Continued)

	Females		Males		Sex Difference in Wage Gap	
	Black (1)	Hispanic (2)	Black (3)	Hispanic (4)	Black (5)	Hispanic (6)
Combined segregation effect:						
Race	-0.011	-0.019	-0.062	-0.033	—	—
Hispanic ethnicity	-0.010	-0.015	-0.013	-0.070	—	—
With job-cell dummy variables, coefficients	-0.020 (0.003)	-0.026 (0.006)	-0.063 (0.003)	-0.023 (0.005)	0.043 (0.004)	-0.003 (0.008)
N	298,677		339,041			

Note: The other control variables included correspond to those in columns (3) and (7) in table 2. The standard errors in columns (5) and (6) are calculated assuming independent samples.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

included, followed by the combined segregation effects (corresponding to the rows labeled "Segregation by race" and "Segregation by Hispanic ethnicity" in tables 3 and 4). Finally, we report the estimated coefficients of the black and Hispanic dummy variables when we instead include the fixed effects. The black-white wage gap, and nearly one-half of the sex difference in the black-white wage gap.

In both panels A and B we see that the sex difference in the wage gap between blacks and Hispanics is a bit smaller when we use job-cell dummy variables, as opposed to segregation controls.²⁰ However, the results are qualitatively similar; in panel A we see that the sex difference in the wage gap for blacks is 0.052 when using job-cell dummy variables and 0.061 when using segregation controls, whereas for Hispanics the sex differences are 0.003 and 0.013. Thus, we again find that controlling for segregation accounts for most or all of the sex difference in the His

The estimates in panel B are very similar to those in panel A, again confirming that the level of occupational aggregation has little influence on the results. Most important, more of the sex difference in the black-white wage gap remains (0.043), compared with the sex difference in the Hispanic-white wage gap (which actually changes sign, to -0.003 , but is indistinguishable from zero).

5. Conclusions

The goal of this paper is to assemble general evidence on the effects on wages of racial and ethnic segregation along the lines of occupations, industries, establishments, and job cells (i.e., the same jobs within establishments). More specifically, we ask whether larger racial and ethnic wage differences for men than for women are attributable to more severe segregation among men or to more severe effects of this segregation. To generate this evidence, we use a data set we have constructed called the New Worker-Establishment Characteristics Database, which is based on a match of employees to their establishments of employment.

In standard log wage regressions with individual-level controls, black-white and Hispanic-white differentials among women are around 2 percent, while among men the black-white differential is 12 percent and the Hispanic-white differential 7 percent. Our evidence indicates that greater segregation between Hispanic and white men than between Hispanic and white women explains essentially all of the higher Hispanic-white wage gap for men. Similarly, our estimates indicate that greater segregation between black and white men than between black and white women explains a large share (one-third to one-half) of the higher black-white wage gap for men, although the black-white wage gap for men remains sizable (about 6 to 7 percent) after controlling for segregation.

²⁰ The estimated within-job-cell wage gaps are similar and if anything somewhat larger (in absolute value) using the fixed effects, implying that these estimates were not biased away from zero owing to measurement error in the segregation variables.

Section 1. Analysis with Employer-Employee Matched Data
Wage Differentials in Firms

Overall, our results imply that segregation is an important contributor to the lower wages paid to black and Hispanic men relative to white men with similar individual characteristics. It further suggests that equal pay types of laws may offer some scope for reducing the black-white wage differential for men, but little scope for reducing the Hispanic-white wage differential for men. In contrast, policies intended to reduce the latter would appear to be more effective if they targeted segregation into lower paying jobs. More definitive policy conclusions await further analysis of the type of evidence we report in this paper. While we view the NWECD as an important step forward, the construction of other data sets that improve on some of its shortcomings should be pursued, which we intend to do in future research.

References

Altonji, Joseph G., and Blank, Rebecca M. "Race and Gender in the Labor Market." *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card. Amsterdam: North-Holland, forthcoming.

Bayard, Kimberly; Hellerstein, Judith; Neumark, David; and Troske, Kenneth. "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data." Working Paper No. 98-12. Columbia, MO: University of Missouri-Columbia, 1998a.

———. "Why are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database." Working Paper No. 98-13. Columbia, MO: University of Missouri-Columbia, 1998b.

Blau, Francine D. *Equal Pay in the Office*, Lexington, MA: D.C. Heath and Company, 1977.

Cain, Glen C. "The Economic Analysis of Labor Market Discrimination: A Survey." In *Handbook of Labor Economics*, vol. 1, edited by Orley Ashenfelter and Richard Layard. Amsterdam: North-Holland, 1986.

Carrington, William J.; and Troske, Kenneth R. "Sex Segregation in U.S. Manufacturing." *Industrial and Labor Relations Review* 51 (April 1998a): 445-64.

Carrington, William J., and Troske, Kenneth R. "Interfirm Segregation and the Black-White Wage Gap." *Journal of Labor Economics* 16 (April 1998b): 231-60.

Darity Jr., William A., and Mason, Patrick L. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives* 12 (Spring 1998): 63-92.

Groshen, Erica L. "The Structure of the Female-Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?" *Journal of Human Resources* 26 (Summer 1991): 457-72.

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

King, Mary C. "Occupational Segregation by Race and Sex, 1940-88." *Monthly Labor Review* (April 1992): 30-7.

Macpherson, David A., and Hirsch, Barry T. "Wages and Gender Composition: Why Do Women's Jobs Pay Less?" *Journal of Labor Economics* 13 (July 1995): 426-71.

Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women?

K. Bayard, J. Hellerstein, D. Neumark, and K. Troske

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

Oaxaca, Ronald. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (October 1973): 693-709.

Sorensen, Elaine. "Measuring the Effect of Occupational Sex and Race Composition on Earnings." In *Pay Equity: Empirical Inquiries*, edited by Robert T. Michael, Heidi I. Hartmann, and Brigid O'Farrell. Washington, D.C.: National Academy Press 1989a.

———. "Measuring the Pay Disparity Between Typically Female Occupations and Other Jobs: A Bivariate Selectivity Approach." *Industrial and Labor Relations Review* 42 (July 1989b): 624-39.

Trejo, Stephen J. "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy* 105 (December 1997): 1235-68.

Watts, Martin J. "Trends in Occupational Segregation by Race and Gender in the U.S.A., 1983-92: A Multidimensional Approach." *Review of Radical Political Economics* 27 (1995): 1-36.