Workplace Segregation in the United States: Race, Ethnicity, and Skill*

Judith Hellerstein
Department of Economics and MPRC,
University of Maryland,
and NBER

David Neumark
Public Policy Institute of California,
and NBER

May 2004

Abstract: We study workplace segregation in the United States using a unique matched employer-employee data set that we have created. We first present measures of workplace segregation by race and ethnicity, using simulation methods to measure segregation beyond what would occur randomly as workers are distributed across establishments. We then assess the role of skill differentials in generating workplace segregation, as skilled workers may be more complementary with other skilled workers than with unskilled workers, and skill is often correlated with race and ethnicity. Specifically, we measure the extent of segregation by education level and quality, and by language ability, providing informative contrasts with measures of segregation by race and ethnicity. Finally, we attempt to distinguish between segregation by skill based on general crowding of unskilled poor English speakers into a narrow set of jobs, and segregation based on common language for reasons such as complementarity among workers speaking the same language. Our results consistently indicate that workplace segregation by race and ethnicity is driven in large part by skill differences across workers.

*This research was funded by the Russell Sage Foundation and NIH. We are grateful to Megan Brooks, Joel Elvery, Gigi Foster, and Melissa Powell for outstanding research assistance, to Seth Sanders for helpful discussions, and to seminar participants at the Public Policy Institute of California and the Federal Reserve Board for helpful comments. An early version of this paper was prepared for the Color Lines Conference at Harvard University, August 2003. This paper reports the results of research and analysis undertaken while the authors were research affiliates at the Center for Economic Studies at the U.S. Census Bureau. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. It has been screened to ensure that no confidential information is revealed. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau, the Public Policy Institute of California, or the Russell Sage Foundation.

I. Introduction

Labor market differentials by race and ethnicity in the United States have been extensively documented (see, e.g., Donohue and Heckman, 1991; Cain, 1986; Altonji and Blank, 1999; Welch, 1990; and Ihlanfeldt and Sjoquist, 1990). The sources of these differences is a hotly contested question. When it comes to wage differentials, there has been extensive research trying to uncover their sources, with some researchers interpreting them as reflecting discrimination (e.g., Darity and Mason, 1998), and others arguing that they instead reflect skill differences that are unmeasured in many standard data sets (Heckman, 1998; Neal and Johnson, 1996; O'Neill, 1990). On the other hand, while there is widespread agreement that there is labor market segregation by race and ethnicity, and that this segregation accounts–at least in a statistical sense–for a sizable share of wage gaps between white males and other demographic groups (e.g., Carrington and Troske, 1998; Bayard, et al., 1999; King, 1992; Watts, 1995; Higgs, 1977), there has been very little work trying to uncover whether this segregation is due to discrimination or other sources.[1]

Discrimination is one fundamental potential explanation of workplace segregation. Perhaps the most convincing evidence of discrimination in employment comes from audit studies of hiring (Cross, et al., 1990; Turner, et al., 1991).[2] There are, however, other possible sources of labor market segregation. The principal alternative on which we focus in this paper is the potential role of skill in generating workplace segregation. There are numerous models suggesting that employers may segregate workers across workplaces by skill, most likely because of complementarities among workers with more similar

[1] This segregation may occur along industry and occupation lines, as well as at the more detailed level of the establishment or job cell (occupations within establishments). For example, Bayard, et al. (1999) found that, for men, job cell segregation by race accounts for about half of the black-white wage gap and a larger share of the Hispanic-white wage gap. Carrington and Troske (1997) use data sets much more limited in scope than the one we use here to examine workplace segregation by race and sex. In general, the paucity of research on workplace segregation is presumably a function of the lack of data linking workers to establishments.

[2] Heckman (1998) notes that even if there is hiring discrimination–as audit studies suggest–whether or not a wage differential arises depends on the discriminatory behavior of the marginal rather than the average employer. Black (1995) shows that in a search model discriminatory tastes on the part of some employers can result in a wage gap, even when the discriminatory employers do not hire any minorities.

skills. Because in U.S. labor markets skill is often correlated with race and ethnicity, workplace segregation along racial and ethnic lines could reflect segregation by skill.[3] Thus, this inquiry ultimately raises questions about whether or not labor market discrimination is a major determinant of workplace segregation.

This paper has two goals: to use a new matched employer-employee data set to provide the best available measurements of workplace segregation by race and ethnicity in the United States; and to present evidence that helps in understanding the sources of this segregation, in particular the role of skill. We pursue both of these goals using the 1990 Decennial Employer-Employee Database (DEED), a unique data set that we have created. The 1990 DEED is based on matching records in the 1990 Decennial Census of Population to a Census Bureau list of most business establishments in the United States.[4] The matching yields data on multiple workers matched to establishments, providing the means to measure workplace segregation in the United States based on a large, fairly representative data set. In addition, the reliance on the Decennial Census of Population as the source of information on workers creates the capacity to link information on workplace segregation to information on other characteristics of workers. This allows us to examine the role of skill in generating segregation. Thus, the DEED provides unparalleled opportunities to study workplace segregation by race, ethnicity, and skill.

Our empirical analysis proceeds in three steps that exploit these various characteristics of the DEED. First, we present measures of workplace segregation in the United States, focusing on segregation along the lines of race and ethnicity.[5] Rather than considering all deviations from proportional representation across establishments as an "outcome" or "behavior" to be explained, we scale our measured segregation to reflect segregation above and beyond that which would occur by chance if workers are distributed randomly across establishments, using Monte Carlo simulations to generate

---

[3] On the supply side, labor market networks can also generate workplace segregation; we do not focus on labor market networks in this paper.

[4] The 1990 Census of Population is currently the most recent Decennial Census available for matching. The 2000 data have just become available for matching, and we are now beginning the detailed process of constructing a 2000 DEED.

measures of randomly occurring segregation.[6]

Simple calculations of workplace segregation are important in their own right, aside from the questions we consider concerning the sources of workplace segregation. Most research on segregation focuses on residential segregation (e.g., Massey and Denton, 1987; and Cutler et al., 1999). But the boundaries used in studying residential segregation are to some extent arbitrary or even endogenously related to characteristics for which one wants to measure segregation; for example, Census tract boundaries are often generated in order to ensure that the tracts are "as homogeneous as possible with respect to population characteristics, economic status, and living conditions."[7] In contrast, workplaces–specifically establishments–are units of observation that are generated by economic forces and in which people clearly do interact in a variety of ways, including work, social activity, labor market networks, etc.[8] Thus, while it is more difficult to study workplace segregation because of data constraints, measuring workplace segregation may be more useful than measuring residential segregation, as traditionally defined, for describing the interactions that arise in society between different groups in the population.[9] Of course similar arguments to those about workplaces could be made about other settings, such as schools, religious institutions, etc. (e.g., James and Taeuber, 1985).

Second, our main inquiry probes the sources of racial and ethnic segregation. Numerous models suggest that employers find it useful to group workers of similar skills together. For example, Kremer and Maskin (1996) develop a model in which employers have incentives to segregate workers by skill when workers of different skill levels are not perfect substitutes and different tasks within firms are

---

[5] We focus on Hispanic ethnicity.

[6] This distinction between comparing measured segregation to a no-segregation ideal vs. segregation that is generated by randomness is discussed in other work (see, e.g., Cortese, et al., 1976; Winship, 1977; Boisso, et al., 1994; and Carrington and Troske, 1997).

[7] U.S. Census Bureau, www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed May 10, 2004).

[8] For a discussion of the importance of the workplace as a venue for social interaction between groups see Estlund (2003).

[9] Moreover, industry code, the closest proxy in public-use data to an establishment identifier, is a very crude measure to use to examine segregation. We calculate that racial and ethnic segregation at the three-digit industry level in the DEED is on the order of less than half as large (and sometimes much less) as the establishment-level segregation we document below.

differentially sensitive to skill.[10] Saint-Paul (2001) generates skill segregation across firms by assuming that there are productivity-related spillovers among workers within an establishment.[11] Cabrales and Calvó-Armengol (2002) show that when workers' utility depends on interpersonal comparisons with nearby workers (such as those in the same firm), segregation by skill results.[12] And, of course, there are potential benefits to employers from grouping together workers who speak the same language.

Because race and ethnicity are correlated with skill (for example, blacks and Hispanics have less education and Hispanics have lower English proficiency), racial and ethnic segregation may not reflect discrimination, but may be generated by segregation along skill lines. Because of educational differences between blacks and whites, we begin by measuring the extent of segregation between blacks and whites once we condition on both the quantity and the quality of education. We use years of education as a measure of the quantity of education, and measure the extent of workplace segregation by race within education groups. We also probe segregation that arises on the basis of quality (rather than quantity) of education. To do this, we follow Card and Krueger (1992) by considering only workers who work in the North, and using the fact that older blacks educated in the South received lower quality education relative to older blacks educated in the North, a phenomenon that is not particularly true for younger blacks.

In considering the source of workplace segregation by Hispanic ethnicity, we measure the extent

---

[10] For example, let the production function be $f(L_1, L_2) = L_1^c L_2^d$, with $d > c$. Assume that there are two types of workers: unskilled workers with labor input equal to one efficiency unit, and skilled workers with efficiency units of $q > 1$. Kremer and Maskin show that for low q, it is optimal for unskilled and skilled workers to work together, but above a certain threshold of q (that is, a certain amount of skill inequality), the equilibrium will reverse, and workers will be sorted across firms according to skill. Hirsch and Macpherson (1999) do not posit a formal model of sorting by skill, but assume that employers tend to hire workers of similar skills, and use this assumption–coupled with an assumption that blacks are on average less skilled than whites in terms of both observed and unobserved (to the researcher) skills–to suggest that the wage penalty associated with working in establishments with a large minority share in the workforce in part reflects lower unobserved skills of workers in such establishments.

[11] For example, positive spillovers may be reflected in each worker's productivity being the product of his productivity and an increasing function of the establishment's average skill level. Negative spillovers may arise because of fixed factors of production. All that is required for segregation in Saint-Paul's model is that over some range of average skill levels of an establishment's workforce there are increasing returns to skill.

[12] These authors also discuss evidence consistent with sorting by skill across employers, including Brown and Medoff (1989) and Davis, et al. (1991).

of segregation by English language ability, treating language ability as another important dimension of skill.[13] Because language is associated with ethnicity, language skills could underlie segregation by ethnicity, just as education may underlie segregation by race.[14]

Third, language is associated not only with skill, but also with country of origin, immigrant status, and assimilation. Consequently, if discriminatory forces lead to the segregation of blacks or Hispanics from whites, they can also operate to segregate workers with poor English skills (immigrants, most likely) from other workers, in which case segregation by language would not reflect skill complementarities. We probe this question by exploring segregation among those whose English proficiency is poor, but whose native (and spoken) languages differ.

Our results point to workplace segregation by race, and more so by ethnicity, even when accounting for the segregation that can occur randomly. We also, however, document substantial segregation by both the quantity and quality of education and by language ability, as well as by language differences among those who speak poor English. These findings consistently imply that these skill-related factors may be driving much of the segregation by race or ethnicity.

II. Data

The analysis in this paper is based on the DEED, which we have created at the Center for Economic Studies at the U.S. Bureau of the Census. The DEED is formed by matching workers to establishments. The workers are drawn from the 1990 Sample Edited Detail File (SEDF), which contains all individual responses to the 1990 Decennial Census of Population one-in-six Long Form. The establishments are drawn from the Standard Statistical Establishment List (SSEL), an administrative database containing information for all business establishments operating in the United States in 1990. Here we provide a brief overview of the construction of the DEED; more details regarding the matching

[13] We first documented segregation by language ability and explored its consequences for wages in Hellerstein and Neumark (forthcoming). Because language may reflect things other than skill, there may be additional influences on hiring by language, including customer discrimination or the need for workers to speak the same language as customers, which, coupled with residential patterns, lead to this form of workplace segregation.

[14] Of course education differences cold also contribute to workplace segregation of Hispanics and whites.

of the data are provided in Hellerstein and Neumark (forthcoming).

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. The file containing this employer name and address information is referred to as the "Write-In" file, and had previously been used only for internal Census Bureau purposes. The Write-In file contains the information written on the questionnaires by Long-Form respondents, but not actually captured in the SEDF. The SSEL is an annually-updated list of all business establishments with one or more employees operating in the United States. The Census Bureau uses the SSEL as a sampling frame for its Economic Censuses and Surveys, and continuously updates the information it contains. The SSEL contains the name and address of each establishment, geographic codes based on its location, its four-digit SIC code, and an identifier that allows the establishment to be linked to other establishments that are part of the same enterprise, and to other Census Bureau establishment- or firm-level data sets that contain more detailed employer characteristics. We can therefore use employer names and addresses for each worker in the Write-In file to match the Write-In file to the SSEL. Because the name and address information on the Write-In file is also available for virtually all employers in the SSEL, nearly all of the establishments in the SSEL that are classified as "active" by the Census Bureau are available for matching. Finally, because both the Write-In file and the SEDF contain identical sets of unique individual identifiers, we can use these identifiers to link the Write-In file to the SEDF. Thus, this procedure yields a very large data set with workers matched to their establishments, along with all of the information on workers from the SEDF.

Matching workers and establishments is a difficult task, because we would not expect employers' names and addresses to be recorded identically on the two files. To match workers and establishments based on the Write-In file, we use MatchWare–a specialized record linkage program. MatchWare is comprised of two parts: a name and address standardization mechanism (AutoStan); and a matching system (AutoMatch). This software has been used previously to link various Census Bureau data sets (Foster, et al., 1998). Our method to link records using MatchWare involves two basic steps. The first step is to use AutoStan to standardize employer names and addresses across the Write-In file and the SSEL. Standardization of addresses in the establishment and worker files helps to eliminate differences in how data are reported. For example, a worker may indicate that she works on “125 North Main Street,” while her employer reports “125 No. Main Str.” The standardization software considers a wide variety of different ways that common address and business terms can be written, and converts each to a single standard form.

Once the software standardizes the business names and addresses, each item is parsed into components. To see how this works, consider the case just mentioned above. The software will first standardize both the worker- and employer-provided addresses to something like “125 N Main St.” Then AutoStan will dissect the standardized addresses and create new variables from the pieces. For example, the standardization software produces separate variables for the House Number (125), directional indicator (N), street name (Main), and street type (St). The value of parsing the addresses into multiple pieces is that we can match on various combinations of these components

We supplemented the AutoStan software by creating an acronym for each company name, and added this variable to the list of matching components. We noticed that workers often included only the initials of the company for which they work (e.g., “ABC Corp”), while the business is more likely to include the official corporate name (e.g., “Albert, Bob, and Charlie Corporation”).

The second step of the matching process is to select and implement the matching specifications. The AutoMatch software uses a probabilistic matching algorithm that accounts for missing information, misspellings, and even inaccurate information. This software also permits users to control which matching variables to use, how heavily to weight each matching variable, and how similar two addresses must be in order to constitute a match. AutoMatch is designed to compare match criteria in a succession of “passes” through the data. Each pass is comprised of “Block” and “Match” statements. The Block statements list the variables that must match exactly in that pass in order for a record pair to be linked. In each pass, a worker record from the Write-In file is a candidate for linkage only if the Block variables agree completely with the set of designated Block variables on analogous establishment records in the

SSEL. The Match statements contain a set of additional variables from each record to be compared. These variables need not agree completely for records to be linked, but are assigned weights based on their value and reliability.

For example, we might assign “employer name” and “city name” as Block variables, and assign “street name” and “house number” as Match variables. In this case, AutoMatch compares a worker record only to those establishment records with the same employer name and city name. All employer records meeting these criteria are then weighted by whether and how closely they agree with the worker record on the street name and house number Match specifications. The algorithm applies greater weights to items that appear infrequently. So, for example, if there are several establishments on Main St. in a given town, but only one or two on Mississippi St., then the weight for “street name” for someone who works on Mississippi St. will be greater than the “street name” weight for a comparable Main St. worker. The employer record with the highest weight will be linked to the worker record conditional on the weight being above some chosen minimum. Worker records that cannot be matched to employer records based on the Block and Match criteria are considered residuals and we attempt to match these records on subsequent passes using different criteria.

It is clear that different Block and Match specifications may produce different sets of matches. Matching criteria should be broad enough to cover as many potential matches as possible, but narrow enough to ensure that only high probability matches are linked. Because the AutoMatch algorithm is not exact there is always a range of quality of matches, and we were therefore cautious in accepting linked record pairs. Our general strategy was to impose the most stringent criteria in the earliest passes, and to loosen the criteria in subsequent passes, while always maintaining criteria that erred on the side of avoiding false matches. We did substantial experimentation with different matching algorithms, and visually inspected thousands of matches as a guide to help determine cutoff weights. In total, we ran 16 passes, obtaining most of our matches in the earliest passes. Finally, we engaged in a number of procedures to fine-tune the matching process, involving hand-checking of thousands of matches and subsequent revision of the matching procedures.

The final result is an extremely large data set of workers matched to their establishment of employment. The DEED consists of information on 3.3 million workers matched to nearly one million establishments, which account for 27 percent of workers in the SEDF and 19 percent of establishments in the SSEL. Descriptive statistics for the matched workers and establishments, along with comparisons to the full SEDF and SSEL, respectively, are provided in Tables 1 and 2.[15]

As reported in Table 1, the means of the demographic variables in the DEED are quite close to the means in the SEDF. For example, female workers comprise 46 percent of the SEDF and 47 percent of the DEED. The distribution of workers across races and ethnicities is also relatively similar; in the SEDF, white, Hispanic, and black workers account for 82, 7, and 8 percent of the total, respectively. The comparable figures for the DEED are 86, 5, and 5 percent. Similarly, there is a close parallel between the distributions of workers across education categories in the two data sets. The distributions of workers across industries paint a slightly different picture, as approximately 25 percent of all workers in the SEDF are employed in the manufacturing sector, a figure that is somewhat greater in the DEED (33 percent). Retail workers comprise 20 percent of all workers in the SEDF, and 17 percent in the DEED.

In addition to comparing worker-based means, it is useful to examine the similarities across establishments in the SSEL and the DEED. Table 2 shows descriptive statistics for establishments in each data set. There are 5,237,592 establishments in the SSEL; of these, 972,436 (19 percent) also appear in the DEED. Because only workers who are sent Decennial Census Long Forms are eligible for matching to their employers, it is far more likely that at least one worker in large establishments will be sent a Long Form, and consequently more likely that such establishments are included in the DEED. One can see evidence of the bias toward larger employers by comparing the means across data sets for total

[15] For both data sets, we have excluded individuals as follows: with missing wages; who did not work in the year prior to the survey year (1989); who worked in public administration or were self-employed; who were not classified in a state of residence; or who were employed in an industry that was considered "out-of-scope" in the SSEL. ("Out-of-scope" industries do not fall under the purview of Census Bureau surveys. They include many agricultural industries, urban transit, the U.S. Postal Service, private households, schools and universities, labor unions, religious and membership organizations, and government/public administration. The Census Bureau does not validate the quality of SSEL data for businesses in out-of-scope industries.)

employment. (No doubt this also influences the distribution of workers and establishments across industries.) On average, establishments in the SSEL have 18 employees, while the average in the DEED is 53 workers. The distributions of establishments across industries in the DEED relative to the SSEL are similar to those in the worker sample. For example, manufacturing establishments are somewhat over-represented in the DEED, constituting 13 percent of establishments, relative to 6 percent in the SSEL. Overall, analyses reported in Hellerstein and Neumark (forthcoming) indicate that the DEED sample is far more representative than previous detailed matched data sets for the United States, and it is the largest national matched employer-employee database covering the United States that contains detailed demographic information on workers.[16]

III. Methods

We do our analysis for two measures of segregation. The first is based on measures of the percentages of workers in an individual's establishment, or workplace, in different demographic groups. For example, if Hispanics and whites are perfectly segregated, then Hispanics work with 100 percent Hispanics and zero percent whites, and conversely whites work with 100 percent whites and zero percent Hispanics. For a dichotomous classification of workers (e.g., whites and Hispanics), we define two segregation variables: the average percentage of Hispanic workers with which Hispanic workers work, denoted H_H; and the average percentage of Hispanic workers with which white workers work, denoted W_H.[17] The difference between these,

[16] Another national matched employer-employee data set currently under construction at the U.S. Census Bureau is the Longitudinal Employer Household Database (LEHD). The LEHD is very rich in that it contains observations on all workers in covered establishments (not limited to the one-in-six sample of Census Long-Form respondents) and is longitudinal in nature. As of now, however, the LEHD does not contain detailed demographic information on workers, and only covers a handful of states (although some of the largest ones). In addition, it matches workers to firms rather than establishments, so that workers can only be matched to establishments when the establishment is not part of a multi-unit firm. There were also some earlier data sets with information on employees within establishments–the Bureau of Labor Statistics Industry Wage Surveys and Area Wage Surveys–although these were not matched data sets per se. For research using these data sets, see Blau (1977) and Groshen (1991). These were specialized, non-representative data sets that covered particular industries or cities.

[17] In the sociological literature, the percentage of Hispanics with which Hispanics work is often called the "isolation index" and the percentage of whites with which Hispanics work is called the "exposure index."

$CW = H_H - W_H$

is our measure of observed "co-worker segregation," and measures the extent to which Hispanics are more likely to work with other Hispanics.[18]

The second measure of segregation is the traditional Duncan index (Duncan and Duncan, 1955). For two groups–for example, Hispanics H and whites W–this is defined as:

$DI = (½)\cdot\Sigma_i|H_i - W_i|$,

where H_i is the share of Hispanic workers in the Hispanic population who work in establishment i, and W_i is the share of whites in the white population who work in establishment i. The Duncan index ranges from zero to 100, and measures the percentage of workers in one of the two groups that would have to change establishments in order to create a perfectly integrated workforce in which, for example, the percentage of Hispanics in each establishment equaled the overall percentage in the workforce.[19]

Naturally, these two measures have some different features. The co-worker segregation measure (CW) is sensitive to the proportions of each group in the workforce. For example, if the distribution of Hispanics across establishments remains constant, but the number of Hispanics doubles, CW will rise. In contrast, the Duncan index (DI) is invariant to such a proportional change in the representation of any group. Depending on the question one is asking, one measure may be more appropriate than the other. While the Duncan index is widely used, we have some preference for the co-worker segregation measure because its units are simpler to interpret and because we think it better captures the composition of the workforce–as illustrated, for example, by the invariance of the Duncan index to proportional changes in the representation of a demographic group.

[18] Note that the term "co-worker" is slightly incorrect in that we factor in an individual's own ethnicity in calculating H_H and W_H. Also, we could equivalently define the percentages of white workers with which Hispanic or white workers work, H_W and W_W, which would simply be 100 minus these percentages, in which case CW would simply be the opposite sign.

[19] There are other common indexes of segregation such as the Gini coefficient, which would yield similar qualitative results to those we report below for the Duncan index. There are also multi-dimensional versions of Duncan indexes, where workers are classified into more than two groups. These indexes are harder to interpret, and we find that the traditional dichotomous classifications we use here describe well the phenomena of interest to us.

For each measure of segregation, we first report observed segregation, which is simply the sample estimate of the segregation measure. We denote these measures by appending an 'O' superscript to the segregation measures, i.e., CW^O or DI^O. One important point that is often overlooked, however, is that some segregation occurs randomly, and we are presumably most interested in the segregation that occurs systematically–i.e., that which is greater than would be expected to result from randomness. This is particularly problematic in a data set like ours in which small establishments are covered, and only subsets of workers in each establishment are sampled and matched, which together imply that we are often working with small numbers of workers per establishment. To see this quite starkly, suppose we have a large sample of establishments all of which have only two workers. Suppose further that the workforce is one-quarter Hispanic and three-quarters white. Then if workers are randomly assigned to establishments we expect the following ethnic composition of establishments to occur:

Probabilities of Alternative Workforce Compositions, Two Workers per Establishment, Random Assignment		
	First worker:	
Second worker:	Hispanic	White
Hispanic	1/16	3/16
White	3/16	9/16

In this example both measures indicate segregation. Computing the co-worker segregation measures, $H_H = 62.5$ and $W_H = 12.5$, so that $CW^O = 50$. The Duncan index is 75, indicating that at least 75 percent of either Hispanic or white workers have to change establishments to integrate the workforce.[20] Of course in a sense there is segregation, but it occurs randomly and therefore is presumably not the type of segregation that we would want to attribute to discrimination, skill complementarities, or any other

[20] In this extreme example, reallocating exactly 75 percent of, say, Hispanic workers so as to achieve a Duncan index of zero would require some establishments to have fractional numbers of Hispanics. This

behavioral explanation.[21]

Rather, what we care about from the perspective of understanding how behavior or policy shape workplace segregation is the segregation above and beyond what is generated by randomness. For example, if the Hispanic-Hispanic cell in this example contained much more than 1/16th of the observations, we would regard this as "real" segregation. Therefore, in the calculations we present, we first report observed segregation CW^O and DI^O. We then conduct a Monte Carlo simulation to determine how much segregation would occur randomly (the algorithm is described below), which we label "simulated segregation" and denote CW^S and DI^S.[22]

Following Carrington and Troske (1997), to measure segregation beyond that which would occur randomly, we compute the difference between observed segregation and the mean level of simulated segregation, and scale the difference by the maximum segregation that can occur. We refer to this as "effective segregation." In terms of the co-worker segregation measure, for example, when $CW^O > CW^S$ (which almost always occurs), the segregation measure is:

$$[\{CW^O - CW^S\}/\{100 - CW^S\}]\times 100 .$$

The denominator, $100 - CW^S$, is the maximum by which observed segregation can exceed simulated segregation, and so the scaling converts the difference $CW^O - CW^S$ into the share of this maximum possible segregation that is actually observed.[23] The corresponding measures for the Duncan

illustrates well that an ideal of full integration is rather silly in the context of small establishments.

[21] Nonetheless, even randomly generated segregation might lead to outcomes that could be considered sub-optimal from a social or policy standpoint.

[22] This was suggested and demonstrated by Boisso, et al. (1991) and implemented in Carrington and Troske (1997).

[23] When $CW^O < CW^S$ (which occurs once in this paper), the effective segregation measure is:

$$[\{CW^O - CW^S\}/CW^S]\times 100 ,$$

which measures the extent to which observed segregation is more even than the segregation that would be expected to occur randomly, which is CW^S. Consequently, in this case the effective segregation measure can be interpreted as capturing "excess evenness."

index are calculated in the same way.[24]

The example above of establishments with only two observed matched workers is, in fact, not unrealistic in our data, as many such establishments occur in the DEED. Indeed, if we simply take the full DEED, we have numerous cases of establishments with only one worker matched. For our empirical analysis, we emphasize results using the subsample of establishments with a minimum of two matched workers. In Appendix A, however, we also report results without this lower bound (so that establishments with only one worker matched are included), as well as with a more restrictive limit of at least five matched workers per establishment. The presentation of these alternative results serves two goals. First, it provides an indication of the robustness of the results. Second, it helps in assessing whether our methods for capturing the segregation above and beyond that which occurs randomly are successful. If so, we should see considerable differences in observed segregation as the cutoff for the number of matched workers varies–with more observed segregation the lower the cutoff. But the simulated segregation measures should also reflect these cutoffs in the same way, and effective segregation should not be influenced by the cutoff used.[25] The one exception is if the extent of segregation differs by establishment size, given that imposing a higher cutoff on the number of matched workers will also tend to drop small establishments; in such cases we can also examine directly the sensitivity of the results to establishment size.

For the Monte Carlo simulations that generate measures of random segregation, we first define the geographic unit–which for most of the analysis is metropolitan areas. We then compute from our data

[24] In line with the earlier example, the Duncan index provides a good illustration of why the scaled segregation measure is more meaningful than the simple difference DI^O - DI^S. This index is computed at the establishment level, and many establishments have only two or three matched workers. This leads to a high simulated (random) Duncan index, which means that there is little scope for observed segregation to exceed actual segregation. Thus DI^O - DI^S might appear low, whereas it is large relative to the maximum possible segregation beyond randomness that could occur.

[25] When establishments with only one matched worker are also included, the fact that all such establishments are perfectly segregated is taken account of by our techniques, because in the simulation the same number of establishments are assigned a single worker in proportion to the composition of the corresponding workforce. However, there is an obvious sense in which two matched workers is the minimum required to meaningfully refer to the characteristics of co-workers.

the numbers of workers in each category for which we are doing the simulation–for example, blacks and whites–as well as the number of establishments and the size distribution of establishments (in terms of sampled workers). Within a metropolitan area, we then randomly assign workers to establishments, ensuring that we generate the same size distribution of establishments within a metropolitan area as we have in the sample. One "round" of random assignment is considered a single simulation. We do the simulation 100 times, and compute the simulated segregation measures as the means over these 100 simulations. It turns out that the simulated segregation measures are very precise; in all cases the standard deviations were trivially small (ranging to a high of .04 for some of the smallest samples we study), and observed segregation was well the outside the 99-percent confidence interval for the simulated measures.

As noted above, most of our analysis focuses on metropolitan areas. We use U.S. Census Bureau measures of metropolitan areas, because these are defined to some extent based on areas within which substantial commuting to work occurs.[26] We first look at segregation within Consolidated Metropolitan Statistical Areas (CMSAs) and Metropolitan Statistical Areas (MSAs) for which there is no CMSA, and then at MSAs and Primary Metropolitan Statistical Areas (PMSAs), which are parts of CMSAs. The restrictions to workers in CMSAs/MSAs or MSAs/PMSAs reduce the sample by about one-third. In addition, for the calculations of the segregation measures the sample is further restricted to the subset of groups considered (e.g., Hispanics and whites only), and for the indexes by metropolitan area, a metropolitan area is dropped from the relevant segregation calculation if there is no worker in the DEED from the subset considered (e.g., Hispanics) in that urban area. When we perform the simulations for disaggregated regions and calculate segregation indexes, we condition on geographic area of residence (and work) so that we have region-specific segregation measures. We then calculate "conditional" national segregation measures by weighting over the whole sample (that is, summary measures for the extent of workplace segregation, where we condition on the metropolitan area where the worker lives and

[26] See U.S. Census Bureau, http://www.census.gov/geo/lv4help/cengeoglos.html (viewed July 3, 2003). This is not to say that residential segregation at a level below that of MSAs and PMSAs may not influence workplace segregation. However, an analysis of this question requires somewhat different methods. For example, in conducting the simulations it is not obvious how one should limit the set of

works).[27] For descriptive purposes we also present some "unconditional" nationwide segregation measures where we do not first condition on metropolitan area, and where in the simulations we randomly assign workers to establishments anywhere in the country. For comparability, when we construct these unconditional segregation indexes we use only the workers included in the CMSA/MSA or MSA/PMSA samples.

The segregation measures for each region in the sample have to be weighted to construct conditional national segregation indexes. The co-worker observed and simulated segregation measures are worker based, so our national conditional co-worker indexes are simply the overall sample means across all workers, effectively weighting metropolitan areas by the number of workers in them. As a result, for the observed segregation measures the conditional and unconditional measures yield identical results; only the simulations differ. For the Duncan indexes, the observed segregation measures change for the within-MSA analysis, because the index is calculated relative to the workforce in the metropolitan area, so observed segregation conditional on region generally differs from the unconditional measure. To arrive at conditional national Duncan indexes of observed and simulated segregation, we average the area-specific indexes, weighting by the size of the workforce in the metropolitan area. With these conditional national indexes of observed and simulated segregation, we then construct the effective segregation indexes, which measure the level of effective segregation to which the average worker is subjected, conditional on the distribution of workers across metropolitan areas.

Finally, in addition to constructing estimates of effective segregation in the workplace along various dimensions, we are interested in comparisons of measures of effective segregation across different samples. Given the complicated method of measuring effective segregation, and given also that we are

establishments within a metropolitan area in which a worker could be employed.

[27] In all cases, a worker in must live and work in the same geographical region, or the worker is dropped from the sample. When we compute Duncan indexes within metropolitan areas, we define the workforce shares represented by demographic groups relative to the workforce in the metropolitan area, rather than nationally. That way, the Duncan index for each metropolitan area has the correct within-area interpretation.

sometimes comparing estimates across samples that have some overlap,[28] we can only assess statistical significance of measures of effective segregation or differences between them using bootstrap methods. These methods are very computationally intensive because within each iteration of the bootstrap we have to do the set of simulations needed to construct measures of simulated segregation, and our samples are often very large. The methods for doing this for a number of cases and the results are detailed in Appendix B. Briefly, the evidence indicates that our estimates are quite precise, and that the differences between the effective segregation indexes discussed in detail in the next section are generally strongly statistically significant.

IV. Results

Workplace Segregation by Race

The analysis begins with measures of workplace segregation by race (black and white). Table 3 reports results for black-white segregation, using the baseline sample of establishments with two or more matched workers. To provide a sense of overall segregation, column (1) provides the various segregation measures at the unconditional national level, looking at all urban areas (CMSAs and MSAs) as a whole. Column (2) presents the conditional national segregation indexes that are constructed by weighting up to the national level each individual CMSA/MSA segregation measure. Column (3) repeats the unconditional national measures, but for the subsample of workers who live and work in the often-smaller MSA/PMSA urban areas, and column (4) provides the conditional MSA/PMSA segregation measures.

In column (1), looking first at the observed co-worker segregation measures, we see extensive segregation. In particular, black workers on average work in establishments in which the matched workforce is 33.4 percent black, whereas whites work in establishments with workforces that are only 5.0 percent black on average. Below these figures we present the calculations from the simulations. These reveal that a considerable amount of segregation arises randomly. In particular, random assignment would lead blacks to work in establishments with workforces that are on average 23.1 percent black,

[28] For example, we compare effective segregation between Hispanics who speak English poorly and Hispanics who speak English well, to effective segregation between Hispanics who speak English poorly

versus an average percent black of 5.7 percent for whites. Based on the comparison between observed and simulated segregation, the effective segregation measure is 13.4 percent, meaning that just over 13 percent of the maximum amount of segregation that could arise due to non-random factors is actually observed in the data. The Duncan indexes reported at the bottom of column (1) tell a similar story. The observed Duncan index of 68.5 suggests extensive segregation, and although simulated segregation is also quite high (an index of 54.1), the effective segregation measure is still a substantial 31.4.

Column (2) looks at segregation within urban areas defined as CMSAs/MSAs. As noted earlier, observed co-worker segregation is the same within and across urban areas. Observed segregation measured with the Duncan index is virtually identical to that in column (1) because blacks are reasonably evenly dispersed across CMSAs. Simulated segregation is somewhat higher in column (2) than in column (1) because workers are reallocated for the simulation only within urban areas; as a result, the effective segregation measures are smaller in column (2) than in column (1), by about 30 percent for the co-worker measure and 40 percent for the Duncan index. The resulting effective segregation measures are 9.5 for co-worker segregation, and 19.1 for the Duncan index.

Columns (3) and (4) of Table 3 repeat this analysis for the urban areas defined by MSAs/PMSAs. As we would expect, simulated segregation often turns out to be a shade higher within MSAs/PMSAs, but overall this has relatively little impact on the estimates or on the qualitative conclusions.

Segregation by Skill versus Segregation by Race

To this point, we have taken the relatively standard approach of studying segregation from the perspective of race. Such analyses are often motivated by the fact that discriminatory behavior on the part of economic agents can lead to workplace segregation by race (and ethnicity). For example, any of Becker's (1971) models of discrimination (by employers, employees, or customers) can lead to workforce segregation. What is often ignored in discussions of segregation, however, is whether the racial (and ethnic) dimensions along which segregation is measured are simply proxies for other characteristics of workers along which employers segregate their workforces for non-discriminatory (profit-maximizing)

---

and non-Hispanics who speak English poorly.

reasons.

By way of contrast, there has been a tremendous amount of attention paid to whether skill differences drive race differences in wages. For example, O'Neill (1990) and Neal and Johnson (1996) show that wage differences between young black and white males in the National Longitudinal Survey of Youth can be largely eliminated when one controls in a wage regression for age-adjusted AFQT test score. But unlike the rich data sets containing detailed information on individuals' wages and skill levels, there previously have been few data sets containing information on where workers work that, coupled with measures of workers' skills, could allow a thorough investigation of the role of skill in generating racial segregation in the workplace.

As noted in the Introduction, there may be incentives for employers to segregate workers along skill lines.[29] The obvious skill distinction between blacks and whites is in formal education. For example, in the samples used in Table 3, about 60 percent of whites have more than just a high school degree, versus about 50 percent of blacks, while just under 20 percent of blacks do not have a high school diploma, compared with about 10 percent of whites. Moreover, there is research indicating that the quality of black education was particularly low in the South for those educated in the pre-Civil Rights years of the 20th century, as a result of public school segregation and unequal resources (Card and Krueger, 1992).[30] Thus, aside from overall education levels, there are differences in school quality by race (and region) across cohorts that can be exploited to test for the importance of skill differences in generating workplace segregation.

Given these considerations, we turn to an analysis of workplace segregation by education levels

---

[29] This same issue is taken up with regard to residential segregation by race and ethnicity in Bayer, et al. (2002), who explore the extent to which this segregation is driven by household characteristics such as immigration status, education, and income. Paralleling our discussion of benefits from language similarities in the workplace, they suggest that residential segregation by language may bring some benefits (at least in the short run) by easing communication of those with poor English skills.

[30] While *Brown v. Board of Education*, in 1954, spurred desegregation, Card and Krueger show that school segregation remained high for cohorts born through 1949. Furthermore, they show that three school quality measures–pupil-teacher ratios, term length, and teacher pay–strongly favored white schools in this period, although the gaps were widest in the first half of the century and began to converge somewhat prior to *Brown*.

and educational quality in order to try to disentangle the contribution of skill to the segregation by race documented above. For this analysis, we look only within metropolitan areas, so we are already conditioning on any segregation by skill that occurs across metropolitan areas. Given that the CMSA/MSA and MSA/PMSA results were so similar in the earlier table, we use only the latter here.

In Table 4 we examine black-white segregation conditioning in various ways on years of education. The first column repeats the previous estimates on overall black-white segregation, from column (4) of Table 3, for purposes of comparison. In column (2) of Table 4 we report benchmark segregation measures for the extent of segregation by skill for whites, where we compute segregation measures for highly-educated whites (more than a high school degree) relative to less-educated whites (at most a high school degree). These segregation measure serve two purposes. First, they put into context the scale of the numbers presented in Table 3 documenting the extent of segregation by race. Second, they provide a kind of lower bound for the extent of segregation by education across the races. Because quality differences in education between whites and blacks may lead to (mostly unobservable) skill differences across the races, even conditional on education it is possible that some segregation by race is a function of skill differences. Therefore, skill segregation among whites should be lower than skill segregation between blacks and whites.

The effective co-worker segregation by education for whites in column (2) is 17.0, almost 80 percent higher than the effective co-worker segregation of 9.6 between all blacks and all whites. For the Duncan index, the effective segregation measure for white segregation by education is 24.5, larger than the 18.4 effective Duncan index for black-white segregation. That is, segregation by education is larger–and sometimes a lot larger–than black-white segregation.

In column (3) we report segregation by education for low-educated (high school or less) blacks relative to high-educated (more than high school) whites. There are two important comparisons to be made with this column, both of which imply racial segregation that is driven largely by skill differences. First, segregation between low-educated blacks and high-educated whites is much more pronounced than overall black-white segregation; for example, the effective co-worker segregation measure is 18.0, nearly double the figure for overall black-white segregation. It is worth noting that this difference across columns is driven largely by the difference in the observed segregation measures across the columns, rather than differences in the extent of simulated segregation that scale the effective measures. Second, segregation between low-educated blacks and high-educated whites is almost identical to that between low-educated whites and high-educated whites. To illustrate, the effective co-worker segregation measure in column (3) is 18.0, while the comparable figure in column (2) for whites only is 17.0. Here, the similarity in the indexes is driven by both the numerator and denominator of the effective measures being very similar. In total, then, the comparison between columns (2) and (3) shows that black and white workers with low levels of education are segregated in virtually identical (effective) amounts from high-educated white workers.

In column (4) we compute the extent of segregation by education for black workers. It turns out that segregation by education for blacks is similar, although slightly smaller, than segregation by education for whites, again suggesting that skill differences play a large role in generating workplace segregation. For example, the effective co-worker segregation by education for blacks is 14.1, relative to the corresponding measure of 17.0 for whites.

Finally, in columns (5) and (6) of Table 4 we return to workplace segregation by race, this time conditioning on education. In column (5) we report the extent of segregation by race for less-educated workers; column (6) reports the segregation measures for workers with more than a high school degree. The segregation measures in column (5) reveal that segregation among less-educated blacks and whites is larger than overall black-white segregation, but is of the same magnitude as the extent of segregation by skill. In contrast, column (6) shows that segregation among high-education blacks and whites is less than among blacks and whites overall, by about 25 to 30 percent. Therefore, while conditioning on education does not eliminate racial segregation in the workplace, racial segregation is of an order of magnitude no larger (and sometimes much smaller) than skill segregation for whites, and skill segregation appears to play a large role in generating racial segregation in the workplace.[31]

Table 5 presents the analysis in which we use information on where black men were schooled to examine the role of schooling quality differentials in generating racial segregation. In particular, we first restrict attention to those born before 1950, to capture the period when school quality differentials were sharp. We distinguish black workers by whether they were educated in the South or not (based on the 18 Southern states defined by Card and Krueger). We then ask whether low-educated black workers who were educated (born) in the South are more segregated from high-educated white workers than are low-educated black workers who were educated (born) in the North. Because blacks born in the South are more likely to still reside in the South, we restrict attention to those blacks and whites currently residing in the North, to preclude the results being influenced by different segregation patterns in the South versus the North (in particular, more racial segregation in the South, which would bias the results toward finding more segregation associated with skill). Finally, to examine whether any differences are in fact attributable to school quality, we repeat the analysis for those born after 1950, for which education quality differences associated with region of birth should be smaller if they exist at all.

All measures of segregation in Table 5 are lower than the corresponding measures in Table 4 because, overall, segregation by race is lower in the North. Column (1) of Table 5 reports the estimates for those born before 1950, with blacks educated in the North, while column (2) reports the corresponding estimates for blacks educated in the South. For either the co-worker segregation measure or the Duncan index, effective segregation by race is considerably higher for blacks educated in the South, consistent with skill playing an important role. Indeed, for the co-worker segregation measure there is virtually no segregation by race for blacks educated in the North; the effective measure is 2.8, versus 11.7 for blacks educated in the South. In contrast, in columns (3) and (4), which report results for those born after 1950, the differences are smaller and the direction is reversed. The finding, in the latter estimates, that there is

[31] These results are echoed when we compute (unreported in the tables) segregation by race within four education categories: less than high school, high school, some college, and college or more. Segregation decreases monotonically as education increases, but at its largest (for less than high school) is of the same magnitude as segregation by education for whites.

somewhat more segregation of low-education Northern-born blacks from high-education whites than of low-education Southern-born blacks from high-education whites is not surprising, as the adjustment costs of moving would be expected to lead to selective migration of (unobservably) higher-skilled black workers from South to North, and will dominate any small regional school quality differences that may remain. Thus, the evidence reported in Table 5 is again consistent with the conclusion that skill differences play a central role in generating workplace segregation by race.

Workplace Segregation by Ethnicity

We now turn to an examination of the extent and causes of workplace segregation by Hispanic ethnicity. The baseline estimates for the extent of Hispanic-white segregation are reported in Table 6. The first thing to note is that the segregation figures for the unconditional national indexes indicate somewhat more segregation by ethnicity than their counterparts for race as reported in Table 3. For example, in column (1) of Table 6, the average percentage of Hispanics with whom Hispanics work is 48.9 percent, versus a comparable figure of 33.4 percent for blacks. This difference is not attributable to differences in randomly generated segregation; for U.S. metropolitan areas as a whole, simulated segregation is very similar for black-white and Hispanic-white segregation. For example, in column (1) of Table 6 the simulated share of black workers in the establishment is 23.5 percent for Hispanics, versus a comparable share of 23.1 percent for blacks.

However, when we look within metropolitan areas, we see that randomly-generated segregation is quite a bit higher among Hispanics than among blacks. In column (2) of Table 6, for example, the simulated share Hispanic for Hispanic workers is 36.7 percent, compared with a simulated share black for black workers of 26.5 percent in column (2) of Table 3. This difference occurs because Hispanics are much less evenly dispersed across metropolitan areas than are blacks, so that random sorting of workers into establishments only within urban areas generates more workplace segregation among Hispanics.

The net result is that the effective co-worker segregation measures are higher for Hispanics than for blacks. For example, the effective measure for the CMSA/MSA sample is 19.2, versus 9.5 for blacks. The difference is in the same direction but much less pronounced using the Duncan index. For example, for the corresponding calculation, the effective segregation measure based on the Duncan index is 21.3 for Hispanics, versus 19.1 for blacks.

Segregation by Skill versus Segregation by Ethnicity

For Hispanics and whites, the obvious skill distinction is English language ability. In the samples used in Table 6, virtually 100 percent of whites report speaking English well or very well, compared with a bit over 80 percent of Hispanics.[32] Table 7 therefore looks at Hispanic-white segregation, comparing Hispanic-white segregation for Hispanics with good English skills to Hispanic-white segregation for Hispanics with poor English skills (as opposed to showing simply that workers are segregated on the basis of language skills, which could be attributable just to segregation along lines of Hispanic ethnicity). Column (1) repeats the information from column (4) of Table 6 to provide a benchmark. Then in columns (2) and (3) we report segregation calculations first for white workers and Hispanics with good English skills, and then for white workers and Hispanics with poor English skills. The basic message from these calculations is quite clear. Segregation between whites and Hispanics with good English skills is minor relative to the very pronounced segregation between whites and Hispanics with poor English skills. For example, the estimates in column (2) indicate that while observed segregation between whites and Hispanics with good English skills is high (for example, the average Hispanic in this subsample works in an establishment that is 43.5 percent Hispanic), the co-worker effective segregation measure is only 11.9. In contrast, for whites and Hispanics with poor English skills the corresponding effective segregation measure is 44.8, more than double the figure for Hispanic-white segregation overall. The same finding is echoed in the Duncan indexes, for which segregation is considerably higher (by a factor of more than four) between whites and Hispanics with poor English skills.[33]

---

[32] While there are observed education differences between Hispanics and whites, skill may be embodied differently in the education of Hispanics who are born outside of the United States. We therefore focus on education as a measure of skill differences between blacks and whites, and English language proficiency as a measure of skill differences between Hispanics and whites.

[33] We note, however, that the simulated Duncan index in column (3) of Table 7 is 88.0 percent, so that even a small difference between observed and simulated indexes would lead to a large effective segregation measure. Nonetheless, the same is not true of the corresponding co-worker segregation measure in Table 7, so the overall point remains that segregation from white workers of Hispanics with

The figures on language skills and segregation among whites and Hispanics reported in Table 7 suggest that a major share of Hispanic-white workplace segregation is attributable to, or at least associated with, language differences. Indeed, if Hispanics with good English had English skills identical to those of white workers, the reported segregation measures in column (2) would presumably be even smaller.

Understanding Workplace Segregation by Skill

For both black and Hispanic workers we have documented that substantial workplace segregation is generated by skill differences. One interpretation of this evidence is that employers have good reasons to pursue such segregation, and because skills are correlated with race and ethnicity, segregation by skill arising for non-discriminatory reasons generates segregation by race and ethnicity. Another possibility, though, is that the proxies for skill that we have examined are associated with other dimensions along which employers discriminate–such as national origin or socioeconomic factors–and on the basis of which they crowd workers into a subset of jobs (typically jobs that pay less). It is difficult to distinguish between these competing hypotheses.[34] In the case of language skills, however, we believe some progress can be made on this question.

In particular, to test whether there are legitimate economic reasons for segregation by language skill, as opposed to simple segregation of those with poor English into a subset of jobs, we can consider employment patterns for workers who speak poor English but who also speak different languages. If Hispanic poor English speakers (who generally speak Spanish) are not segregated from non-Hispanic poor English speakers (who speak a language other than Spanish), then this would suggest that those with low skills are clustered in the same workplaces for reasons other than efficiency gains from grouping

---

poor English ability is much larger than that of Hispanics with good English ability.

[34] This is potentially true in many contexts, even though it is often ignored. For example, Bertrand and Mullainathan (2003) provide evidence from an audit study that employers are less likely to interview job candidates with "black-sounding" names. This may be because of race discrimination per se, or because of discrimination against workers whose names suggest a certain cultural and socioeconomic upbringing (or the intersection of the two), but the paper has been interpreted as providing evidence of discrimination on the basis of race. (See also Fryer and Levitt, 2003.)

workers who speak the same language; such segregation would be more consistent with simple segregation of "less desirable" workers into a subset of jobs. In contrast, if Hispanic poor English speakers are segregated from those who have poor English skills but speak languages other than Spanish, then segregation by language skills may be arising for reasons of complementarity between workers who speak the same language (or a related economic incentive to segregate workplaces by language).[35]

The results of this analysis are reported in Table 8. Column (1) reports calculations for segregation between Hispanic workers with poor English skills and Hispanic workers with good English skills. There is some segregation among these workers, although considerably less than the segregation between white workers and Hispanics with poor English skills (see column (3) of Table 7). For example, effective co-worker segregation among Hispanics with good English skills and poor English skills is 28.5, compared with 44.8 (from Table 7) among white workers and Hispanic workers with poor English skills. This makes sense since the language skills between these two groups of Hispanics may not differ if Hispanics with good English skills also speak Spanish, as some no doubt do. In contrast, column (2) reports calculations for segregation between Hispanics with poor English skills and non-Hispanics with poor English skills. These figures indicate much more extensive segregation–considerably more than in column (1) (48.4 versus 28.5 for the effective co-worker measure), and much more similar to the segregation between whites and Hispanics with poor English skills. Thus, this evidence suggests that much of the segregation of Hispanics with poor English skills arises because of factors other than the general crowding of low-skilled workers into the same set of low-paying workplaces.

However, some caution is in order in interpreting these estimates. Unlike the case with education differences overall or English language skills among Hispanics, residential segregation between Hispanics and other groups with poor English might be quite strong. If residential segregation by language drives workplace segregation along similar lines, we might expect that simulated segregation would be higher

[35] However, the latter finding would not necessarily be decisive, because such segregation by language may be a function of residential segregation and/or hiring networks where workers who speak the same language have access to the same subset of employers. Network relationships can themselves be efficiency enhancing if they make it easier for workers to find jobs or for employers to find workers.

and effective segregation lower for units of analysis smaller than the MSA/PMSA level. Therefore, to further address whether the segregation by language among poor English speakers is due to complementarities in the workplace, we explore differences in language segregation by establishment size. After all, in larger establishments there may be considerably more scope for segregating workers within establishments, so that across-establishment segregation is not as critical in achieving language complementarities.[36] Table 9 reports similar calculations to those in Table 8, but for establishments with different minimum total employment cutoffs–roughly the 25th, 50th, and 75th percentiles of the size distribution. As the table shows, segregation of Hispanics from non-Hispanics when both groups have poor English skills falls as the minimum establishment size is raised. This pattern suggests that language complementarities do in fact contribute to workplace segregation by language among those who speak poor English.

V. Conclusions

We use a unique data set of employees matched to establishments to study workplace segregation in the United States. We document the rather extensive observed segregation by race and more so by ethnicity in the United States, but note that much observed segregation overstates actual segregation when we are studying many small units–as occurs with a representative sample of establishments, and more so in our case when we are matching only a sample of workers to establishments.

Our analysis focuses on whether this racial and ethnic segregation reflects race or ethnicity per se, likely stemming from discrimination, or instead is attributable to skills that differ across race and ethnic groups and along which employers might find it useful to segregate workers. These analyses suggest that sizable shares of racial and ethnic segregation appear to be attributable to skill. We show that segregation of low-educated whites from high-educated whites is nearly as high as segregation of low-educated blacks

[36] As an anecdotal example, an article in the *New York Times* describes a Texas factory that nearly completely segregates its Hispanic and Vietnamese workers into two different departments in the factory (with the Hispanics working in the lower-paying department). This article also points to the role of language complementarities between workers and supervisors, as one of the company's defenses of this practice is that the supervisor of the higher-paying department speaks Vietnamese but not Spanish (Greenhouse, 2003).

from high-educated whites and only slightly smaller than segregation of low-educated blacks from high-educated whites. Although among blacks and whites with low education, racial segregation is of the same order of magnitude as between whites of different education levels, among blacks and whites with high education racial segregation is only about one-third as large as segregation between whites of different education levels. All of this implies that skill plays a very large role in generating racial segregation.

In order to probe the role of skill in generating ethnic segregation we focus on language skill differences between whites and Hispanics. For example, segregation between whites and Hispanics with good English skills is only about two-thirds as large as overall white-Hispanic segregation, and about one-quarter as large as segregation between whites and Hispanics with poor English skills. Like the results for race, these findings suggest that skill plays an important role in generating ethnic segregation in the workplace.

Finally, we ask whether segregation by skill likely arises due to the consignment of less-skilled workers to the same subset of workplaces, perhaps because of discrimination against workers on the basis of numerous characteristics associated with low skills, or whether other factors such as skill-based complementarities lead certain types of workers to work together. Providing evidence inconsistent with the first hypothesis, we find that Hispanics with poor English skills are considerably more segregated from workers with poor English skills who speak other languages than from Hispanics with good English skills. It therefore appears that the process by which workers are sorted into workplaces is not simply one whereby low-skilled workers are relegated to the same set of (low-paying) workplaces.

None of these findings deny the reality of racial and ethnic segregation in U.S. workplaces. Nor do they imply that efforts to directly influence racial and ethnic hiring patterns cannot help to reduce this segregation. But they do suggest that the important, perhaps even the main, culprits in generating this segregation are not discriminatory hiring practices based on race and ethnicity. Rather, skill differences between white, black, and Hispanic workers appear to account for substantial shares of workplace segregation.

To summarize such findings in terms of one segregation measure we present, consider co-worker segregation, which measures the difference between the average percentage of Hispanic (black) workers with whom Hispanics (blacks) work, and the average percentage of Hispanic (black) workers with whom whites work. Within metropolitan areas, the effective segregation estimate is 18.7 for Hispanics and 9.6 for blacks. Compared to the baseline effective segregation measure of 18.7 for Hispanics, the effective measure of segregation between whites and Hispanics with good English skills is 11.9, while the segregation among whites and Hispanics with poor English skills is 44.8. Looking only at whites and blacks with some college education, the effective segregation measure is only 6.8, while segregation among whites with low education and whites with high education is 17.0. Clearly, then, skill differences account for major shares of workplace segregation by race and ethnicity.

As we find in so many other studies of race and ethnic differences (e.g., earnings inequality), closing skill gaps between whites, blacks, and Hispanics is likely to contribute importantly to equalizing other outcomes. There are two corollaries of this. First, trying to reduce workplace segregation without confronting skill gaps may entail efficiency costs, if indeed there are good reasons to segregate workers by skill. On the other hand, given skill complementarities, when employers are faced with workers with divergent skills, either because of the available labor pool or because of government anti-discrimination efforts, employers may have some incentives to attempt to close skill gaps between workers.[37] However, most skill gaps are probably generated considerably earlier in the life-cycle, suggesting that efforts to equalize schooling and other early opportunities for skill acquisition are critical.

[37] For example, an article in the *Washington Post* describes efforts of employers of large numbers of Hispanic workers to learn Spanish (Rivera, 2003).

Appendix A: Robustness of the Results with Respect to the Minimum Number of Matched Workers

In the tables covered in the main text, all the results were generated from the sample of establishments with two or more matched workers. In this appendix, we explore the robustness of the results to using different minimum numbers of matched workers. Perhaps more importantly, though, this helps to illustrate the importance of considering the deviation of observed segregation from random (simulated) segregation, because with a lower floor on the number of workers matched there should be considerably more segregation generated randomly, affecting both observed and simulated segregation, and vice versa. If, however, the underlying process generating segregation is the same regardless of establishment size, the effective segregation measures should be similar. Of course, they may still differ if the forces that lead to segregation differ in larger versus smaller establishments.

In Appendix Table A1, we repeat the within MSA/PMSA analysis from columns (4) of Tables 3 and 6, but using the larger subsample of all establishments (which includes establishments with one matched worker), and the smaller subsample of establishments for which five or more workers are matched.[38] The results show that observed co-worker segregation is considerably higher for the subsample that requires only one matched worker, and conversely considerably lower for the subsamples with more matched workers. The same types of differences are observed for simulated segregation. In contrast, the effective segregation measures are much more robust across sample sizes for both blacks and Hispanics. The one exception is that the effective segregation measure is noticeably lower for blacks in the sample of all establishments (by about half using the co-worker segregation measure), suggesting that the employment patterns of blacks are generated by more segregation in larger establishments.

In Table A2 we repeat the calculations from Table 4 for black-white segregation by skill, but for the alternative samples based on the minimum number of matched workers. The qualitative conclusions are quite similar across samples, with segregation by education almost always larger than segregation by

[38] Note that the restriction to establishments with five or more matched workers retains about two-thirds of the workers but only about one-quarter of the establishments. This difference reflects the larger

race, segregation by race conditional on education somewhat less extreme than overall segregation by race, and segregation between less-educated blacks and more-educated whites considerably sharper than overall racial segregation.

Paralleling our other sensitivity analyses, Appendix Table A3 repeats the analysis of racial segregation in Table 5 for the larger sample of all establishments, and the smaller sample of establishments with five or more workers. The qualitative pattern of the evidence is the same as in the main table. When we look at low-educated blacks and high-educated whites residing in the North, segregation of blacks from whites is higher for blacks educated (born) in the South, when attention is focused on those born before 1950 when black schools in the South were measurably inferior.[39]

Appendix Table A4 parallels Table 7, showing calculations of skill segregation among whites and Hispanics classified by English proficiency, for the sample of all establishments and the subsample of establishments with five or more matched workers. These alternative samples yield very similar conclusions with regard to the difference between the segregation of whites from Hispanics with poor English skills and from Hispanics with good English skills. In addition, the estimates are quite robust across the alternative samples.

Finally, Appendix Table A5 parallels Table 8, reporting the analysis of segregation based on language skills and Hispanic versus non-Hispanic ethnicity. In all cases, segregation is higher among Hispanics and non-Hispanics with poor English skills. However, while the estimates (and the contrast) for the sample of all establishments are quite similar to those in Table 8, for the subsample of establishments with five or more matched workers the contrast is less sharp, and in particular segregation among Hispanics and non-Hispanics with poor English skills falls quite a bit. Given that the samples with five or more matched workers are on average larger, this is consistent with the results in Table 9

---

establishments retained with the cutoff of five matched workers as opposed to two matched workers.

[39] Note that for the sample of all establishments, effective segregation is actually negative when we focus on blacks born in the North before 1950 (column (1)), indicating excess evenness. Regardless, though, the evidence points to lower segregation in this case than when we focus on blacks born in the South before 1950.

indicating less segregation by language in larger establishments.

In total, the calculations for the establishments with different minimum numbers of matched workers demonstrate that observed segregation can be quite misleading, especially when there are many establishments with few workers (or few matched workers). At the same time, the calculations suggest that looking at the differences between observed and simulated segregation largely solves the problem of spurious segregation generated by randomness.

Appendix B: Statistical Significance

From the point of view of drawing statistical inferences, we need to be able to assess the statistical significance of our effective segregation measures and of differences between them. Given the precision of the simulated segregation measures as discussed in Section III, the effective segregation measures are also likely relatively precise. To assess this more formally, we explore bootstrapped distributions for the effective segregation measures.

To carry out this procedure, at each iteration of the bootstrap we draw a sample with replacement of the original size of the sample. We sample establishments, not workers. This ensures that we maintain the size distribution of establishments, and in particular that we maintain the restriction that all establishments have at least two matched workers. The bootstrap sample at each iteration is then the workers in these establishments. We then calculate the observed segregation measures, and compute simulated segregation the same way as described earlier, with 100 Monte Carlo simulations, so that there are 100 iterations within each iteration of the bootstrap. Finally, we collect the information on the empirical distribution of the effective segregation measures.

Given that this procedure is very intensive computationally, we did not carry it out for all of the estimates presented in the paper. Instead, because the estimates in Table 8 are based on the smallest samples, we carried out a detailed analysis for these estimates. We computed the 100 bootstraps for each of the samples in columns (1) and (2) of Table 8. Looking at the results for the sample of Hispanic and non-Hispanic workers who speak English poorly, in column (2), the effective segregation measures were estimated reasonably precisely, with a standard deviation of 1.27 for the co-worker measure and 1.50 for the Duncan index.[40] The standard deviations of the effective segregation measures for the larger sample of Hispanics with good English relative to those with poor English (column (1)) were smaller, at 0.57 for

[40] However, the observed empirical distribution was relatively far from normal. For the co-worker segregation measure the 95-percent confidence interval ranged from 2.8 below the estimate of 48.37 to 1.5 above (so the range is less than plus or minus 1.96 times the standard deviation, but the distribution is not symmetric). For the Duncan index the 95-percent confidence interval ranged from 3.4 below the estimate of 52.77 to 1.4 above (again non-symmetric about the estimate, and again with the range smaller than plus or minus 1.96 times the standard deviation).

the co-worker measure and 0.76 for the Duncan index.

Finally, in order to assess whether the differences in estimated effective segregation between samples are statistically significant, we pair each of the 100 bootstraps across the two samples, calculate the difference in the segregation measures across the samples for each bootstrap, and calculate the standard deviation of the difference in the segregation measures across columns. So while the effective co-worker segregation measures in Table 8 differ across the two samples by 19.9 (28.5 versus 48.4), the standard deviation of the bootstrapped difference in these measures across columns is 1.46, indicating that the observed difference is highly statistically significant. Similarly, the difference in the effective Duncan index across the two columns is 15.4, while the standard deviation of the bootstrapped difference is 1.73.

Thus, especially given the far large sample sizes in the other tables in the paper, it seems clear that the differences in effective segregation measures that we obtain in the paper are generally highly statistically significant. To reinforce this conclusion, for one particular case we verified that with the much larger samples used in many of our analyses, the precision of the estimates is even greater. Specifically, for the analysis of overall Hispanic-white segregation, based on data on over 1.7 million workers in over 300,000 establishments, the standard deviation was 0.24 for the co-worker segregation measure and 0.50 for the Duncan index.

References

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In Handbook of Labor Economics, Vol. 3, eds. Ashenfelter and Card (Amsterdam: Elsevier), pp. 3143-259.

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 1999. "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." In The Creation and Analysis of Employer-Employee Matched Data, eds. Haltiwanger, Lane, Spletzer, Theeuwes, and Troske (Amsterdam: Elsevier Science B.V.), pp. 175-203.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2002. "What Drives Racial Segregation? Evidence from the San Francisco Bay Area Using Micro-Census Data." Unpublished paper, Yale University.

Bertrand, Marianne, and Sendhil Mullainathan. 2003. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." National Bureau of Economics Working Paper No. 9873.

Becker, Gary S. 1971. The Economics of Discrimination, Second Edition (Chicago: University of Chicago Press).

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics*, Vol. 13, No. 2, April, pp. 309-34.

Blau, Francine D. 1977. Equal Pay in the Office (Lexington, MA: Heath).

Boisso, Dale, Kathy Hayes, Joseph Hirschberg, and Jacques Silber. 1994. "Occupational Segregation in the Multidimensional Case." *Journal of Econometrics*, Vol. 61, No. 1, March, pp. 161-71.

Brown, Charles, and James Medoff. 1989. "The Employer Size Wage Effect." *Journal of Political Economy*, Vol. 97, No. 5, October, pp. 1027-59.

Cabrales, Antonio, and Antoni Calvó-Armengol. 2002. "Social Preferences and Skill Segregation." Unpublished paper, Universitat Pompeu Fabra.

Cain, Glen. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey." In Handbook of Labor Economics, Vol. 1, eds. Ashenfelter and Layard (Amsterdam: North-Holland), pp. 693-785.

Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics*, Vol. 107, No. 1, February, pp. 151-200.

Carrington, William J., and Kenneth R. Troske. 1997. "On Measuring Segregation in Samples with Small Units." *Journal of Business & Economic Statistics*, Vol. 15, No. 4, October, pp. 402-9.

Carrington, William J., and Kenneth R. Troske. 1998. "Interfirm Segregation and the Black-White Wage Gap." *Journal of Labor Economics*, Vol. 51, No. 2, April, pp. 445-64.

Cortese, Charles, F., R. Frank Falk, and Jack K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review*, Vol. 51, No. 4, August, pp. 630-7.

C ross, Harry, Genevieve Kenney, Jane Mell, and Wendy Zimmerman. 1990. Employer Hiring Practices:

Differential Treatment of Hispanic and Anglo Job Seekers (Washington, DC: Urban Institute Press).

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." Journal of Political Economy, Vol 107, No. 3, June, pp. 455-506.

Darity, William A., Jr., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 63-92.

Davis, Steve J., John Haltiwanger, Lawrence F. Katz, and Robert Topel. 1991. "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986." *Brookings Papers on Economic Activity: Microeconomics*, Vol. 1, pp. 115-200.

Donohue, John J., and James Heckman. 1991. "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks." *Journal of Economic Literature*, Vol. 29, No. 4, December, pp. 1603-43.

Duncan, Otis D., and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indices." *American Sociological Review*, Vol. 20, No. 2, April, pp. 210-7.

Estlund, Cynthia. 2003. Working Together: How Workplace Bonds Strengthen a Diverse Democracy (New York: Oxford University Press).

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 1998. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." NBER Working Paper No. 6803.

Fryer, Roland G., and Steven D. Levitt. 2003. "The Causes and Consequences of Distinctively Black Names." NBER Working Paper No. 9938.

Greenhouse, Steven. 2003. "At a Factory in Houston, Hispanics Fight to Work in Coveted Department." *New York Times*, February 9, p. 14.

Groshen, Erica L. 1991. "The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?" *Journal of Human Resources*, Vol. 26, No. 3, Summer, pp. 457-72.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 101-16.

Hellerstein, Judith, and David Neumark. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." Forthcoming in *Annales d'Economie et de Statistique*.

Higgs, Robert. 1977. "Firm-Specific Evidence on Racial Wage Differentials and Workforce Segregation." *American Economic Review*, Vol. 67, No. 2, March, pp. 236-45.

Hirsch, Barry T., and David A. Macpherson. 2003. "Wages, Sorting on Skill, and the Racial Composition of Jobs." IZA Discussion Paper No. 741.

Ihlanfeldt, Keith, and David Sjoquist. 1990. "Job Accessibility and Racial Differences in Youth Employment Rates." *American Economic Review*, Vol. 80, No. 1, March, pp. 267-76.

James, Daniel R., and Karl E. Taeuber. 1985. "Measures of Segregation." In Sociological Methodology, ed. Brandon Tuma (San Francisco: Jossey-Bass), pp. 1-32.

King, Mary C. 1992. "Occupational Segregation by Race and Sex, 1940-1988." *Monthly Labor Review*, April, pp. 30-7.

Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." National Bureau of Economic Research Working Paper No. 5718.

Massey, Douglas, and Nancy Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." American Sociological Review, Vol. 52, No. 6, pp. 802-25.

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy*, Vol. 104, No. 5, October, pp. 869-95.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." *Journal of Economic Perspectives*, Vol. 4, No. 4, Fall, pp. 25-45.

Rivera, Elaine. 2003. "Area Bosses Try to Bridge Language Gaps." *Washington Post*, May 6, p. B1.

Saint-Paul, Gilles. 2001. "On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks." *Journal of Political Economy*, Vol. 109, No. 1, February, pp. 1-37.

Turner, Margery Austin, Michael Fix, and Raymond J. Struyk. 1991. Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring (Washington, DC: Urban Institute Press).

U.S. Census Bureau. "Census Geographic Glossary." http://www.census.gov/geo/lv4help/cengeoglos.html (viewed July 3, 2003).

U.S. Census Bureau, "Census Tracts and Block Numbering Areas." http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed May 10, 2004).

Watts, Martin J. 1995. "Trends in Occupational Segregation by Race and Gender in the U.S.A., 1983-92: A Multidimensional Approach." *Review of Radical Political Economics*, Vol. 27, No. 4, Fall, pp. 1-36.

Welch, Finis. 1990. "The Employment of Black Men." *Journal of Labor Economics*, Vol. 8, No. 2, April, pp. S26-S75.

Winship, Christopher. 1977. "A Revaluation of Indexes of Residential Segregation." *Social Forces*, Vol. 55, No. 4, June, pp. 1058-66.

Table 1: Means of Worker Characteristics

	SEDF (1)	Full DEED (2)
Age	37.08	37.51
	(12.78)	(12.23)
Female	0.46	0.47
Married	0.60	0.65
White	0.82	0.86
Hispanic	0.07	0.05
Black	0.08	0.05
Full-time	0.77	0.83
Number of kids (if female)	1.57	1.53
	(1.62)	(1.55)
High school diploma	0.34	0.33
Some college	0.30	0.32
B.A.	0.13	0.16
Advanced degree	0.05	0.05
Ln(hourly wage)	2.21	2.30
	(0.70)	(0.65)
Hourly wage	12.10	12.89
	(82.19)	(37.07)
Hours worked in 1989	39.51	40.42
	(11.44)	(10.37)
Weeks worked in 1989	46.67	48.21
	(11.05)	(9.35)
Earnings in 1989	22,576	25,581
	(26,760)	(29,475)
Industry:		
Mining	0.01	0.01
Construction	0.07	0.04
Manufacturing	0.25	0.33
Transportation	0.08	0.05
Wholesale	0.05	0.07
Retail	0.20	0.17
FIRE	0.08	0.08
Services	0.26	0.24
Observations	12,143,183	3,291,213

Standard deviations of continuous variables are reported in parentheses.

Table 2: Means for Establishments

	SSEL	Full DEED
Total employment	17.57	52.68
	(253.75)	(577.39)
Establishment size:		
1 - 25	0.88	0.65
26 - 50	0.06	0.15
51 - 100	0.03	0.10
101 +	0.03	0.10
Industry:		
Mining	0.00	0.01
Construction	0.09	0.07
Manufacturing	0.06	0.13
Transportation	0.04	0.05
Wholesale	0.08	0.11
Retail	0.25	0.24
FIRE	0.09	0.10
Services	0.28	0.26
In MSA	0.81	0.82
Census Region:		
North East	0.06	0.06
Mid Atlantic	0.16	0.15
East North Central	0.16	0.20
West North Central	0.07	0.08
South Atlantic	0.18	0.16
East South Central	0.05	0.05
West South Central	0.10	0.10
Mountain	0.06	0.05
Pacific	0.16	0.15
Payroll ($1000)	397	1,358
	(5,064)	(10,329)
Payroll/total employment	21.02	24.24
	(1,385.12)	(111.79)
Share of employees matched	–	0.17
Multi-unit establishment	0.23	0.42
Observations	5,237,592	972,436

Standard deviations of continuous variables are reported in parentheses. 55 establishments in the DEED sample do not have valid county data from the SSEL. For these 55, the workers reported place of work was used to determine MSA status.

Table 3: Black-White Segregation

	Establishment racial composition:			
	U.S., CMSA/MSA sample	Within CMSA/MSA	U.S., MSA/PMSA, sample	Within MSA/PMSA
	%Black	%Black	%Black	%Black
	(1)	(2)	(3)	(4)
Co-worker segregation				
Observed segregation				
Black workers (B_B^O)	33.4	33.4	34.4	34.4
White workers (W_B^O)	5.0	5.0	5.0	5.0
Difference (CW^O)	28.4	28.4	29.4	29.4
Simulated segregation				
Black workers (B_B^S)	23.1	26.5	23.5	27.5
White workers (W_B^S)	5.7	5.5	5.9	5.6
Difference (CW^S)	17.3	21.0	17.7	21.9
Effective segregation, $[\{CW^O - CW^S\}/\{100 - CW^S\}]\times 100$	**13.4**	**9.5**	**14.2**	**9.6**
Duncan index				
Observed (DI^O)	68.5	66.4	69.2	66.7
Simulated (DI^S)	54.1	58.4	54.3	59.3
Effective segregation, $[\{DI^O - DI^S\}/\{100 - DI^S\}]\times 100$	**31.4**	**19.1**	**32.6**	**18.4**
Number of workers	1,735,614	1,735,614	1,618,876	1,618,876
Number of establishments	300,908	300,908	285,988	285,988

Calculations are for establishments with two or more matched workers. For the CMSA/MSA (MSA/PMSA) sample of workers, the median number of workers matched to an establishment is 9 (9), and the median share of the workforce matched 8.1 (7.9) percent. For the sample of establishments, the median number of matched workers is 3 (3), and the median share of the workforce matched is 9.1 (8.8) percent. While the median numbers of workers matched are low, this arises because there are many small establishments in the data; the shares of the workforce matched range from 8.0 to 9.1 percent, relative to a hypothetical maximum of 16.7 percent, given that only 1/6 of workers receive the Census long form. All medians are reported as “fuzzy medians” to comply with confidentiality restrictions; but they are extremely close to actual medians.

Table 4: Black-White Segregation by Skill Level, Within MSA/PMSA

	Establishment racial and skill composition:										
Black workers-white workers		White workers, low educ.-white workers, high educ.		Black workers, low educ.-white workers, high educ.		Black workers, low educ.-black workers, high educ.		Black workers, low educ.-white workers, low educ.		Black workers, high educ.-white workers, high educ.	
	%Black		%White, low educ.		%Black, low educ.		%Black, low educ.		%Black, low educ.		%Black, high educ.
	(1)		(2)		(3)		(4)		(5)		(6)
Co-worker segregation											
Observed segregation											
Black workers	34.4	White workers, low educ.	61.8	Black workers, low educ.	40.9	Black workers, low educ.	67.9	Black workers, low educ.	44.5	Black workers, high educ.	31.6
White workers	5.0	White workers, high educ.	26.9	White workers, high educ.	3.8	Black workers, high educ.	32.0	White workers, low educ.	5.8	White workers, high educ.	4.6
Difference	29.4		34.9		37.1		35.9		39.1		27.0
Simulated segregation											
Black workers	27.5	White workers, low educ.	53.9	Black workers, low educ.	27.9	Black workers, low educ.	62.6	Black workers, low educ.	35.4	Black workers, high educ.	26.6
White workers	5.6	White workers, high educ.	32.4	White workers, high educ.	4.7	Black workers, high educ.	37.3	White workers, low educ.	6.3	White workers, high educ.	4.9
Difference	21.9		21.5		23.2		25.4		29.1		21.7
Effective segregation	**9.6**		**17.0**		**18.0**		**14.1**		**14.1**		**6.8**
Duncan index											
Observed	66.7		48.2		74.0		46.7		73.6		66.4
Simulated	59.3		31.4		62.9		35.3		66.0		61.1
Effective segregation	**18.4**		**24.5**		**29.9**		**17.7**		**22.4**		**13.7**
Number of workers	1,618,876		1,500,322		854,140		83,401		588,920		860,229
Number of establishments	285,988		273,084		164,073		19,062		131,387		163,231

Calculations are for establishments with two or more matched workers. "Low education" refers to those with no higher than a high school degree. The "black workers-white workers" calculations in column (1) are the same as in Table 3, column (4).

Table 5: Black-White Segregation by Skill Level, Region of Birth, and Birth Cohort, Within MSA/PMSA, Currently Residing in North

	Establishment racial and skill composition:							
	Black workers, low educ., born in North before 1950- white workers, high educ., born before 1950		Black workers, low educ., born in South before 1950- white workers, high educ., born before 1950		Black workers, low educ., born in North after 1950- white workers, high educ., born after 1950		Black workers, low educ., born in South after 1950- white workers, high educ., born after 1950	
		%Black, low educ., born North before 1950		%Black, low educ., born South before 1950		%Black, low educ., born North after 1950		%Black, low educ., born North after 1950
		(1)		(2)		(3)		(4)
Co-worker segregation								
Observed segregation								
Black workers	Black workers, high educ.	26.0	Black workers, low educ.	34.5	White workers, low educ.	33.1	Black workers, low educ.	26.9
White workers	White workers, high educ.	1.6	White workers, high educ.	2.7	White workers, high educ.	1.6	Black workers, high educ.	0.6
Difference		24.5		31.8		31.6		26.3
Simulated segregation								
Black workers	Black workers, high educ.	23.9	Black workers, low educ.	25.8	White workers, low educ.	23.6	Black workers, low educ.	22.1
White workers	White workers, high educ.	1.6	White workers, high educ.	3.1	White workers, high educ.	1.8	Black workers, high educ.	0.6
Difference		22.3		22.8		21.9		21.5
Effective segregation		**2.8**		**11.7**		**12.5**		**6.1**
Duncan index								
Observed		83.0		79.1		83.6		89.9
Simulated		80.2		72.2		78.0		87.6
Effective segregation		**14.1**		**24.6**		**25.7**		**18.5**
Number of workers		157,192		162,913		311,553		299,932
Number of establishments		33,817		35,020		66,120		63,486

Calculations are for establishments with two or more matched workers. “Low education” refers to those with no higher than a high school degree.

Table 6: Hispanic-White Segregation

	Establishment ethnic composition:			
	U.S., CMSA/MSA sample	Within CMSA/MSA	U.S., MSA/PMSA, sample	Within MSA/PMSA
	%Hispanic	%Hispanic	%Hispanic	%Hispanic
	(1)	(2)	(3)	(4)
Co-worker segregation				
Observed segregation				
Hispanic workers (H_H^O)	48.9	48.9	50.7	50.7
White workers (W_H^O)	3.9	3.9	3.7	3.7
Difference (CW^O)	45.0	45.0	47.1	47.1
Simulated segregation				
Hispanic workers (H_H^S)	23.5	36.7	23.8	39.4
White workers (W_H^S)	5.8	4.8	5.7	4.5
Difference (CW^S)	17.7	32.0	18.1	34.8
Effective segregation, $[\{CW^O - CW^S\}/\{100 - CW^S\}]\times 100$	**33.2**	**19.2**	**35.4**	**18.7**
Duncan index				
Observed (DI^O)	78.4	73.5	79.7	74.2
Simulated (DI^S)	54.8	66.3	55.7	67.9
Effective segregation, $[\{DI^O - DI^S\}/\{100 - DI^S\}]\times 100$	**52.2**	**21.3**	**54.1**	**19.8**
Number of workers	1,747,719	1,747,719	1,625,953	1,625,953
Number of establishments	309,357	309,357	293,989	293,989

Calculations are for establishments with two or more matched workers. For the CMSA/MSA (MSA/PMSA) sample of workers, the median number of workers matched to an establishment is 9 (8), and the median share of the workforce matched is 8.0 (7.8) percent. For the sample of establishments, the median number of matched workers is 3 (3), and the median share of the workforce matched is 9.1 (8.8) percent.

Table 7: Skill Segregation Among Whites and Hispanics, Within MSA/PMSA

	Establishment ethnic and skill composition:				
White workers- Hispanic workers		White workers- Hispanic workers, good English		White workers- Hispanic workers, poor English	
	%Hispanic		%Hispanic, good English		%Hispanic, bad English
	(1)		(2)		(3)
Co-worker segregation					
Observed segregation					
Hispanic workers	50.7	Hispanic workers, good English	43.5	Hispanic workers, bad English	61.0
White workers	3.7	White workers	3.4	White workers	0.6
Difference	47.1		40.1		60.4
Simulated segregation					
Hispanic workers	39.4	Hispanic workers, good English	35.9	Hispanic workers, bad English	29.4
White workers	4.5	White workers	3.8	White workers	1.1
Difference	34.8		32.1		28.3
Effective segregation	**18.7**		**11.9**		**44.8**
Duncan index					
Observed	74.2		74.0		95.4
Simulated	67.9		69.8		88.0
Effective segregation	**19.8**		**13.8**		**62.1**
Number of workers	1,625,953		1,601,390		1,327,021
Number of establishments	293,989		289,719		244,534

Calculations are for establishments with two or more matched workers. “Good English” means that the respondent reports speaking English “very well” or “well”; “bad English” means the respondent reported speaking English “poorly” or “not at all.” The “white workers-Hispanic workers” calculations in column (1) are the same as in Table 6, column (4).

Table 8: Language Segregation, Within MSA/PMSA

	Establishment ethnic and skill composition:		
Hispanic workers, poor English- Hispanic workers, good English		Hispanic workers, poor English- non-Hispanic workers, poor English	
	%Hispanic, poor English		%Hispanic, poor English
	(1)		(2)
Co-worker segregation			
Observed segregation			
Hispanic workers, poor English	61.7	Hispanic workers, poor English	93.1
Hispanic workers, good English	11.4	Non-Hispanic workers, poor English	17.8
Difference	50.3		75.4
Simulated segregation			
Hispanic workers, poor English	46.4	Hispanic workers, poor English	86.7
Hispanic workers, good English	15.9	Non-Hispanic workers, poor English	34.5
Difference	30.5		52.3
Effective segregation	**28.5**		**48.4**
Duncan index			
Observed	69.8		82.7
Simulated	51.8		63.4
Effective segregation	**37.4**		**52.8**
Number of workers	**81,595**		**19,926**
Number of establishments	**21,933**		**6,393**

Calculations are for establishments with two or more matched workers.

Table 9: Language Segregation, Within MSA/PMSA, Sensitivity to Establishment Size

	Establishment ethnic and skill composition:						
Hispanic workers, poor English- Hispanic workers, good English				Hispanic workers, poor English- non-Hispanic workers, poor English			
	%Hispanic, poor English				%Hispanic, poor English		
	Employment > 10	Employment > 60	Employment > 170		Employment > 10	Employment > 60	Employment > 170
	(1)	(1')	(1")		(2)	(2')	(2")
Co-worker segregation							
Observed segregation							
Hispanic workers, poor English	70.6	62.7	54.1	Hispanic workers, poor English	95.9	94.2	91.4
Hispanic workers, good English	7.1	8.8	9.2	Non-Hispanic workers, poor English	7.5	10.4	13.6
Difference	63.5	53.9	44.9		88.4	83.8	77.8
Simulated segregation							
Hispanic workers, poor English	61.9	51.2	40.5	Hispanic workers, poor English	91.3	89.0	85.4
Hispanic workers, good English	9.2	11.6	11.9	Non-Hispanic workers, poor English	15.7	19.7	23.2
Difference	52.7	39.6	28.6		75.7	69.4	62.2
Effective segregation	**22.9**	**23.6**	**22.9**		**52.3**	**47.0**	**41.3**
Duncan index							
Observed	80.3	74.4	69.6		90.8	86.7	81.0
Simulated	69.9	60.1	52.0		78.8	72.2	63.5
Effective segregation	**34.7**	**35.9**	**36.5**		**56.5**	**52.3**	**47.9**
Number of workers	122,180	78,358	46,080		36,707	23,162	12,184
Number of establishments	63,927	30,469	12,648		23,080	12,248	5,346

Calculations are for establishments with two or more matched workers. The employment cutoffs chosen are approximately the 25th, 50th, and 75th percentiles of the establishment size distribution.

Appendix Table A1: Black-White and Hispanic-White Segregation, Within MSA/PMSA, Sensitivity to Alternative Minimum Numbers of Matched Workers

	All establishments	≥ 5 matched workers	All establishments	≥ 5 matched workers
	%Black	%Black	%Hispanic	%Hispanic
	(1)	(2)	(3)	(4)
<u>Co-worker segregation</u>				
Observed segregation				
Hispanic workers (H_H^O)/Black workers (B_B^O)	44.3	26.4	60.9	41.3
White workers (W_H^O)/(W_B^O)	4.0	6.2	3.0	4.1
Difference (CW^O)	40.3	20.2	57.9	37.2
Simulated segregation				
Hispanic workers (H_H^S)/Black workers (B_B^S)	41.8	18.6	52.8	29.6
White workers (W_H^S)/(W_B^S)	4.2	6.9	3.6	4.9
Difference (CW^S)	37.6	11.7	49.2	24.7
Effective segregation, $[\{CW^O - CW^S\}/\{100 - CW^S\}]\times 100$	**4.3**	**9.6**	**17.1**	**16.6**
<u>Duncan index</u>				
Observed (DI^O)	72.9	55.0	79.2	65.2
Simulated (DI^S)	67.9	43.2	74.4	56.7
Effective segregation, $[\{DI^O - DI^S\}/\{100 - DI^S\}]\times 100$	**15.6**	**20.9**	**18.8**	**19.6**
Number of workers	2,006,415	1,083,322	2,019,727	1,074,570
Number of establishments	672,242	76,013	686,835	77,822

Note that in adding establishments with only one matched worker, the number of workers and the number of establishments do not increase (relative to Tables 3 and 6) by the exact same amount. This is because of the handful of small metropolitan areas that are dropped from each sample if the workers in that metropolitan area are completely homogeneous. With different cutoffs for the minimum number of matched workers, the set of metropolitan areas so affected can change slightly.

Appendix Table A2: Black-White Segregation by Skill Level, Within MSA/PMSA, Sensitivity to Alternative Minimum Numbers of Matched Workers

	Establishment racial and skill composition:																
Black workers-white workers			White workers, low educ.-white workers, high educ.			Black workers, low educ.-white workers, high educ.			Black workers, low educ.-black workers, high educ.			Black workers, low educ.-white workers, low educ.			Black workers, high educ.-white workers, high educ.		
	%Black			%White, low educ.			%Black, low educ.			%Black, low educ.			%Black, low educ.			%Black, high educ.	
	All est.'s	≥ 5 matched workers		All est.'s	≥ 5 matched workers		All est.'s	≥ 5 matched workers		All est.'s	≥ 5 matched workers		All est.'s	≥ 5 matched workers		All est.'s	≥ 5 matched workers
	(1)	(1')		(2)	(2')		(3)	(3')		(4)	(4')		(5)	(5')		(6)	(6')
Co-worker segregation																	
Observed segregation																	
Black workers	44.3	26.4	White workers, low educ.	69.6	55.6	Black workers, low educ.	54.7	29.0	Black workers, low educ.	80.3	59.0	Black workers, low educ.	57.5	35.3	Black workers, high educ.	45.7	22.3
White workers	4.0	6.2	White workers, high educ.	21.6	29.4	White workers, high educ.	2.8	4.6	Black workers, high educ.	19.7	37.4	White workers, low educ.	3.7	7.0	White workers, high educ.	3.4	5.8
Difference	40.3	20.2		48.0	26.2		51.9	24.4		60.6	21.6		53.8	28.3		42.3	16.4
Simulated segregation																	
Black workers	41.8	18.6	White workers, low educ.	63.1	47.2	Black workers, low educ.	46.8	16.5	Black workers, low educ.	77.0	53.8	Black workers, low educ.	55.3	24.9	Black workers, high educ.	45.5	16.2
White workers	4.2	6.9	White workers, high educ.	26.2	35.0	White workers, high educ.	3.3	5.5	Black workers, high educ.	23.0	42.1	White workers, low educ.	3.9	8.1	White workers, high educ.	3.4	6.3
Difference	37.6	11.7		36.9	12.1		43.5	11.1		54.0	11.7		51.3	16.8		42.1	9.9
Effective segregation	**4.3**	**9.6**		**17.6**	**16.0**		**14.9**	**15.0**		**14.3**	**11.1**		**5.0**	**13.8**		**0.3**	**7.2**
Duncan index																	
Observed	72.9	55.0		58.8	40.6		80.6	62.5		67.4	34.0		81.3	61.1		74.8	51.9
Simulated	67.9	43.2		45.0	20.1		73.0	46.1		60.2	21.4		77.3	48.2		72.0	42.3
Effective segregation	**15.6**	**20.9**		**25.1**	**25.7**		**28.1**	**30.3**		**18.1**	**16.0**		**17.5**	**24.9**		**10.0**	**16.8**
Number of workers	2,006,415	1,083,322		1,877,716	986,702		1,157,105	532,682		135,617	46,068		839,348	331,634		1,161,477	542,241
Number of estab.'s	672,242	76,013		650,478	70,857		463,464	36,242		71,243	3,904		379,387	29,072		462,340	36,265

The analysis in this table parallels that in Table 4.

Appendix Table A3: Black-White Segregation by Skill Level, Region of Birth, and Birth Cohort, Within MSA/PMSA, Currently Residing in North, Sensitivity to Alternative Minimum Numbers of Matched Workers

	Establishment racial and skill composition:											
	Black workers, low educ., born in North before 1950- white workers, high educ., born before 1950			Black workers, low educ., born in South before 1950- white workers, high educ., born before 1950			Black workers, low educ., born in North after 1950- white workers, high educ., born after 1950			Black workers, low educ., born in South after 1950- white workers, high educ., born after 1950		
		%Black, low educ., born North before 1950			%Black, low educ., born South before 1950			%Black, low educ., born North after 1950			%Black, low educ., born North after 1950	
		All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers
		(1)	(1')		(2)	(2')		(3)	(3')		(4)	(4)
Co-worker segregation												
Observed segregation												
Black workers	Black workers, high educ.	48.5	14.5	Black workers, low educ.	53.4	23.6	White workers, low educ.	56.2	16.5	Black workers, low educ.	52.8	11.4
White workers	White workers, high educ.	1.0	2.1	White workers, high educ.	1.6	3.6	White workers, high educ.	1.0	1.8	Black workers, high educ.	0.4	0.7
Difference		47.5	12.4		51.8	20.0		55.2	14.7		52.4	10.7
Simulated segregation												
Black workers	Black workers, high educ.	52.5	10.9	Black workers, low educ.	53.2	13.4	White workers, low educ.	48.5	10.1	Black workers, low educ.	48.0	8.4
White workers	White workers, high educ.	0.9	2.2	White workers, high educ.	1.7	4.1	White workers, high educ.	1.2	1.9	Black workers, high educ.	0.4	0.7
Difference		51.6	8.7		51.5	9.3		47.3	8.2		47.6	7.7
Effective segregation		**-7.9**	**4.1**		**0.5**	**11.8**		**14.9**	**7.1**		**9.3**	**3.3**
Duncan index												
Observed		89.6	72.0		86.9	66.9		89.1	73.5		93.3	83.3
Simulated		88.7	65.8		84.0	53.8		85.1	65.7		91.6	80.1
Effective segregation		**7.7**	**18.0**		**18.3**	**28.3**		**26.9**	**22.6**		**20.3**	**16.2**
Number of workers		260,014	88,521		266,185	93,010		464,652	179,387		447,107	171,763
Number of establishments		134,163	6,334		136,956	6,626		216,290	13,326		208,343	12,739

The analysis in this Table parallels that in Table 5. For the negative effective segregation measure in column (1'), see footnote 23.

Appendix Table A4: Skill Segregation Among Whites and Hispanics, Within MSA/PMSA, Sensitivity to Alternative Minimum Numbers of Matched Workers

	Establishment ethnic and skill composition:							
White workers- Hispanic workers			White workers- Hispanic workers, good English			White workers- Hispanic workers, poor English		
	%Hispanic			%Hispanic, good English			%Hispanic, bad English	
	All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers
	(1)	(1')		(2)	(2')		(3)	(3')
Co-worker segregation								
Observed segregation								
Hispanic workers	60.9	41.3	Hispanic workers, good English	56.3	32.5	Hispanic workers, bad English	70.7	50.9
White workers	3.0	4.1	White workers	2.7	3.7	White workers	0.5	0.6
Difference	57.9	37.2		53.6	28.8		70.2	50.3
Simulated segregation								
Hispanic workers	52.8	29.6	Hispanic workers, good English	50.9	25.3	Hispanic workers, bad English	47.8	16.3
White workers	3.6	4.9	White workers	3.0	4.1	White workers	0.8	1.0
Difference	49.2	24.7		47.9	21.2		47.0	15.3
Effective segregation	**17.1**	**16.6**		**10.9**	**9.6**		**43.8**	**41.3**
Duncan index								
Observed	79.2	65.2		79.0	64.5		96.4	93.5
Simulated	74.4	56.7		75.9	58.7		90.7	83.7
Effective segregation	**18.8**	**19.6**		**12.9**	**14.0**		**61.3**	**60.4**
Number of workers	2,019,727	1,074,570		1,993,569	1,057,102		1,692,339	841,494
Number of establishments	686,835	77,822		680,970	76,184		590,424	61,287

The analysis in this table parallels that in Table 7.

Appendix Table A5: Language Segregation, Within MSA/PMSA, Sensitivity to Alternative Minimum Numbers of Matched Workers

	Establishment ethnic and skill composition:				
Hispanic workers, poor English- Hispanic workers, good English			Hispanic workers, poor English- non-Hispanic workers, poor English		
	%Hispanic, poor English			%Hispanic, poor English	
	All estab.'s	≥ 5 matched workers		All estab.'s	≥ 5 matched workers
	(1)	(1')		(2)	(2')
Co-worker segregation					
Observed segregation					
Hispanic workers, poor English	72.6	53.3	Hispanic workers, poor English	96.2	91.1
Hispanic workers, good English	6.3	14.6	Non-Hispanic workers, poor English	6.4	29.5
Difference	66.3	38.7		89.8	61.6
Simulated segregation					
Hispanic workers, poor English	65.4	35.7	Hispanic workers, poor English	91.7	86.3
Hispanic workers, good English	7.9	20.1	Non-Hispanic workers, poor English	14.2	45.7
Difference	57.5	15.6		77.5	40.6
Effective segregation	**20.7**	**27.4**		**54.7**	**35.4**
Duncan index					
Observed	82.5	58.2		91.8	70.1
Simulated	73.7	31.7		80.3	46.8
Effective segregation	**33.5**	**38.7**		**58.4**	**43.8**
Number of workers	140,543	37,086		41,354	6,147
Number of establishments	80,653	3,924		27,348	827

The analysis in this table parallels that in Table 8.