Interfirm Segregation and the Black/White Wage Gap
Author(s): William J. Carrington and Kenneth R. Troske
Source: *Journal of Labor Economics*, Vol. 16, No. 2 (April 1998), pp. 231-260
Published by: The University of Chicago Press on behalf of the Society of Labor Economists and the NORC at the University of Chicago
Stable URL: http://www.jstor.org/stable/10.1086/209888

# Interfirm Segregation and the Black/White Wage Gap

William J. Carrington, *Charles River Associates*

Kenneth R. Troske, *University of Missouri—Columbia*

This article studies interfirm racial segregation in two newly developed firm-level databases. We find that the interfirm distribution of black and white workers is close to what would be implied by random assignment. We also find that black workers are clustered in employers where managers, owners, and customers are also black. These findings may be reconciled by the facts that (*a*) there are not enough black employers to generate much segregation and that (*b*) other forces may systematically integrate black and white workers. Finally, we find that the black/white wage gap is primarily a within-firm phenomenon.

## I. Introduction

Interracial contact is an important index of social health. Yet while there are innumerable studies of residential racial segregation, we know very little about the extent to which blacks and whites are integrated at work. This is unfortunate because employed adults spend a large fraction

We thank Bill Evans and seminar participants at the Center for Economic Studies, the University of Maryland, and the National Bureau of Economic Development for helpful comments. The opinions expressed here are solely ours and do not reflect the opinions of the U.S. Census Bureau. The data used in this article were collected under the provisions of Title 13 U.S. Code and are available only at the Center for Economic Studies, U.S. Census Bureau. To obtain access to these data, contact The Chief, Center for Economic Studies, U.S. Census Bureau, RM 211/WPII, Washington, DC 20233.

[*Journal of Labor Economics,* 1998, vol. 16, no. 2]

0734-306X/98/1602-0003$02.50

of their time at work and because perhaps an even larger fraction of nonfamily social interaction takes place at work. Thus, an understanding of racial integration in society as a whole requires some understanding of racial segregation at work.

This article advances our understanding of interfirm racial segregation in three ways. First, there are very few studies of interfirm racial segregation, largely, we believe, because the appropriate data have not been available. Thus, our work simply lengthens a short literature on this topic. Second, such studies as exist use data from the mid-1970s (Becker 1980), the late 1960s (Flanagan 1973), or from the early part of the twentieth century (Higgs 1977). In contrast, we use data from the late 1980s. Finally, commonly used measures of segregation conflate systematic segregation, which occurs in some models of labor market discrimination, for example, with segregation due to the random allocation of workers to firms. In contrast, we use new methods that better sort out the systematic and random components of segregation.

Our first results simply measure interfirm segregation. We find that black and white workers are quite segregated by conventional measures. However, when we look within metropolitan statistical areas (MSAs), we find that the distribution of black and white workers is surprisingly close to what would be implied by random allocation of workers to firms. We find virtually no systematic segregation in our first sample of large manufacturing establishments, and while we find some systematic segregation within our second sample of smaller firms drawn from a range of industries, segregation is still far from complete. This is particularly true among workers stratified by industry, occupation, or educational attainment. Indeed, in some cases we find that black and white workers of similar skill are actually more *integrated* than random allocation would predict. In sum, we find that the distribution of black and white workers, particularly when workers are grouped by skill, can be modeled reasonably well by a random hiring model.

Our second set of results examines the matching of workers to firms. Our main finding is that black workers are disproportionately sorted into firms in which owners, managers, and customers are also black. These findings are at first contradictory, because how can the distribution of workers look nearly random, while at the same time black workers are sorted into "black" employers (i.e., those with black owners, managers, and customers)? There are two complementary answers to this question. First, while the relationship between the race of employers and the race of workers is moderately strong, there are not many black employers, so the relationship does not generate much segregation. Second, it is possible that Title VII, affirmative action, and other difficult-to-identify forces systematically integrate black and white workers. On balance, these weak opposing forces induce an interfirm distribution of workers that is roughly consistent with the random allocation of workers to employers.

Our final results decompose the black/white wage gap in the manufacturing industry into within- and between-plant components. The central finding is that most of the black/white wage gap among men is accounted for by within-plant differences in pay. In fact, we find that both black men and black women are disproportionately sorted into plants with above-average wages. In addition, we find that the wages for all workers tend to increase with the fraction of their coworkers that are of the opposite race. Thus, whites earn the most in plants with many blacks, and blacks earn the most in plants that are nearly all white.

This article proceeds as follows. Section II describes our data drawn from the Worker-Establishment Characteristics Database and the Characteristics of Business Owners survey. Section III discusses our approach to measuring the systematic component of segregation, and Section IV uses this approach to measure and interpret interfirm racial segregation. Section V then analyzes the determinants of interfirm segregation, and Section VI assesses the role of segregation in accounting for the black/white wage gap. Although this article is primarily descriptive, Section VII discusses some implications of the results.

## II. Data Sources

This section describes the two data sets used in this article, the first of which is the Worker-Establishment Characteristics Database (WECD).[1] In the WECD workers are identified through household responses to the 1990 census long form, which contains standard demographic information as well as the location and industry code for each respondent's place of work.[2] All workers located in an industry-location cell with a unique establishment are matched to that establishment, whereas workers in cells without a unique establishment are not included in the sample. Establishments themselves are identified as being unique or not based on our analysis of the Census Bureau's list of manufacturing establishments. This matching process results in a data set consisting of 199,558 workers matched to 16,144 plants. Since the WECD matches workers only to employers that are unique in an industry-location cell, it tends to eliminate small plants that are less likely to meet this criterion. Thus, our WECD results describe segregation in large manufacturing plants, and the results may not generalize to the economy at large. However, Bound and Free-

[1] See Troske (in press) for a more complete description of the WECD.

[2] Industry information is recorded at the three-digit level. For establishments in urban areas (primarily MSAs), a plant's location is coded at the block level. For establishments in rural areas, a plant's location is coded at the place level.

man (1992) argue that the post-1975 decline in the relative wages of young black men is partially attributable to the employment declines in manufacturing, so manufacturing is a particularly interesting sector to study from the perspective of racial segregation and wage differences.

The large plant focus of the WECD leads us to also study segregation among the smaller firms surveyed in the Characteristics of Business Owners (CBO) database. The CBO is the result of a 1987 Census Bureau survey of 125,000 small business owners.[3] The CBO is essentially a survey of firm *owners,* but one can merge owner responses to obtain a firm-level data set (Carrington and Troske 1995). The resulting database includes information on the firm itself (location, industry, receipts, capital), the firm's owners (income, race, sex, age, education), the firm's customers (% minority), and the firm's workforce (number of employees, % minority, % female, payroll).[4] These latter data allow calculations of interfirm racial segregation. Compared with the WECD, the CBO is limited in that it has no information on workers' characteristics other than race and sex and reveals nothing about the within-firm distribution of pay. However, it complements the WECD with demographic information on owners and customers and is not restricted to the manufacturing industry. It bears emphasis that in the CBO we analyze segregation across small firms, while in the WECD we analyze segregation across large plants. In

[3] Owners were included in the sample frame if they filed their tax return with one of the following IRS forms: 1040 (Schedule C), 1065, or 1120S. The 1040 (Schedule C) returns correspond to individual proprietorships, or unincorporated businesses owned by an individual. The 1065 returns include unincorporated businesses owned by two or more persons. Finally, the 1120S returns correspond to subchapter S corporations, which are legally incorporated businesses with 35 or fewer shareholders who elect to be taxed as individuals rather than corporations. Corporations filing a regular 1120 tax return were excluded from the sample. The CBO comprises five equal-sized panels of business owners drawn from five demographic groups: Hispanics, blacks, other minorities, women, and nonminority men. The equal size of the panels obviously required that the CBO oversample certain ownership groups, but sample weights can be used to recover the attributes of a random sample.

[4] Linking owner characteristics to the firm is trivial for firms owned by one person, but multiowner firms are slightly tricky because not all owners are alike. Following the work of previous CBO users (Bates 1988; Carrington and Troske 1995), we use the cross-owner mean for continuous variables (such as education) and the cross-owner mode for discrete variables (such as sex or race). In cases of ties for the discrete variables, we use the mode containing the owner that reports spending the most hours per week at the business. For example, if a firm has two white and one black owners, then we describe the firm as being "white-owned." If a firm has one white and one black owner, then we describe the firm as "white-owned" if the white reports working more weekly hours at the firm, and as "black-owned" if not. Single-owner firms account for 57% of the firms and 39% of the employment in our sample.

statements referring to both data sets, we generally speak of "interfirm" segregation, though this applies directly only to the CBO.[5]

One complication with the CBO is that employees are only identified as "minority" or not.[6] Since only about 50% of all minorities (i.e., those reported as black, Asian, or Hispanic) are black, we imposed two restrictions to minimize the number of nonblack minorities in our CBO sample of employees. First, we excluded CBO firms owned by Asians and Hispanics because we suspect that minority employees of such firms are much less likely to be black. Second, we restricted the CBO sample to states and MSAs where blacks account for over 75% of all minorities.[7] This amounts to restricting the sample to southern states such as Mississippi, Georgia, and North Carolina, and to certain northern cities such as Detroit.[8] In addition, this restriction increases the black share of em-

[5] Because the CBO oversamples black and minority business owners, we use sample weights in our CBO analysis. The sample weights in the CBO are inversely proportional to the estimated probability that each firm entered the sample. We scaled the weights so that they summed to the sample size.

[6] More particularly, employers reported the fraction of "white non-Hispanics" in their workforce.

[7] We treated the non-MSA portions of each state as a single independent MSA (e.g., rural Texas).

[8] These states and MSAs were identified from our analysis of the outgoing rotation groups of the Current Population Survey (CPS) for the months of January 1987 to December 1991. We restricted the sample to those employed at the time of the survey and to those aged 16 or older. In addition, we restricted the sample to those in their first year of CPS interviews (out of a maximum of two outgoing CPS interviews) to avoid double-counting individuals. The resulting sample included roughly 570,000 individuals, of whom approximately 150,000 were minorities. Blacks account for a small fraction of all minorities in almost all western cities, 21% in Los Angeles, for example. In contrast, blacks account for a very high share of minorities in the South, with the exception of the Miami MSA and the MSAs in Texas. For example, blacks account for 99% of the minorities in Birmingham, Alabama. While there is a sharp distinction between the South and West along these lines, MSAs within the Northeast and Midwest are more varied. Minorities are relatively heterogeneous in some northern cities such as New York and Chicago, but blacks account for a very large share of all minorities in other northern cities. For example, blacks account for 90% of all minorities in the Detroit metropolitan area. Thus, restricting the sample to MSAs where blacks account for a large share of all minorities is equivalent to restricting the sample to all southern MSAs save for Miami and those in Texas and to a smattering of MSAs in the North. It is worth noting that black workers are disproportionately located in states and MSAs where blacks account for the bulk of all minorities. For example, 75% of all nonrural blacks (i.e., those who live in an MSA) live in MSAs where blacks account for at least 54% of all minorities. Similarly, 50% of blacks live in MSAs where blacks account for more than 83% of all minorities, and 25% of blacks live in MSAs where blacks account for more than 92% of all minorities.

ployment in our CBO sample to roughly 30%. While this does not completely avoid the confusion of blacks with other minorities, these restrictions provide some legitimacy to the interpretation of our CBO results as reflecting black/white differences.[9]

Since neither the WECD nor the CBO contain a random sample of workers, it is important to ask whether patterns of segregation found in these data are likely to parallel the rest of the economy. Of course, a precise answer to this question would require a random sample of workers that contained employer identifiers, and it is the absence of such data that led us to the WECD and the CBO in the first place. Nevertheless, table 1 considers the measurable similarities between workers in the WECD, the CBO, and two relatively representative data sets. Columns 1 and 2 present summary statistics for manufacturing workers in the 1990 Decennial Census, for blacks and whites separately. Columns 3 and 4 present analogous statistics for the WECD, columns 5 and 6 present results for the May 1988 Current Population Survey, and columns 7 and 8 present results for the CBO.[10] The comparisons between the Decennial Census data and the WECD suggest several ways in which WECD workers are different from the manufacturing industry at large. First, WECD workers are more likely to live in the Northeast and Midwest and less likely to live in the South and West. Second, average hourly wages and annual earnings for workers are higher in the WECD than in the Decennial Census. Finally, the black/white wage and earnings gap is somewhat lower in the WECD than in the Decennial Census. These facts suggest that, while the WECD is not wildly different from the rest of the manufacturing industry, there is some legitimate concern about the representativeness of the WECD.

Table 1 also illustrates some special features of our CBO sample. First, the black population share is much larger than that in the WECD or in the general U.S. population. This arises because we have lumped all "minorities" into the "black" group and because we have restricted the

[9] Our CBO analysis is also complicated by the fact that while most multiunit CBO firms operate in a single MSA, a small fraction (280 out of more than 7,000 firms in the CBO) have establishments in more than one MSA. This is potentially problematic when we later assign each firm as belonging to a particular MSA, particularly because these multi-MSA firms are somewhat bigger than average and because 20% of their employment occurs outside of their main MSA. Unfortunately, the available data do not allow us to break out each firm's employment by MSA and race simultaneously. Thus, we assigned all employment to the firm's main MSA even though this is not strictly accurate. We examined the sensitivity of our results to the inclusion of these multi-MSA firms, and our results are not substantively dependent on whether they are included in the sample.

[10] We chose the May 1988 CPS for comparison because it included questions about the number of employees in establishments where respondents work.

**Table 1**
**Sample Characteristics of Workers**

| Characteristic | 1990 Decennial Census Manufacturing White (1) | 1990 Decennial Census Manufacturing Black (2) | WECD White (3) | WECD Black (4) | May 1988 CPS White (5) | May 1988 CPS Black (6) | CBO White (7) | CBO Black (8) |
|---|---|---|---|---|---|---|---|---|
| Population share | 92.2 | 7.8 | 93.4 | 6.6 | 92.0 | 8.0 | 72.0 | 28.0 |
| Age | 39.2 | 38.6 | 40.3 | 40.0 | 36.5 | 35.6 | ... | ... |
| Education: | | | | | | | | |
| Less than high school | 18.6 | 28.3 | 17.0 | 24.6 | 14.6 | 25.8 | ... | ... |
| High school diploma | 39.0 | 41.9 | 42.9 | 42.9 | 44.0 | 45.6 | ... | ... |
| Some college | 26.1 | 23.3 | 25.7 | 25.7 | 21.7 | 18.5 | ... | ... |
| College | 12.3 | 5.22 | 11.1 | 5.6 | 12.8 | 7.5 | ... | ... |
| Advanced degree | 4.0 | 1.37 | 3.3 | 1.3 | 6.8 | 2.5 | ... | ... |
| Experience | 21.0 | 21.1 | 22.2 | 22.4 | 17.7 | 17.6 | ... | ... |
| % Women | 30.7 | 41.5 | 27.6 | 37.9 | 45.6 | 51.8 | ... | ... |
| Region: | | | | | | | | |
| Northeast | 21.5 | 11.7 | 28.6 | 13.7 | 23.2 | 14.0 | 11.3 | 8.5 |
| South | 28.7 | 64.0 | 19.7 | 50.4 | 30.3 | 62.7 | 66.7 | 73.1 |
| Midwest | 35.4 | 18.9 | 45.8 | 33.8 | 29.0 | 17.8 | 22.0 | 18.4 |
| West | 14.4 | 5.5 | 5.9 | 2.1 | 17.4 | 5.4 | .0 | .0 |
| Occupational shares: | | | | | | | | |
| Professionals/technicans/managers | 19.6 | 6.7 | 17.7 | 6.1 | 22.3 | 10.7 | ... | ... |
| Sales and service | 7.6 | 3.8 | 5.8 | 3.3 | 16.2 | 10.9 | ... | ... |
| Clerical | 13.7 | 12.9 | 13.4 | 11.1 | 28.6 | 36.2 | ... | ... |
| Craftsmen | 20.2 | 16.2 | 21.2 | 17.1 | 13.9 | 10.2 | ... | ... |
| Operatives | 33.9 | 51.9 | 35.9 | 53.8 | 14.1 | 23.6 | ... | ... |
| Laborers | 5.1 | 8.5 | 6.0 | 8.7 | 4.8 | 8.4 | ... | ... |
| Establishment size: | | | | | | | | |
| 1–9 | ... | ... | 2.2 | 1.4 | 22.5 | 15.5 | 13.3 | 12.7 |
| 10–24 | ... | ... | 3.8 | 2.2 | 15.0 | 11.4 | 18.3 | 15.9 |
| 25–49 | ... | ... | 4.9 | 3.3 | 12.8 | 11.8 | 16.4 | 16.1 |
| 50–99 | ... | ... | 9.1 | 7.3 | 10.9 | 10.4 | 15.1 | 17.5 |
| 100–249 | ... | ... | 17.8 | 14.5 | 13.1 | 14.3 | 18.2 | 20.7 |
| 250+ | ... | ... | 62.2 | 71.3 | 25.7 | 36.6 | 18.7 | 17.2 |
| Hourly wages | 12.86 | 10.24 | 13.56 | 11.96 | 9.93 | 7.56 | ... | ... |
| Log(hourly wages) | 2.40 | 2.19 | 2.48 | 2.35 | 2.08 | 1.82 | ... | ... |
| Annual earnings | 28,275 | 21,396 | 29,863 | 25,124 | 19,897 | 14,121 | 14,283 | 12,927 |
| Log(annual earnings) | 10.07 | 9.82 | 10.16 | 9.99 | 9.53 | 9.18 | 9.29 | 9.21 |

NOTE.—Results for cols. 1–2 are based on a sample of manufacturing workers drawn from the Sample Detail File. Results for cols. 3–4 are based on all workers in the Worker-Establishment Characteristics Database (WECD), who are by construction employed in the manufacturing industry. Results for cols. 5–6 are based on workers in the May 1988 Current Population Survey (CPS), with no restriction on industry. For each of these samples, workers were included only if they had a reasonable degree of labor force attachment. Finally, results for cols. 7–8 are based on our sample from the Characteristics of Business Owners (CBO) database. Since we have very crude information on workers in the CBO, we could not apply any restrictions associated with labor force attachment.

sample to locales with large black populations. Second, these geographic restrictions have led to a sample overrepresented in the South and in certain cities in the Northeast and Midwest. Third, CBO firms are small, as less than 20% of CBO workers are employed at firms with more than 250 employees. Finally, CBO earnings are lower than in the rest of the economy, partly because of the focus on the South but also because wages and earnings are lower in the small firms overrepresented in the CBO. The lack of information on workers' age and education also illustrates an important weakness of the CBO; all we really know about the workers is their race and sex.

Table 2 presents summary statistics on the firms in our two samples. The left side of the table presents data on our WECD sample of manufacturing plants. The average plant in our sample employs roughly 155 workers and ships about $130,000 worth of goods per worker per year. Note that while the average establishment represented in the WECD is large, the number of workers per establishment *in our sample* is only about 12. The incomplete sampling within establishments occurs because workers had to both answer the long form of the 1990 Decennial Census and give accurate information about their employer's address. The right side of table 2 presents information on the firms in our CBO sample. The table reports figures for all firms and reports them separately for black- and white-owned businesses. Rows 1 and 2 show that the CBO firms are obviously much smaller than our WECD firms. The rows also show that the white-owned businesses are larger than those owned by

**Table 2**
**Characteristics of Firms and Owners**

| WECD: Characteristic | *M* |
|---|---|
| 1. Total workers in plant | 155.3 |
| 2. Sample workers in plant | 12.0 |
| 3. Shipments/worker | 131,606 |
| 4. Annual earnings/worker ($) | 24,542 |
| 5. % multiunit | 46 |
| 6. % in MSA | 88 |
| 7. Region (in %): | |
| Northeast | 32.9 |
| South | 22.9 |
| Midwest | 36.2 |
| West | 8.1 |

| CBO: Characteristic | All Firms | Black-Owned | White-Owned |
|---|---|---|---|
| 1. Workers in firm | 6.4 | 4.8 | 6.4 |
| 2. Receipts/worker | 77,196 | 57,776 | 77,790 |
| 3. Average earnings ($) | 11,271 | 8,826 | 11,346 |
| 4. Education of owner (in %): | | | |
| Elementary school | 4.6 | 10.3 | 4.4 |
| Some high school | 10.7 | 15.5 | 10.6 |
| High school/General Equivalency Degree | 29.3 | 25.9 | 29.4 |
| Some college | 19.9 | 17.4 | 20.0 |
| College graduate | 14.6 | 11.5 | 14.7 |
| Graduate school | 20.8 | 19.4 | 20.9 |
| 5. Age of owner (in %): | | | |
| Under 25 | .5 | .8 | .4 |
| 25–34 | 11.7 | 9.0 | 11.8 |
| 35–44 | 32.2 | 26.1 | 32.4 |
| 45–54 | 26.1 | 28.5 | 26.0 |
| 55–64 | 21.7 | 24.0 | 21.6 |
| 65+ | 7.8 | 11.6 | 7.7 |
| 6. Region (in %): | | 7.1 | |
| Northeast | 12.8 | | 13.0 |
| South | 60.8 | 75.0 | 60.4 |
| Midwest | 26.4 | 17.9 | 26.7 |
| West | .0 | .0 | .0 |
| 7. % in MSA | 70.4 | 72.9 | 70.3 |
| 8. Number of owners | 1.1 | 1.1 | 1.1 |

NOTE.—The unit of observation in this table is an establishment (in the WECD) or a firm (in the CBO). MSA = metropolitan statistical area.

blacks. The table also reveals that we know the education, age, race, and sex of CBO owners. Since CBO owners presumably make hiring decisions, this information is useful in assessing the role of employer discrimination in generating interfirm segregation.

## III. On Measuring Segregation

This section considers a methodological question that we need to address before we can adequately analyze the data discussed in Section II. Economists typically summarize segregation patterns with a segregation index, which is simply a statistic that summarizes the extent to which two groups interact in a sample. Although there are many such indices, our work focusses on the Gini coefficient of segregation because of its desirable properties (Hutchens 1991). This index measures the extent to which the distribution of blacks and whites across firms deviates from an *even* distribution in which each group is proportionately represented in each firm. To present this more formally, let there be $T$ firms, let $s_{\mathrm{b}i}$ and $s_{\mathrm{w}i}$ be firm $i$'s share of the black and white sample populations, respectively, and let the firms be sorted in ascending order of $s_{\mathrm{b}i}/s_{\mathrm{w}i}$. Then the Gini coefficient is

$$G = 1 - \sum_{i=1}^{T} (s_{\mathrm{b}i}) \left( s_{\mathrm{w}i} + 2 \sum_{j=i+1}^{T} s_{\mathrm{w}j} \right). \tag{1}$$

The Gini coefficient varies between zero and one, with zero corresponding to complete evenness and one corresponding to complete unevenness, which occurs when blacks and whites never work for the same employer.[11]

An often overlooked feature of the Gini coefficient of segregation (and other popular indices) is that it can be positive when workers are allocated randomly across firms.[12] This occurs for two reasons. First, there is a simple integer constraint in that each worker must be uniquely allocated to one firm. In a sample with 10 black workers and 20 firms, for example, evenness is unobtainable because it is impossible for each firm to get half a black worker. Second, and more important, the random allocation of

[11] If every firm has exactly the same share of black workers, say 10%, then the sample is completely even and $G = 0$. In contrast, if every firm is either all white or all black, then the sample is completely uneven, and $G = 1$. The Gini coefficient of segregation has the same geometric interpretation as the Gini coefficient of income inequality, as the indices are analytically equivalent.

[12] This discussion is drawn from Carrington and Troske (1997).

workers to firms typically generates some deviation from evenness.[13] It is easy to show that such random unevenness leads to positive and sometimes large Gini coefficients, particularly in samples with small firms and small minority shares. This is a problem in our study because the average firm has 6.4 workers in the CBO and 12.0 workers in the WECD and because blacks are a small minority, particularly in the WECD. Thus, random allocation would lead to Gini coefficients on the order of .65 for the WECD and .40 for the CBO.

This is unfortunate because we would like to use the term ‘‘segregated’’ only in instances where the distribution of workers across firms is more uneven than would be implied by random allocation.[14] Carrington and Troske (1997) propose the following modifications of the Gini coefficient as a means of distinguishing between systematic and random segregation. Let the *Gini coefficient of random segregation* $G^*$ be the Gini coefficient that would occur if a very large number of workers were allocated randomly to employers, taking the black population share and the size distribution of plants as determined by the sample.[15] Put slightly differently, $G^*$ is the average Gini coefficient obtained if sample workers are assigned randomly to sample plants. Thus, if a sample contains mostly large plants and roughly equal numbers of blacks and whites, then $G^*$ will be close to zero. In contrast, if the sample contains mostly small plants and is either mostly black or mostly white, then $G^*$ will be closer to one.

We use the Gini coefficient of random segregation to adjust the standard Gini coefficient so that it accounts for the role of random assignment

[13] As a stark example, consider a large sample of two-person firms that, in aggregate, employ a 50/50 mix of blacks and whites. Random allocation of workers to firms will result in 25% of the firms employing two blacks, 50% of the firms employing one black and one white, and 25% of the firms employing two whites. The Gini coefficient of segregation would be .75 in this instance.

[14] Our interest in separating systematic from random effects, as well as our fix to the problem, are similar in spirit to Ellison and Glaeser's (1997) work on indices of geographical concentration.

[15] See Carrington and Troske (1997) for a complete discussion of this issue. The basic idea behind the computation of $G^*$ is as follows. The first step is to calculate the *empirical* number of firms in each size class $s$, $\hat{g}(s)$. Within any size class $s$, random allocation implies that the binomial function $B(m; s, p)$ is the fraction of firms that will have $m$ blacks if $p$ is the black population share. Thus, random allocation implies that the number of sample units with size $s$ and $m$ blacks should be $N(m, s; p) = B(m; s, p)\hat{g}(s)$. Thus, the support of $N(m, s; p)$ is $m = 0$ to $s$ for every $s$ in the support of $\hat{g}(s)$. This artificial distribution corresponds to what would be expected if workers were allocated randomly, given $p$ and $\hat{g}(s)$. The Gini coefficient of segregation computed from this artificial distribution is what we call the *Gini coefficient of random segregation.*

in causing unevenness. In particular, we define the *Gini coefficient of systematic segregation* as

$$\hat{G} = \frac{G - G^*}{1 - G^*} \quad \text{if } G - G^* \geq 0$$

and

$$\hat{G} = \frac{G - G^*}{G^*} \quad \text{if } G - G^* < 0.$$

If the sample is more uneven than random allocation would imply—that is, $G > G^*$—then $\hat{G} > 0$ is simply excess unevenness $(G - G^*)$ expressed as a fraction of the maximum amount of such excess segregation that could possibly occur $(1 - G^*)$. When $\hat{G} = 1$, it is analogous to complete unevenness, as with the standard gini coefficient, but $\hat{G} = 0$ implies that the sample is equivalent to random allocation. If, in contrast, there is excess evenness—that is, $G < G^*$—then $\hat{G}$ is negative and represents excess evenness in the sample $(G - G^*)$, expressed as a fraction of the maximum amount of such excess evenness that could possibly occur $(G^*)$.[16]

To summarize, our $\hat{G}$ is different from the standard Gini coefficient in two ways. First, we have set the baseline of zero to correspond to random allocation rather than complete evenness. Second, we have remapped values of $G$ that are greater than $G^*$ into the $[0, 1]$ interval and remapped values of $G$ that are less than $G^*$ into the $[-1, 0]$ interval. We think that this modified index provides more useful information than the traditional Gini coefficient. However, we recognize that some readers will be more comfortable with the traditional index. Thus, in the work that follows we report traditional indices of segregation, indices of random segregation, and the indices of systematic segregation that we developed here. Together, these indices provide a useful summary of segregation patterns.

[16] The example from n. 13 above illustrates the difference between $G$ and $\hat{G}$. First, consider a large sample of two-person plants with equal aggregate numbers of men and women and a traditional Gini coefficient of .75. In this case, $\hat{G} = 0$ because .75 is exactly what random assignment would imply. Second, consider a large sample of 1,000-worker plants with equal aggregate numbers of men and women and a traditional Gini coefficient of .30. In this case, $\hat{G}$ would be very close to .30 as well, as random assignment implies very little unevenness. As these examples illustrate, the extent to which $\hat{G}$ differs from $G$ is entirely dependent on the likely role of random allocation in the particular sample at hand.

## IV. Interfirm Racial Segregation

Table 3 begins this section's analysis of interfirm racial segregation with results from both the WECD (rows 1–13) and the CBO (rows 14–20). Rows 1 and 14 report Gini coefficients for the entire WECD and CBO samples, respectively, while the other rows present results for subsets of the data.[17] For example, rows 2 and 3 analyze the WECD data when the manufacturing industry is broken into durables and nondurables, and rows 4–9 analyze the WECD data when workers are stratified by occupation. Note that while both the CBO and the WECD allow separate analyses by industry, only the WECD admits to separate analyses by workers' occupation. This is a result of the relatively crude demographic information available in the CBO.

While the rows of table 3 vary by data set and by samples defined by characteristics of workers or firms, the columns of table 3 vary on two separate dimensions. First, columns 1–3 differ from columns 4–6 in the geographic sampling scheme under consideration. (Table 4 considers two additional ways of geographically organizing the data.) Second, within any geographic scheme, we report estimates of the traditional Gini coefficient, the Gini coefficient of random segregation, and the Gini coefficient of systematic segregation. In addition to the index values, bootstrap standard errors are reported in parentheses to the right of each index value.[18] Thus, if we look at row 1 of column 1, we see that the Gini coefficient for all workers in the WECD is .78 with a bootstrap standard error of .01.[19]

There is a range of views on the causal relationship between residential segregation and workplace segregation. At one extreme, one might view all residential decisions as the outcome of employment opportunities, in which case residential segregation is entirely due to employment segregation. At the opposite extreme, one might view residential decisions as completely exogenous to the labor market, in which case the causality runs entirely from residential segregation to employment segregation. We

[17] We have conducted an analogous set of calculations using the dissimilarity index (Duncan and Duncan 1955). In spite of its problems (Hutchens 1991), the dissimilarity index remains popular because of its intuitive interpretation. In particular, the dissimilarity index reports the share of black (or white) workers that would have to switch firms in order for the sample to be completely even. The results of our analysis using the dissimilarity index are broadly similar to those reported here, but full results are available from us on request.

[18] Boisso et al. (1994) have shown that bootstrapping provides a useful measure of sampling variation in this context.

[19] We also tested the hypothesis of random allocation for each of our samples with the $\chi^2$ test proposed by Blau (1977). Most entries in tables 3 and 4 represent a statistically significant departure from random allocation, although we will argue that the differences are not always economically significant.

**Table 3**
**Gini Coefficients of Interfirm Racial Segregation**

| | Geographic Scheme | | | | | |
|---|---|---|---|---|---|---|
| | National | | | Within MSAs | | |
| | Standard Gini (1) | Random Gini (2) | Systematic Gini (3) | Standard Gini (4) | Random Gini (5) | Systematic Gini (6) |
| A. WECD: | | | | | | |
| 1. All workers | .78 (.01) | .48 (.01) | .57 (.01) | .60 (.02) | .56 (.02) | .09 (.03) |
| Within industry: | | | | | | |
| 2. Nondurable manufacturing | .81 (.01) | .51 (.01) | .61 (.02) | .63 (.02) | .62 (.01) | .06 (.02) |
| 3. Durable manufacturing | .74 (.01) | .46 (.02) | .52 (.02) | .53 (.02) | .51 (.02) | .02 (.03) |
| Within occupation: | | | | | | |
| 4. Professionals/technicans/managers | .86 (.03) | .86 (.02) | −.03 (.06) | .73 (.05) | .80 (.05) | −.13 (.07) |
| 5. Sales and service | .93 (.01) | .92 (.01) | .06 (.06) | .82 (.03) | .89 (.02) | −.05 (.05) |
| 6. Clerical | .89 (.01) | .86 (.01) | .20 (.03) | .79 (.02) | .85 (.01) | −.06 (.02) |
| 7. Craftsmen | .85 (.01) | .73 (.01) | .45 (.02) | .71 (.02) | .74 (.02) | −.00 (.03) |
| 8. Operatives | .84 (.01) | .58 (.01) | .62 (.02) | .68 (.02) | .66 (.01) | .11 (.03) |
| 9. Laborers | .90 (.01) | .75 (.01) | .60 (.03) | .76 (.01) | .78 (.01) | .08 (.03) |
| Within plant size group: | | | | | | |
| 10. 15 or fewer employees | .97 (.00) | .97 (.00) | .23 (.07) | .95 (.01) | .97 (.00) | .15 (.04) |
| 11. Between 16 and 50 employees | .95 (.01) | .91 (.01) | .41 (.05) | .90 (.01) | .92 (.01) | .13 (.04) |
| 12. Between 51 and 125 employees | .90 (.01) | .76 (.01) | .56 (.02) | .82 (.01) | .82 (.01) | .15 (.02) |
| 13. More than 125 employees | .72 (.02) | .31 (.01) | .59 (.02) | .47 (.02) | .41 (.01) | .06 (.04) |
| B. CBO: | | | | | | |
| 14. All workers | .87 (.01) | .25 (.01) | .83 (.01) | .81 (.02) | .28 (.01) | .74 (.02) |
| Within industry: | | | | | | |
| 15. Nondurable manufacturing | .76 (.04) | .13 (.02) | .72 (.04) | .48 (.07) | .28 (.06) | .16 (.13) |
| 16. Durable manufacturing | .86 (.02) | .21 (.01) | .82 (.02) | .54 (.05) | .30 (.04) | .30 (.09) |
| 17. Construction | .88 (.02) | .26 (.01) | .84 (.02) | .70 (.05) | .33 (.04) | .54 (.07) |
| 18. Wholesale trade | .81 (.03) | .29 (.02) | .74 (.04) | .63 (.04) | .34 (.02) | .42 (.12) |
| 19. Retail trade | .86 (.02) | .28 (.01) | .81 (.03) | .68 (.04) | .35 (.03) | .47 (.07) |
| 20. Services | .90 (.01) | .23 (.01) | .87 (.01) | .76 (.04) | .39 (.04) | .62 (.05) |

NOTE.—The geographic schemes vary as follows. The national measures of segregation presented in cols. 1–3 are computed for the entire United States at a single step. The within-MSA measures of cols. 4–6 are the employment-weighted average of the indices for each MSA. Note that (3) = col. 3 = (col. 1 − col. 2)/(1 − col. 2) by definition. However, col. 6 ≠ (col. 4 − col. 5)/(1 − col. 5) because each of the three indices is the average across MSAs, and the relationship has to hold only within MSAs, and not for the averages. Numbers in parentheses are bootstrap standard errors. See text for description of the Gini coefficient of segregation, the expected Gini coefficient of segregation, and the systematic Gini coefficient of segregation. Also see the text for sample descriptions. In rows 10–13, plants are classified by their *total* employment, not by the sample of their employees that are in our data.

will examine both of these extremes, as well as some intermediate views. As a first step, columns 1–3 of table 3 adopt the "employment segregation causes residential segregation" view by studying interfirm segregation at a national level. The traditional Gini coefficients of column 1 suggest that there is substantial interfirm segregation of black and white workers in the United States, as none of the Gini coefficients are less than .72. For example, the Gini coefficients for all workers are .78 in the WECD and .87 in the CBO, and both are precisely measured. One explanation for

**Table 4**
**Gini Coefficients of Interfirm Racial Segregation**

| | Geographic Scheme | | | | | |
|---|---|---|---|---|---|---|
| | Within Relatively Integrated MSAs | | | Within Small MSAs | | |
| | Gini (1) | Expected Gini (2) | Systematic Gini (3) | Gini (4) | Expected Gini (5) | Systematic Gini (6) |
| A. WECD: | | | | | | |
| 1. All workers | .60 (.05) | .63 (.05) | −.07 (.06) | .59 (.03) | .62 (.03) | −.03 (.04) |
| Within industry: | | | | | | |
| 2. Nondurable manufacturing | .64 (.03) | .68 (.03) | −.08 (.03) | .58 (.04) | .63 (.03) | −.13 (.05) |
| 3. Durable manufacturing | .48 (.06) | .58 (.06) | −.26 (.09) | .52 (.04) | .60 (.03) | −.17 (.05) |
| Within occupation: | | | | | | |
| 4. Professionals/ technicians/managers | .63 (.16) | .68 (.16) | −.26 (.20) | .83 (.03) | .91 (.02) | −.22 (.08) |
| 5. Sales and service | .74 (.08) | .83 (.06) | −.27 (.13) | .89 (.03) | .95 (.01) | −.25 (.09) |
| 6. Clerical | .77 (.07) | .83 (.07) | −.20 (.09) | .82 (.03) | .89 (.02) | −.09 (.06) |
| 7. Craftsmen | .80 (.04) | .84 (.03) | .01 (.09) | .73 (.04) | .79 (.03) | −.10 (.08) |
| 8. Operatives | .74 (.02) | .76 (.02) | .05 (.05) | .69 (.02) | .71 (.02) | .03 (.05) |
| 9. Laborers | .85 (.02) | .88 (.02) | .11 (.12) | .75 (.04) | .81 (.03) | −.05 (.08) |
| Within plant size group: | | | | | | |
| 10. 15 or fewer employees | .96 (.01) | .98 (.01) | .14 (.12) | .96 (.01) | .97 (.01) | .03 (.18) |
| 11. Between 16 and 50 employees | .91 (.02) | .93 (.04) | −.01 (.08) | .90 (.03) | .93 (.02) | −.02 (.11) |
| 12. Between 51 and 125 employees | .81 (.02) | .30 (.02) | −.10 (.02) | .81 (.02) | .86 (.01) | −.12 (.05) |
| 13. More than 125 employees | .49 (.05) | .52 (.05) | −.15 (.08) | .49 (.03) | .52 (.03) | −.08 (.05) |
| B. CBO: | | | | | | |
| 14. All workers | .70 (.07) | .28 (.03) | .58 (.12) | .67 (.06) | .38 (.06) | .51 (.09) |
| Within industry: | | | | | | |
| 15. Nondurable manufacturing | .22 (.14) | .44 (.09) | −.79 (.16) | .24 (.13) | .42 (.10) | −.91 (.07) |
| 16. Durable manufacturing | .37 (.15) | .47 (.08) | −.27 (.32) | .24 (.18) | .22 (.18) | .00 (.34) |
| 17. Construction | .73 (.14) | .31 (.08) | .49 (.28) | .26 (.13) | .57 (.08) | −.68 (.20) |
| 18. Wholesale trade | .42 (.09) | .31 (.05) | .01 (.23) | .40 (.07) | .35 (.06) | −.10 (.22) |
| 19. Retail trade | .38 (.08) | .32 (.07) | −.04 (.16) | .52 (.10) | .52 (.10) | −.09 (.17) |
| 20. Services | .80 (.05) | .46 (.07) | .60 (.13) | .68 (.12) | .64 (.09) | .19 (.25) |

NOTE.—To be included in col. 1, MSAs must be below the median of the residential Gini coefficient distribution or have a residential Gini coefficient of .712 or less. To be included in col. 2, MSAs must be below the median of the MSA population distribution, or have a population of 255,301 or less. Numbers in parentheses are bootstrap standard errors. See text for description of the Gini coefficient of segregation, the expected Gini coefficient of segregation, and the systematic Gini coefficient of segregation. Also see the text for sample descriptions.

this pattern is that blacks and whites have different skills and that workers of all races are segregated by skill (Kremer and Maskin 1994). However, the other rows of column 1 show that these indices are not much reduced when we look at relatively homogeneous groups of workers. For example, row 9 shows that the national Gini coefficient among laborers is .90 in the WECD, even higher than the Gini coefficient for all workers. Similarly, the Gini coefficient among construction-industry workers is .88 in the CBO, which is again even higher than that observed among all workers. Thus, by this traditional measure, the U.S. workforce is quite segregated.

Are these patterns evidence of *systematic* sorting of blacks and whites to different employers? Column 2 of table 3 addresses this question by reporting the Gini coefficient of random segregation, which again is the Gini coefficient that would arise if workers were randomly allocated to employers. Inspection shows that these random Gini coefficients are often quite high. For example, the random Gini coefficient for the entire sample is .48 for the WECD and .25 for the CBO, and generally even higher in the subsamples.[20] This suggests that much of the unevenness measured by the Gini coefficient is potentially attributable to random allocation rather than to systematic forces such as employer discrimination. However, the systematic Gini coefficients of column 3 show that there is an important systematic component to segregation in the United States. For example, the systematic Gini coefficient of .57 in row 1 implies that actual excess unevenness ($G - G^*$) is 57% of the maximum that could possibly be observed ($1 - G^*$). Similarly, the systematic Gini coefficient is substantial within each industry of the WECD and the CBO and for most of the other subgroups. Yet there are some subgroups in which the systematic Gini coefficient is not substantially different from zero. For example, the systematic Gini coefficient in the WECD is −.03 for professionals, technicians, and managers and .06 for sales and service occupations. Thus, for these subsamples, the traditional Gini coefficient is quite close to what would be implied by random allocation. These results are consistent with the notion that the market for such skilled labor is relatively nationalized and, therefore, less likely to reflect historical racial differences in residential patterns.

Much of the segregation measured in columns 1–3 of table 3 is due to the dissimilar distributions of blacks and whites across states and MSAs. For example, blacks are relatively likely to live in the South and in the central cities of the North and relatively unlikely to live in states like Colorado, Utah, and Montana. To the extent that such patterns reflect historical and cultural factors not directly related to the contemporary labor market, columns 1–3 do not tell us much about current labor

[20] The role of random allocation varies across samples for two reasons. First, random allocation generates more unevenness when the black sample share is small. This leads CBO indices to be smaller than those in the WECD. Second, the role of random allocation varies with the number of sample employees per firm, which varies considerably across samples. In particular, note that per-firm sample sizes are much reduced when we restrict attention to the WECD subsamples stratified by worker characteristics. This causes the Gini coefficients of random segregation for these samples to be higher than those observed in the entire data set. It may seem odd that the expected Gini coefficient has a standard error, but this too is stochastic because the expected Gini coefficient takes the empirical firm size distribution as fixed, and this distribution varies across bootstrap replications.

market segregation. Therefore, columns 4–6 of table 3 take a more refined geographic approach by measuring segregation *within* MSAs. In particular, we computed our three Gini coefficients separately for each MSA and then reported the mean of these indices across MSAs.[21] While this measure may still reflect within-MSA residential segregation, it is not affected by the uneven distribution of blacks and whites across MSAs.

The traditional Gini coefficients of column 4 of table 3 show that there is substantial unevenness within the typical MSA. For example, the average within-MSA Gini coefficient among all workers is .60 in the WECD and .81 in the CBO. However, it is not always clear that this within-MSA unevenness represents a substantial departure from random allocation. For example, column 5 of row 1 shows that the within-MSA Gini coefficient of random segregation is .56 in the WECD as a whole, which implies that observed unevenness within MSAs is only marginally greater than that implied by random allocation. In fact, in none of the rows of panel A is workplace segregation substantially greater than that predicted by random allocation. Thus, there is little evidence in the WECD of any important systematic component to within-MSA segregation. In contrast, there is substantially more unevenness than random allocation would imply in the CBO. Row 14 shows that in these data the traditional Gini coefficient is .81 within MSAs, while random allocation implies an index of .28.

The other rows of columns 4–6 analyze racial segregation within relatively homogenous groups of workers. These specialized results are analogous to the general results of rows 1 and 14: there is systematic segregation within-MSAs in the CBO but not in the WECD. In particular, there are several groups for which there is significantly *less* within-MSA segregation in the WECD than random allocation would predict. For example, the "sales and service" occupation within manufacturing has a Gini coefficient of .82, whereas random allocation predicts a Gini coefficient of .89. In contrast, the within-industry CBO estimates of rows 15–20 show that there is significantly more unevenness than random allocation would imply in these data, particularly in the service and trade industries.

Why do we find more systematic segregation in the CBO than in the WECD? The difference is partly driven by the CBO's broader industrial coverage, as the systematic Gini coefficients for CBO manufacturing industries are substantially lower than those of other CBO industries. The difference is also driven by the greater size of the WECD plants. Rows 10–13 of table 3 show that there is more systematic segregation among smaller plants in the WECD, and these plants are closer in size to those

[21] Note that the move to an MSA-based sample excludes the roughly 28% of whites and 17% of blacks who do not live in an identifiable MSA.

typical of the CBO. Thus, the comparisons both across data sets and within the WECD suggest that large firms are less systematically segregated. Finally, the difference may also be driven by the greater frequency of family-owned firms in the CBO, as owners of such businesses are probably more likely to hire members of the same ethnic group. In contrast, the difference between the CBO and the WECD is apparently not driven by the CBO sample's focus on the South. We base this conclusion on within-MSA indices calculated separately by region in the WECD.[22] These figures show that segregation is, if anything, less systematic in the South than in the rest of the country.[23]

The WECD results of table 3 show that, once we restrict attention to within-MSA patterns, black and white workers are not systematically segregated. Nevertheless, the CBO results suggest that perhaps there is some systematic component to these patterns. To what extent is this systematic interfirm segregation the outcome of "spatial mismatch," which occurs when (*a*) blacks and whites are *residentially* segregated within MSAs and (*b*) all workers tend to find jobs near their home (Wilson 1987)? Table 4 addresses this question by examining segregation within MSAs where spatial mismatch is likely to be a minor factor. In particular, for spatial mismatch of black workers and jobs to explain workplace racial segregation, it must be that blacks and whites are residentially segregated. Thus, looking at MSAs where there is little residential segregation is one way to reduce the effect of within-MSA spatial mismatch.

In this spirit, columns 1–3 of table 4 repeat the within-MSA analysis of table 3 for a sample of MSAs that are relatively residentially integrated. In particular, we used the residential Gini coefficients of Harrison and Weinberg (1992) to restrict the sample to MSAs below the median of the distribution of MSA (residential) Gini coefficients. The results for the WECD are similar to those of table 4, as there is little evidence that within-MSA segregation is more pervasive than random allocation would imply. If anything, the WECD results suggest that there is *less* segregation than random allocation would predict. In contrast, the CBO results differ slightly from table 3. It is still true that the traditional Gini coefficient for all workers in the CBO (.70) is higher than that implied by random allocation (.28). However, when we restrict attention to industries where

[22] The average within-MSA systematic Gini coefficients were .13 for the Northeast, .13 for the Midwest, .06 for the South, and −.29 for the West.

[23] We are aware of one factor that works in the opposite direction. Since the WECD is a sample of establishments while the CBO is a sample of firms, establishment level segregation *within* firms will be missed in the CBO but not in the WECD. This factor appears to be overwhelmed by the other forces creating more segregation in the CBO.

workers are likely to be relatively homogenous, there is no longer much evidence of systematic segregation. In fact, there is some evidence that the distribution of workers is *less* uneven than randomness would imply. For example, among workers in nondurable manufacturing, the average Gini coefficient within MSAs is .22, while random allocation implies a Gini coefficient of .44. Only in the construction and service industries are black and white workers systematically segregated in an important way.[24]

Of course, even "relatively residentially integrated MSAs" are rather segregated. For example, Bloomington, Indiana, is a relatively integrated city, but its residential Gini coefficient is still .342. Therefore, an alternate approach to minimizing the role of spatial mismatch is to measure segregation in relatively small MSAs. While living on one side of the Los Angeles MSA may preclude holding a job on the other side, there are smaller MSAs where even a cross-town commute takes only a few minutes. This reasoning suggests that spatial mismatch is less likely to explain workplace segregation in small MSAs. Therefore, columns 4–6 of table 4 repeat the within-MSA analysis for MSAs below the median of the MSA population distribution. The results are quite similar to those of columns 1–3 of the same table. Namely, there is some evidence of systematic segregation among all workers in the CBO. For all other samples from both data sets, however, there is no evidence that black and white workers are distributed across employers more unevenly than would be suggested by random allocation.[25]

The results of this section suggest the following view. The national distribution of black and white employees across employers is far from even, as some employers have predominantly white workforces while others are predominantly black. However, systematic national segregation is largely due to black/white differences in MSA residence. When we look within MSAs, there is still substantial interfirm segregation by conventional measures, but this is mostly explained by racial differences in occupations and industry and by simple random allocation. Of the modest amount of within-MSA segregation that remains unexplained by random

[24] We should note that the samples of both data sets become smaller as we restrict the sample to particular MSAs or to particular groups of workers. Thus, our power to distinguish between alternative hypotheses is reduced in these subsamples, as reflected in the increased standard errors.

[25] We also made an attempt to explain the cross-MSA variation in interfirm segregation. In particular, we regressed both the standard and the systematic Gini coefficients on log MSA population and on the residential Gini coefficient. We did this for both the CBO and the WECD. Few consistent patterns emerged, although MSAs with large residiential Gini coefficients tended to have larger systematic Gini coefficients of interfirm segregation. This provides some support for the spatial mismatch hypothesis.

allocation and group skill differences, much appears to be attributable to spatial mismatch.[26] However, it is worth noting that while the data are *consistent* with random allocation in many ways, it is entirely possible that there are strong systematic forces (e.g., discrimination) at work. All we can really say is that the data do not radically reject the hypothesis of random allocation.

These results bear comparison with the earlier work of Flanagan (1973), Higgs (1977), and Becker (1980). Higgs (1977) found substantial racial segregation in a sample of Virginia firms that were surveyed in 1900 and 1909. However, data constraints prevented Higgs from assessing the importance of intrastate residential segregation, and Higgs's analysis did not reduce patterns of segregation into a single summary index. These facts make it difficult to compare his results directly with those presented here. Flanagan (1973) studied segregation within a sample of Chicago firms surveyed in 1967 as part of the Equal Employment Opportunity Commission's (EEOC) monitoring process. As with the Higgs study, methodological differences preclude a numerical comparison with our study. However, Flanagan notes that "observed segregation of blacks is closer to the random (allocation model) than to the general equilibrium prediction of the utility analysis" (p. 468); that is, a random allocation model did a reasonable job of explaining the data. Finally, Becker (1980) analyzed segregation in a nationwide sample of employers drawn from the 1975 EEOC data. He too finds substantial segregation, but the results of his paper are particularly difficult to compare with those presented here. This is because the data are not disaggregated by MSA, because the paper used a somewhat exotic index, and because the role of random allocation was not discussed. In sum, previous studies have found substantial interfirm racial segregation in years past, at least by traditional measures, but the evidence is mixed as to whether the data were inconsistent with random allocation. Unfortunately, methodological differences preclude a more direct quantitative comparison between them and the present study.

[26] Since neither the CBO nor the WECD is a random sample of all employers or employees, we wondered whether these figures were representative of the broader population. The National Survey of Black Americans (NSBA) is of some assistance in this regard (Jackson and Gurin 1987). Administered in 1979 and 1980, the NSBA asked a national sample of blacks about the fraction of blacks in their "workplace," which was probably interpreted as the worker's establishment but could also have been interpreted as their firm. In any case, their answers were tabulated as follows: all black (17.7%), mostly black (26.3%), about half black (20.6%), mostly white (23.6%), all white except you (11.7%). While the qualitative nature of the NSBA question and the lack of firm size information preclude a precise comparison, the figures appear to be closer to those of the CBO than the WECD.

## V. Are Black and White Workers Sorted into Different Types of Employers?

Section IV showed that the distribution of black and white workers across firms is not too far from what would be implied by random allocation. Yet it is important to recognize that the observed patterns could be systematic even though they look random. In an effort to look for systematic patterns, we present in table 5 estimates of a WECD establishment-level regression where the dependent variable is the black share of *nonsupervisory* employment in the establishment.[27] Column 1 includes as regressors the share of black *supervisors* in the establishment, the black sample share within each MSA, the log of establishment employment, the average age and education of nonsupervisory employers, and dummy variables for industry and region. Column 2 adds a set of MSA dummy variables and takes out the black share of MSA population and the region dummies. The regressions show that there is a statistically strong relationship between the race of supervisors and nonsupervisors; black workers tend to be supervised by black managers. This relationship is nowhere near one-for-one, however, as managers of one race often supervise workers of another.

The correlation between the race of managers and the race of nonmanagers may be explained by residential segregation, so columns 3 and 4 present two crude ways of trying to minimize its role. In particular, column 3 interacts percent black supervisors with the *residential* Gini coefficient in the MSA, and column 4 interacts percent black supervisors with the log of MSA population.[28] Our earlier arguments suggest that if residential segregation is responsible for the correlation between the race of supervisors and nonsupervisors, then the coefficient on percent black supervisors should be bigger in large or segregated MSAs. The interaction terms in columns 3 and 4 are small and, in the case of column 3, statistically insignificant. While obviously not conclusive, these results suggest that residential segregation is not an attractive explanation for the interplant connection between black supervisors and black nonsupervisors in the WECD.

We pursue a similar analysis in the CBO. The CBO records the fraction of nonminority employees within six brackets (0%–25%, 26%–50%,

[27] To be in this sample, an establishment had to have at least one supervisor and one nonsupervisor in the WECD. This led to a 50% reduction in the number of plants in our sample, as the smallest establishments seldom met this restriction. Regressions are weighted by total employment in the plant.

[28] Note that this specification results in a smaller sample because we throw out non-MSA establishments. The slight change in the coefficient on percent black supervisors between cols. 1 and 2 and cols. 3 and 4 is driven by the change in specification rather than the change in sample.

**Table 5**
**Plant-Level OLS Models of Employee Racial Composition: WECD Dependent Variable = Black Share of Nonsupervisory Employment**

| Independent Variable | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 1. Percent black supervisors: | .261 | .301 | .178 | .179 |
| | (.013) | (.014) | (.013) | (.013) |
| × residential Gini coefficient in MSA | ... | ... | .001 | ... |
| | | | (.001) | |
| × log MSA population | ... | ... | ... | .001 |
| | | | | (.000) |
| 2. Percent black population in MSA sample | .873 | ... | 1.056 | 1.054 |
| | (.024) | | (.024) | (.024) |
| 3. Log of establishment employment | .003 | .005 | .003 | .003 |
| | (.001) | (.001) | (.001) | (.001) |
| 4. 2-digit industry dummies | yes | yes | yes | yes |
| 5. Region dummies | yes | no | yes | yes |
| 6. MSA dummies | no | yes | no | no |
| 7. Number of plants in sample | 7,813 | 7,813 | 6,562 | 6,562 |
| 8. $R^2$ | .386 | .426 | .404 | .405 |

NOTE.—All data are drawn from the WECD. Standard errors are in parentheses. All regressions included controls for the average age and education of nonsupervisory employees, industry, and region. To be in the sample for this table, establishments had to both have at least one supervisor and one nonsupervisor in the WECD. Regressions are weighted by total employment in the plant.

51%–75%, 76%–90%, 91%–99%, and 100%), making it natural to apply the ordered probit model.[29] Table 6 presents estimates of a series of such models for the CBO where the dependent variable is the fraction of black employees in the firm.[30] The specification in column 1 includes the race and sex of the owner, the log of firm employment, the black population share of the MSA, and a set of controls for the minority share of a firm's customers. The specification indicates that the race of the owner has a fairly strong influence on the minority share of the firm's workforce. Given the estimated cut points, the black owner coefficient is sufficient to move the median predicted response of a firm from a 10%–24% black workforce to one that is 50%–74% black. The relationship between race of owner and race of workers is estimated to be particularly strong for male owners. Row 6's coefficients for percent minority customers are relative to the omitted group of firms that did not report this information. The results indicate a strong relationship between the race of a firm's customers and the race of its employees, even after we control for the race of its owner.

The rest of table 6 presents alternative specifications of the model. Column 2 adds MSA dummies and takes out the percent black population

[29] Regressions are weighted by the product of each observations sampling weight and its total employment.

[30] These results are similar to Bates's (1988) analysis of the 1982 CBO.

**Table 6**
**Firm-Level Ordered Probit Models of Employee Racial Composition: CBO**

| Independent Variable | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 1. Black owner: | 1.293 | 1.320 | 1.353 | 1.347 |
| | (.137) | (.139) | (.160) | (.160) |
| × MSA Gini coefficient | . . . | . . . | .056 | . . . |
| | | | (1.267) | |
| × log MSA population | . . . | . . . | . . . | .098 |
| | | | | (.123) |
| × female owner | −.508 | −.450 | −.497 | −.488 |
| | (.237) | (.240) | (.271) | (.270) |
| 2. Female owner | −.072 | −.081 | −.200 | −.198 |
| | (.099) | (.103) | (.116) | (.116) |
| 3. Log(firm employment) | .132 | .148 | .172 | .172 |
| | (.012) | (.013) | (.014) | (.014) |
| 4. Black share of MSA sample population | 2.992 | . . . | 1.605 | 1.646 |
| | (.938) | | (1.015) | (1.016) |
| 5. MSA dummies | no | yes | no | no |
| 6. Minority customer share: | | | | |
| 75%–100% | 1.195 | 1.473 | 1.152 | 1.152 |
| | (.059) | (.063) | (.071) | (.071) |
| 50%–74% | 1.051 | 1.221 | 1.063 | 1.061 |
| | (.086) | (.089) | (.110) | (.110) |
| 25%–49% | .584 | .681 | .374 | .375 |
| | (.058) | (.061) | (.068) | (.068) |
| 10%–24% | .240 | .308 | .126 | .126 |
| | (.049) | (.053) | (.057) | (.057) |
| 1%–9% | −.207 | −.134 | −.264 | −.264 |
| | (.047) | (.050) | (.055) | (.055) |
| 0% | −.332 | −.275 | −.530 | −.530 |
| | (.057) | (.061) | (.069) | (.069) |
| 7. Cut Points: | | | | |
| 0% black → 1%–9% black | 0 | 0 | 0 | 0 |
| 1%–9% black → 10%–24% black | .321 | .341 | .342 | .342 |
| 10%–24% black → 25%–49% black | .668 | .712 | .713 | .713 |
| 25%–49% black → 50%–74% black | 1.109 | 1.182 | 1.253 | 1.253 |
| 50%–74% black → 75%–100% black | 1.535 | 1.634 | 1.701 | 1.701 |
| 8. Number of firms in sample | 6,043 | 6,040 | 4,149 | 4,149 |

NOTE.—All data are drawn from the 1987 CBO. Standard errors are in parentheses. All regressions also included controls for log of establishment employment, 2-digit industry, region (except col. 2), and the age and educational attainment of the business owner. In row 6, the left-out group is those firms for whom minority customer share was missing. Regression weights are the product of each observation's sample weight and its employment. Regression weights were then scaled to sum to sample size.

in the MSA, which increases the coefficient on black ownership and slightly strengthens the relationship between the race of customers and the race of employees. Thus, the racial nexus between owners, employees, and customers is not solely due to the distribution of blacks and whites across MSAs. As before, spatial mismatch arguments suggest that the relationship between the race of owners and customers and the race of employees is stronger in large, segregated MSAs. Thus, columns 3 and 4 interact the MSA residential Gini coefficient and the log MSA population, respectively, with the black owner variable. As in table 5, there is little difference between those MSAs with high and low residential Gini coefficients.

The results of this section suggest that there is a systematic component to the sorting of black and white workers across establishments in our two data sets. In the WECD, black nonsupervisors are more likely than whites to work for black supervisors, and vice versa. In the CBO, black workers are much more likely to work for firms with black owners and customers. If one accepts the view that black supervisors, owners, and customers are less likely than whites to discriminate against black employees, then these results are consistent with the hypothesis that tastes for discrimination cause some systematic workplace segregation. Of course, it is also possible that these correlations are driven by group differences in skill or by residential segregation. The available data do not support these hypotheses, but the analysis is sufficiently crude that we can not rule out their importance.

Section IV found that the distribution of workers to firms in the WECD was roughly consistent with random allocation, while this section documents a modest systematic component to the matching of black and white workers to particular firms. There are two ways to reconcile these results. The first is that, although black employers are more likely to hire black workers, there simply are not enough black employers for this to generate much systematic segregation. This is particularly true in the WECD, where the black share of supervisors has an interplant mean of .026 and an interplant variance of .013. Applying these supervisor shares and the parameters of table 5 to an initially random distribution of black and white workers leads to a trivial amount of segregation. In contrast, the supervisor/nonsupervisor relationship can generate nontrivial segregation in our CBO sample, as there are a relatively large number of "black" firms (i.e., those with black owners) in that data set. This difference between the two databases may be another reason why we find the CBO to be relatively segregated.

The results may also be explained by recognizing that, while employer tastes may induce systematic segregation, other forces might induce systematic *integration.* For example, if black and white labor are complementary inputs, then firms might systematically integrate their workforces even in the presence of discriminatory attitudes. Title VII and affirmative action provide similar incentives to integrate.[31] Thus, the apparent randomness of the distribution of black and white workers may represent a

[31] We explored this idea in the following way. For a subsample of plants in the WECD and for most firms in the CBO, we know the fraction of sales made to the federal government. In these samples, we reestimated the regressions of tables 5 and 6 including the federal government's share of sales as a regressor. In both databases, this added coefficient was significantly positive but not large.

balance between several systematic forces, each of which is rather weak.[32] This brings up an important point, namely that Section IV only demonstrated that the data were roughly consistent with random allocation. We reiterate that this may reflect a balance between several not-well-understood systematic forces that tend to cancel each other.

## VI. Interfirm Segregation and the Black/White Wage Gap

This section investigates two aspects of the relationship between interfirm segregation and the black/white wage gap. Our analysis in this section is based solely on the WECD because it records wages and personal characteristics for each worker individually. In contrast, the CBO only records a firm's total payroll without providing information on how wages are distributed to individual workers, making it relatively unsuitable for the study of black/white wage differences.[33] It bears repeating that the gap between black and white wages is smaller in the WECD than in the economy at large. For example, the difference in mean log hourly wages between blacks and whites is .21 among manufacturing workers in the 1990 Decennial Census, while the analogous gap in the WECD is only .13. Even bigger differences arise if we compare the WECD to the nonmanufacturing component of the Decennial Census.[34] Thus, the ensuing analysis considers a segment of the economy where

[32] The following analogy may help explain our results. Suppose that you observe black and red checkers strewn across a checkerboard and that you are interested in whether there is any systematic pattern to their distribution. Initially suppose that you observe the color of the checkers but not the color of the checkerboard squares on which they lie and that the distribution of checkers looks roughly random. With a relatively small number of checkers per square, this is consistent with some squares having only red checkers and some squares having only black checkers. Then suppose that you observe the color of the checkerboard squares as well as the color of the checkers, and that you find that the squares with only red checkers tend to be red themselves, and that the squares with black checkers tend to be black. This would now suggest some sort of systematic phenomenon. This situation is analogous to what we find in the distribution of workers to firms. When we ignore the "race" of the firms (i.e., the color of the checkerboard squares), we find nothing systematic in the distribution of workers (i.e., the checkers). But when we look at the matching of workers to firms (i.e., the matching of checkers to squares), then systematic patterns emerge.

[33] Carrington and Troske (1995) combines information from the CPS and the CBO to estimate the within- and between-firm components of the wage gap between two groups. However, since this procedure requires difficult-to-verify assumptions on the compatibility of the CBO and the CPS, our analysis here concentrates on the WECD.

[34] The smaller black/white wage gap in the WECD reflects the fact that (*a*) the WECD contains mostly large manufacturing plants and that (*b*) the black/white wage gap is partially accounted for by whites' greater representation in such plants.

black and white earnings are relatively similar, and our results may not apply directly to the broader economy. Nevertheless, there is a substantial black/white wage gap in the WECD, and we believe that a better understanding of it has implications for the rest of the economy.[35]

Our first exercise decomposes the black/white wage gap into a between- and within-plant component. In particular, we regress wages on a set of plant fixed effects before, after, or at the same time that we control for workers' personal characteristics in order to see how much of the black/white wage gap can be explained by the location of black and white workers in different plants. We do this for all workers at one time, but we also perform separate analyses for men and women because the size of the black/white wage gap varies so much by sex. Let $Y$ = log hourly wages; $X$ = a set of personal characteristics including terms in experience, education, sex, marital status, and dummies for occupation, industry, and region; and $Z$ = a set of plant fixed effects. Columns 1–3 of table 7 then report results from a two-step procedure in which we first estimated $Y = X'\beta$ and then regressed the residuals of this first regression on the plant fixed effects $Z$. Column 1 reports results for all workers, while columns 2 and 3 report results for men and women separately. Row 1 shows that the unadjusted difference in log hourly wages between blacks and whites is 12.7% for all workers, 12.5% for men, and 0.5% for women.[36] (While we include full results for women, the fact that there is virtually no black/white wage gap among women leads us to focus our remaining discussion of table 8 on the results for men.)[37]

Row 2 of table 7 reports the residual wage gap that is left after we control for the effect of personal characteristics. This is approximately 2% for all workers and 5% for men, which are somewhat smaller residual wage gaps than one typically sees in these sort of data. Row 3 reports the residual black/white wage gap after controlling for the fixed effects in the second step, *without* controlling for personal characteristics, and row 4 reports the residual wage gap after we control for both the personal

[35] It bears emphasis, however, that patterns of wage variation in the WECD are generally *not* pathological. An appendix table (available from us on request) presents selected coefficients from a log wage regression estimated on our WECD sample. The results are generally quite similar to those obtained from a similar regression estimated on all manufacturing workers in the 1990 Decennial Census.

[36] We restricted the sample for this analysis to firms where we had at least three workers matched to the establishment. The fact that we imposed this requirement at each stage of the analysis (e.g., to be in the men's sample a firm had to have three men in our sample) means that the total sample is not the union of the male and female samples. Thus, there is no requirement that the total wage gap of col. 1 lie between the men's and women's wage gaps of cols. 2 and 3.

[37] The rough equality of the wages of black and white women is not unique to the WECD. Similar results obtain from the Decennial Census or the CPS.

**Table 7**
**Decomposing the Black/White Wage Gap into Within- and Between-Plant Components**

| | Order of the Decomposition | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Step 1: $Y = X'\beta + u_1$ Step 2: $Y - X'\hat{\beta} = Z'\gamma + u_2$ | | | Step 1: $Y = Z'\gamma + u_1$ Step 2: $Y - Z'\hat{\gamma} = X'\beta + u_2$ | | | Step 1: $Y = X'\beta + Z'\gamma + u_2$ | | |
| | Total (1) | Men (2) | Women (3) | Total (4) | Men (5) | Women (6) | Total (7) | Men (8) | Women (9) |
| 1. Unadjusted black/white wage gap $\bar{Y}_w - \bar{Y}_b$ | .127 | .125 | .005 | .127 | .125 | .005 | .127 | .125 | .005 |
| 2. Black/white wage gap adjusted for personal characteristics $(\bar{Y}_w - \bar{X}'_w\hat{\beta}) - (\bar{Y}_b - \bar{X}'_b\hat{\beta})$ | .019 | .049 | −.036 | .052 | .074 | −.015 | .037 | .053 | −.021 |
| 3. Black/white wage gap adjusted for plant fixed effects $(\bar{Y}_w - \bar{Z}'_w\hat{\gamma}) - (\bar{Y}_b - \bar{Z}'_b\hat{\gamma})$ | .137 | .134 | .043 | .118 | .123 | .031 | .137 | .131 | .055 |
| 4. Black/white wage gap adjusted for personal characteristics and plant fixed effects $(\bar{Y}_w - \bar{X}'_w\hat{\beta} - \bar{Z}'_w\hat{\gamma}) - (\bar{Y}_b - \bar{X}'_b\hat{\beta} - \bar{Z}'_b\hat{\gamma})$ | .029 | .058 | .002 | .043 | .063 | .011 | .045 | .059 | .027 |

NOTE.—$Y$ = hourly wages. $X$ = worker characteristics including flexible terms in experience and education, sex (in cols. 1, 4, and 7), marital status, sex × marital status, 10 occupation dummies, 4-digit industry dummies, MSA, region, MSA × region. $Z$ = a set of plant fixed effects. In addition to our previous restrictions, we required that each individual be in a plant with at least three people in our sample. This reduced the sample size by approximately 10%. The remaining samples were (*a*) 172,056 for the total columns, of whom 160,598 were white and 11,458 were black, (*b*) 121,056 for the male sample, 114,076 of whom were white and 6,980 of whom were black, and 44,997 for the female sample, of whom 40,866 were white and 4,131 were black. Observations in the male and female samples do not equal those of the total sample because we imposed the "must have three people in the sample" restriction for each sample. Thus, some plants had three individuals without having either three women or three men in our sample.

characteristics and the plant fixed effects. Rows 3 and 4 suggest that, if anything, black men tend to work in plants that pay slightly above-average wages once we control for personal characteristics.[38]

Columns 4–6 reverse the order of the decomposition by regressing wages on the plant fixed effects in the first step and then regressing the residuals on the personal characteristics in the second step. In effect, this gives the plant effects first crack at explaining the black/white wage gap.

[38] It is worth noting that the plant effects alter black/white wage comparisons for women much more than they do for men. Black and white women have similar average wages in the WECD, but this is a combination of two effects: black women earn somewhat less than white women within any given plant, but this is compensated for by the fact that black women tend to work in plants with higher average salaries.

Row 3 shows that the black/white wage gap that remains after controlling for plant is .118 for all workers, or about 93% of what it was without controlling for these effects. For men, column 5 shows that plant fixed effects explain virtually none of the black/white wage gap. Columns 7–9 regress *Y* on *X* and *Z* simultaneously, so that personal characteristics and plant fixed effects are given equal opportunity to explain the black/white wage gap. The results are similar to those of the previous columns, as black men (and women) are sorted into what are, if anything, high-paying plants.

These results suggest that the black/white wage gap is primarily a *within-plant* phenomenon.[39] In contrast, very little of the black/white wage gap in our data is accounted for by the allocation of black and white workers to firms that pay systematically different salaries. Most of the within-plant pay gap is accounted for by observable characteristics, but a significant component remains unaccounted for. At least in the sample of manufacturing workers studied here, the problem is not that blacks are not getting jobs at the "good" plants but, rather, that they receive lower wages on average within any given plant.[40]

Our second exercise relates the wages of black and white workers to the racial makeup of their coworkers.[41] Table 8 presents estimates of individual-level hourly wage regressions with personal and plant characteristics on the right hand side.[42] The personal characteristics include terms in experience and education, sex, marital status, occupation, and race, and the plant characteristics include total employment and the black share of establishment employment. Columns 1 and 2 present results when we combine men and women into one regression. The regressions differ only in that column 1 includes a plant-level control for labor productivity (defined as value of shipments/employment), while column 2 does not. The coefficients on most regressors are consistent with similar data sets, so we will not discuss them. The unique results are in rows 4

[39] This result is consistent with the findings of Hellerstein, Neumark, and Troske (1996).

[40] To the extent that we could, we conducted a similar analysis in the CBO. There are two problems with the CBO. First, we only know the annual earnings distributed to the firm's employees and not their hourly wages, and second, we only know aggregate earnings and not how they were distributed among the employees. Nevertheless, if we assign firm average earnings to each worker, then white earnings exceed black earnings by roughly 10% in the CBO. It is impossible to know whether this cross-firm racial difference in *earnings* is due to analogous differences in *wages,* and to what extent it is due to black/white differences in hours worked.

[41] See Ragan and Tremblay (1988) for a similar analysis.

[42] The standard errors in table 8 have been corrected for heteroscedasticity and for the clustered sampling design.

**Table 8**
**Individual-Level Models of Wage Determination: WECD**
Dependent Variable = Log Hourly Wage

| | Sample | | | | | |
|---|---|---|---|---|---|---|
| | All Workers | | Men | | Women | |
| Independent Variable | (1) | (2) | (3) | (4) | (5) | (6) |
| 1. Black | −.008 | .001 | −.006 | .003 | .051 | .056 |
| | (.009) | (.010) | (.010) | (.011) | (.011) | (.011) |
| 2. Female | −.145 | −.145 | . . . | . . . | . . . | . . . |
| | (.006) | (.006) | | | | |
| 3. Black × female | .080 | .076 | . . . | . . . | . . . | . . . |
| | (.009) | (.009) | | | | |
| 4. Black share of establishment employment | .158 | .197 | .177 | .223 | .116 | .136 |
| | (.027) | (.032) | (.035) | (.040) | (.033) | (.035) |
| 5. Black × black share of establishment employment | −.349 | −.387 | −.370 | −.412 | −.295 | −.319 |
| | (.035) | (.038) | (.046) | (.051) | (.047) | (.048) |
| 6. Log(establishment employment) | .062 | .062 | .058 | .058 | .071 | .072 |
| | (.003) | (.004) | (.004) | (.004) | (.004) | (.004) |
| 7. Labor productivity (× 1,000) | .200 | . . . | .188 | . . . | .236 | . . . |
| | (.027) | | (.029) | | (.038) | |
| 8. $R^2$ | .40 | .40 | .37 | .37 | .24 | .23 |
| 9. Number of observations | 168,125 | 168,125 | 120,574 | 120,574 | 47,551 | 47,551 |

NOTE.—Each regression also included controls for 4-digit industry, 1-digit occupation, MSA, region, MSA × region, marital status, marital status × sex, five education categories, a quartic in experience, and the interaction of the education and experience terms. Standard errors are in parentheses. All data are drawn from the WECD. Labor productivity is defined as the dollar value of shipments divided by establishment employment.

and 5, which relate wages to the black share of sample employment within each worker's establishment. Since this share is interacted with race, the direct coefficient in row 4 indicates that wages of white workers are increasing in the black share of their coworkers. For example, the results in column 1 imply that a white worker in a plant that is 50% black earns roughly 8% higher wages than an observationally equivalent worker in an all-white plant. Columns 3–6 present analogous regressions for men and women separately. The results are quite similar to those of column 1, although of course the black/white wage gap among women is, if anything, positive. The wages of white workers increase with the black share of establishment employment, while the wages of black workers decrease along the same dimension. How does one reconcile these results with those of table 7, which demonstrated that the black/white wage gap is primarily a within-plant phenomenon? The answer is that, while black-majority plants pay their black employees less than plants where blacks are a minority, these same plants tend to pay their white workers more. On balance, this implies that plants with substantial black minorities do not pay particularly low wages on average but that the black/white wage gap is biggest within such plants.

## VII. Conclusions

This article has shown that interfirm segregation within the average MSA is close to what would be expected by the random allocation of workers to firms; that blacks are more likely to work at firms with black owners, black supervisors, and black customers; and that the black/white wage gap is predominantly a within-plant phenomenon. We would like to briefly speculate about the implications of our results for theories of black/white wage differentials. Becker's (1957) theory of discrimination and Wilson's (1987) spatial mismatch hypothesis are two of the most influential theories of black economic disadvantage, and both imply that blacks and whites will be systematically segregated in the workplace. The fact that we do not find much systematic segregation in our data provides a modest challenge to the ability of these theories to explain the black/white wage gap. The results are particularly damaging to the spatial mismatch hypothesis, as the problem does not appear to be that blacks are spatially separated from high-paying employers but, rather, that they receive low pay even when they work with whites. We also view the results as damaging to the formulation of Becker's theory where whites have a distaste for *physical* proximity to blacks. In contrast, the results may be fully consistent with a version of Becker's theory in which whites have a distaste for certain types of *social* proximity to blacks, such as having a black boss or coworker of equal rank. However, we have relatively little to say about such within-firm segregation.

There are some caveats to these conclusions associated with our data. Both of the databases analyzed here are special in some way, and thus the results found here may not apply to the broader population of firms, establishments, and workers. We hope that a more representative database suitable for the study of interfirm segregation will be developed in the near future. Until that date, however, our results will have to stand as is.

## References

Bates, Timothy. "Do Black-Owned Businesses Employ Minority Workers? New Evidence." *Review of Black Political Economy* 17 (1988): 51–60.

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

Becker, Henry Jay. "Racial Segregation among Places of Employment." *Social Forces* 58 (1980): 761–75.

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

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

Boisso, Dale; Hayes, Kathy; Hirschberg, Joseph; and Silber, Jacques.

"Occupational Segregation in the Multidimensional Case." *Journal of Econometrics* 61 (1994): 161–71.

Carrington, William J., and Troske, Kenneth R. "Gender Segregation in Small Firms." *Journal of Human Resources* 30 (Summer 1995): 503–33.

———. "On Measuring Segregation in Samples with Small Units." *Journal of Business and Economic Statistics* 15 (October 1997): 402–9.

Duncan, Otis, and Duncan, Beverly. "A Methodological Analysis of Segregation Indices." *American Sociological Review* 20 (1955): 210–17.

Ellison, Glenn, and Glaeser, Edward L. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (October 1997): 889–927.

Flanagan, Robert J. "Racial Wage Discrimination and Employment Segregation." *Journal of Human Resources* 8 (1973): 456–71.

Harrison, Roderick J., and Weinberg, Daniel H. "Racial and Ethnic Residential Segregation in 1990." Paper presented at Population Association of America meetings, May 1992.

Hellerstein, Judith; Neumark, David; and Troske, Kenneth. "Wages, Productivity, and Worker Characteristics." Working Paper no. 5626. National Bureau of Economic Research, Cambridge, MA: 1996.

Higgs, Robert. "Firm-Specific Evidence on Racial Wage Differentials and Workforce Segregation." *American Economic Review* 67 (1977): 236–45.

Hutchens, Robert M. "Segregation Curves, Lorenz Curves, and Inequality in the Distribution of People across Occupations." *Mathematical Social Sciences* 21 (1991): 31–51.

Jackson, James S., and Gurin, Gerald. *National Survey of Black Americans, 1979–1980* (machine-readable data file). 1st ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR), 1987.

Kremer, Michael, and Maskin, Eric. "Segregation by Skill and the Rise in Inequality." Working paper. Cambridge, MA: Harvard University, 1994.

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

Troske, Kenneth R. "The Worker-Establishment Characteristics Database." In *Labor Statistics Measurement Issues,* edited by John Haltiwanger, Marilyn Manser, and Robert Topel. Chicago: National Bureau of Economic Research, in press.

Wilson, William Julius. *The Truly Disadvantaged.* Chicago: University of Chicago Press, 1987.