Changes in Workplace Segregation in the United States between 1990 and 2000: Evidence from Matched Employer-Employee Data*

Judith Hellerstein
University of Maryland and NBER

David Neumark
UCI,
NBER, and IZA

Melissa McInerney
University of Maryland and U.S. Census Bureau

September 2006

*This research was funded by NICHD grant R01HD042806. The analysis and results presented in this paper are attributable to the authors and do not necessarily reflect concurrence by the Center for Economics Studies, Bureau of the Census. This paper has undergone a more limited review by the Census Bureau than its official publications. It has been screened to ensure that no confidential data are revealed.

I. Introduction

In recent work we have constructed and described the 1990 Decennial Employer-Employee Dataset (DEED) based on matching records in the 1990 Decennial Census of Population to a Census Bureau list of most business establishments in the United States. We have used the 1990 DEED to estimate earnings and productivity differentials in manufacturing by demographic and skill group (Hellerstein and Neumark, 2004), to study the influence of language skills on workplace segregation and wages (Hellerstein and Neumark, 2003), to document the extent of workplace segregation by race and ethnicity and to assess the contribution of residential segregation as well as skill to this segregation (Hellerstein and Neumark, 2005), and to study spatial mismatch (Hellerstein and Neumark, in progress).

We just recently completed the construction of the 2000 Beta-DEED[1] (based on the 2000 Census of Population). In this paper, we use the 1990 and 2000 DEED's to measure changes in establishment-level workplace segregation over the intervening decade, an analysis for which the DEED's are uniquely well-suited. We study segregation by education, by race and Hispanic ethnicity, and by sex. With respect to segregation by race and ethnicity, this work is complementary to a flurry of research studying changes in residential segregation from 1990 to 2000 (Glaeser and Vigdor, 2001; Iceland and Weinberg, 2002; and McConville, et al., 2001). As we have suggested elsewhere (and see Estlund, 2003), however, workplace segregation may be far more salient for interactions between racial and ethnic groups than is residential segregation. The boundaries used in studying residential segregation may not capture social interactions, and are to some extent explicitly drawn to accentuate segregation among different groups; 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

[1] The 2000 Beta-DEED is an internal U.S. Census Bureau dataset that will ultimately become part of an integrated matched employer-employee database at the U.S. Census Bureau. The new integrated data will have characteristics of the Decennial Employer-Employee Database (DEED) and the Longitudinal Employer-Household Dynamics Program (LEHD). Hereafter, the 2000 Beta-DEED will be referred to as the 2000 DEED.

conditions."[2] 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. 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.[3] 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), but data constraints truly prevent saying much of anything about segregation along these lines.

Segregation is potentially important for a number of reasons. Aside from general social issues regarding integration between different groups, labor market segregation by race and ethnicity 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, 1998a; Bayard, et al., 1999; King, 1992; Watts, 1995; Higgs, 1977), and the same is true of labor market segregation by sex (Bayard, et al., 2003; Blau, 1977; and Groshen, 1991).[4] There has generally been less attention paid to segregation by education, but in our earlier work (Hellerstein and Neumark, 2005), we documented rather extensive segregation by education (as well as language, which we consider in this paper) in the 1990 DEED.

Measuring changes in workplace segregation along these lines is of interest for a number of reasonsl. First, although much attention has been paid to changes in residential segregation – of which there is evidence of modest declines from 1990 to 2000 – changes in workplace segregation may be more

[2] U.S. Census Bureau, www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed April 27, 2005). Echenique and Fryer (2005) develop a segregation index that relies much less heavily on ad-hoc definitions of geographical boundaries.

[3] Moreover, industry code, the closest proxy in public-use data to an establishment identifier, is a very crude measure to use to examine segregation. For example, we calculate that racial and ethnic segregation at the three-digit industry level in the DEED is typically on the order of one-third as large as the establishment-level segregation we document below.

salient to understanding changing social forces. Second, aside from the relative importance of workplace and residential segregation, in the United States there are extensive efforts to reduce labor market discrimination, and therefore measuring changes in workplace segregation is one indicator of the success of these efforts. Finally, there has been speculation that one source of increased wage inequality by education is increased segregation by skill (e.g., Kremer and Maskin, 1996).[5] A comparison of education segregation between 1990 and 2000 possibly can shed some light on this hypothesis, although relatively more of the run-up in wage inequality occurred prior to 1990 (Autor, et al., 2005).

We measure changes in segregation using the 1990 and 2000 Decennial Employer-Employee Databases (DEED's). For each year, the DEED is based on matching records in the Decennial Census of Population for that year to a Census Bureau list of most business establishments in the United States. The matching yields data on multiple workers matched to establishments, providing the means to measure workplace segregation (and changes therein) in the United States based on a large, fairly representative data set. In addition, the data from the Decennial Census of Population provides the necessary information on race, ethnicity, etc. Thus, data from the 1990 and 2000 DEED's provides unparalleled opportunities to study changes in workplace segregation by skill, race, ethnicity, and sex.[6]

II. The 1990 and 2000 DEED's

The analysis in this paper is based on the 1990 and 2000 DEED's, which we have created at the Center for Economic Studies at the U.S. Bureau of the Census. We have described the construction of the 1990 DEED in detail elsewhere (in particular, Hellerstein and Neumark, 2003). The construction of the

[4] 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.

[5] 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 (L_1) with labor input equal to one efficiency unit, and skilled workers (L_2) 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. Thus, as the returns to education rise (q increases), there may be increased segregation by education.

2000 DEED follows the same procedures, and our detailed investigation of the 2000 data thus far has indicated that no new serious problems arise that require different methods for 2000. Thus, in this section we simply provide a quick overview of the construction of the datasets.

The DEED for each year is formed by matching workers to establishments. The workers are drawn from the Sample Edited Detail File (SEDF), which contains all individual responses to the Decennial Census of Population one-in-six Long Form. The establishments are drawn from the Census Bureau's Business Register list (BR), formerly known as the Standard Statistical Establishment Listing or BR, an administrative database containing information for all business establishments operating in the United States in each year.

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, which contains the information written on the questionnaires by Long-Form respondents, but not actually captured in the SEDF. The BR is an annually-updated list of all business establishments with one or more employees operating in the United States. The Census Bureau uses the BR as a sampling frame for its Economic Censuses and Surveys, and continuously updates the information it contains. The BR 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 BR. Because the name and address information on the Write-In file is also available for virtually all employers in the BR, nearly all of the establishments in the BR 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

---

[6] Carrington and Troske (1998a, 1998b) 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.

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 BR. Standardization of addresses in the establishment and worker files helps to eliminate differences in how data are reported. 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. The value of parsing the addresses into multiple pieces is that we can match on various combinations of these components. We supplement the AutoStan software by creating an acronym for each company name, and added this variable to the list of matching components.[7]

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

---

[7] For 2000, we also added standard acronyms or abbreviations for cities, such as NY or NYC and LA. However, this added a negligible number of additional matches, so we did not go back and do the same for the 1990 DEED.

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 BR. 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. 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 matches that are correct with a high probability are linked.[8] Because the AutoMatch algorithm is not exact there is always a range of quality of matches, and we therefore are cautious in accepting linked record pairs. Our general strategy is 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 choose matching algorithms based on substantial experimentation and visual inspection of many thousands of records.

[8] One might also considering trying to impute matches where this strategy fails, by matching based on imputed place of work. However, this turns out to be problematic. Even imputing place of work at the level of the census tract is not easy. For example, there are workers in the SEDF that we are able to match to an employer in the DEED using name and address information whose place of work code actually is allocated in the SEDF. For these workers, the allocated census tract in the SEDF disagrees with the BR census tract of the matched establishment in more than half the cases.

The final result is an extremely large data set, for each year, of workers matched to their establishment of employment. The 1990 DEED consists of information on 3.29 million workers matched to around 972,000 establishments, accounting for 27.1 percent of workers in the SEDF and 18.6 percent of establishments in the BR. The 2000 DEED consists of information on 4.09 million workers matched to around 1.28 million establishments, accounting for 29.1 percent of workers in the SEDF and 22.6 percent of establishments in the BR.[9]

In Table 1 we provide descriptive statistics for the matched workers from the DEED as compared to the SEDF. Columns (1) and (4) report summary statistics for the SEDF for the sample of workers who were eligible to be matched to their establishments, for 1990 and 2000 respectively. Columns (2) and (5) report summary statistics for the full DEED sample. For both years, the means of the demographic variables in the full DEED are quite close to the means in the SEDF across most dimensions. For example, for the 1990 data, female workers comprise 46 percent of the SEDF and 47 percent of the full DEED, and the number of children (for women) is 0.75 in the SEDF and 0.73 in the DEED. Nonetheless, there are cases of somewhat larger differences. For example, the percent female in the 2000 data is 46 in the SEDF, but 50 in the DEED. And the race and ethnic differences are larger in both years; for example, in 2000 the percent white is 78 in the SEDF versus 83 in the DEED, and correspondingly the share black (and also Hispanic) is lower in the DEED.

Part of the explanation for differences in racial and ethnic representation is that there are many individuals who meet our sample inclusion criteria but for whom the quality of the business address information in the Write-In file is poor, and race and ethnic differences in reporting account for part of the differences in representation. We suspect that the differences in business address information partially

[9] For both the DEED and SEDF we have excluded individuals as follows: with missing wages; who did not work in the year prior to the survey year (1989) or in the reference week for the Long Form of the Census; who did not report positive hourly wages; who did not work in one of the fifty states or the District of Columbia (even if the place of work was imputed); who 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 BR. (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 BR data for businesses in out-of-scope industries.)

reflect weaker labor market attachment among minorities, suggesting that the segregation results we obtain might best be interpreted as measuring of the extent of segregation among workers who have relatively high labor force attachment and high attachment to their employers.

The last eight rows of the table report on the industry distribution of workers. We do find some overrepresentation of workers in manufacturing, more so in 1990. The reasons for this are given below when we discuss establishment-level data.

Columns (3) and (6) report summary statistics for the workers in the DEED who comprise the sample from which we calculate segregation measures. The sample size reductions relative to columns (2) and (5) arise for two reasons. First, for reasons explained in the methods section, we exclude workers who do not live and work in the same Metropolitan Statistical Area/Primary Metropolitan Statistical Area (MSA/PMSA). Second, we exclude workers who are the only workers matched to their establishments, as there are methodological advantages to studying segregation in establishments where we observe at least two workers. The latter restriction effectively causes us to restrict the sample to workers in larger establishments, which is the main reason why some of the descriptive statistics are slightly different between the second and third columns (for example, slightly higher wages and earnings in columns (3) and (6).

In addition to comparing worker-based means, it is useful to examine the similarities across establishments in the BR and the DEED for each year. Table 2 shows descriptive statistics for establishments in each data set. As column (1) indicates, there are 5,237,592 establishments in the 1990 BR, and of these 972,436 (18.6 percent) also appear in the full DEED for 1990, as reported in column (2). For 2000, the percentage in the full DEED is somewhat higher (22.6). Because only one in six workers are sent Decennial Census Long Forms, as noted earlier, it is more likely that large establishments will be included in the DEED. One can see evidence of the bias toward larger employers by comparing the means across data sets for total employment. (This bias presumably also influences the distribution of workers and establishments across industries.) On average, establishments in the BR's have 18-19 employees, while the average in the DEED's is 49-53 workers. The distributions of establishments across industries in the DEED relative to the BR are similar to those for workers in the worker sample. In columns (3) and (6) we report descriptive statistics for establishments in the restricted DEED's, corresponding to the sample of workers in columns (3) and (6) of Table 1. In general, the summary statistics are quite similar between columns (2) and (3) and between columns (5) and (6), with an unsurprising right shift in the size distribution of establishments. Overall, the DEED samples are far more representative than previous detailed matched data sets for the United States (see Hellerstein and Neumark, 2003 for more).

Because the DEED captures larger establishments and because of our sample restrictions that accentuate this, our analysis focuses on larger establishments. So, for example, the first quartile of the establishment size distribution for workers in our analysis is approximately 41 workers in 1990 and 39 in 2000, whereas the first quartile of the employment-weighted size distribution of all establishments in the BR for each year is 19 in 1990 and 21 in 2001.[10] Although we acknowledge that it would be nice to be able to measure segregation in all establishments, this is not the data set with which to do that convincingly. Nonetheless, most legislation aimed at combating discrimination is directed at larger establishments; EEOC laws cover employers with 15 or more workers and affirmative action rules for federal contractors cover employers with 50 or more workers. Since policy has been directed at larger establishments, examining the extent of and changes in workplace segregation in larger establishments is important.

III. Methods

We focus our analysis on a measure of segregation that is based on measures of the percentages of workers in an individual's establishment, or workplace, in different demographic groups. Consider for clarity a dichotomous classification of workers (e.g., whites and Hispanics). For each worker in our sample, we compute the percentage of Hispanic workers with which that worker works, excluding the

[10] In order to adhere to U.S. Census Bureau confidentiality rules, these are "pseudo quartiles" based on averages of observations symmetrically distributed around the actual quartiles. The same applies to "pseudo medians" reported in some of the tables.

worker him or herself. Because we exclude an individual's own ethnicity in this calculation, our analysis of segregation is conducted on establishments where we observe at least two workers.

We then average these percentages separately for white workers in our sample and for Hispanic workers. These averages are segregation measures commonly used in the sociology literature. The average percentage of Hispanic workers with which Hispanic workers work, denoted H_H, is called the "isolation index," and the average percentage of Hispanic workers with which white workers work, denoted W_H, is called the "exposure index." We focus more on a third measure, the difference between these, or

$$CW = H_H - W_H,$$

as a measure of "co-worker segregation." CW measures the extent to which Hispanics are more likely than are whites to work with other Hispanics. For example, if Hispanics and whites are perfectly segregated, then H_H equals 100, W_H is zero, and CW equals 100.[11]

We first report observed segregation, which is simply the sample mean of the segregation measure across workers. We denote this measure by appending an 'O' superscript to the segregation measures – i.e., CW^O. One important point that is often overlooked in research on segregation, however, is that some segregation occurs even if workers are assigned randomly to establishments, 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 (Carrington and Troske, 1997). Rather than considering all deviations from proportional representation across establishments as an "outcome" or "behavior" to be explained, we subtract from our measured segregation the segregation that would occur by chance if workers were distributed randomly across establishments, using Monte Carlo simulations to generate measures of randomly occurring segregation. We denote this random segregation CW^R, and then focus on the difference $\{CW^O - CW^R\}$, which measures segregation above and beyond that which occurs

[11] 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, and $CW' = W_W - H_W$.

randomly.[12] Although theoretically one can have $CW^O < CW^R$ (that is, there is *less* segregation than would be generated randomly), or $CW^O > CW^R$, only the latter one occurs in practice in our data. Again following Carrington and Troske, we scale this difference by the maximum segregation that can occur, or $\{100 - CW^R\}$, which we refer to as "effective segregation." Thus, the effective segregation measure is:

$$[\{CW^O - CW^R\}/\{100 - CW^R\}]\cdot 100,$$

which measures the share of the maximum possible segregation that is actually observed.

There are two reasons that we exclude the worker's own ethnicity when computing the fraction of Hispanics with which he or she works, First, this ensures that in large samples of workers if workers are randomly allocated across establishments, H_H and W_H both equal the share Hispanic in the population. That is, in the case of random allocation we expect to have CW^R equal to 0. This is a natural scaling to use, and stands in contrast to what happens when the worker is included in the calculations, where CW^R will exceed 0 because Hispanic workers are treated as working with "themselves." Second, and perhaps more important, when the own worker is excluded our segregation measures are invariant to the sizes of establishments studied. To see this in a couple of simple examples, first consider a simple case of an economy with equal numbers of Hispanics and whites all working in two-person establishments. Establishments can therefore be represented as HH (for two Hispanic workers), HW, or WW. With random allocation, 1/4 of establishments are HH, 1/2 are WH, and 1/4 are WW. Thus, excluding the own worker, $H_H = (1/2)\cdot 1 + (1/2)\cdot 0 = 1/2$, $W_H = (1/2)\cdot 1 + (1/2)\cdot 0 = 1/2$, and $CW = 0$. If we count the individual, then $H_H = (1/2)\cdot 1 + (1/2)\cdot(1/2) = 3/4$, $W_H = (1/2)\cdot(1/2) + (1/2)\cdot 0 = 1/4$, and $CW = 1/2$. With three-worker establishments and random allocation, 1/8 of establishments are HHH (employing 1/4 of Hispanic workers), 1/8 are WWW (employing 1/4 of white workers), 3/8 are HWW (employing 1/4 of Hispanic and 1/2 of white workers), and 3/8 are HHW (employing 1/2 of Hispanic and 1/4 of white workers). Going through the same type of calculation as above, if we include the worker, then H_H =

[12] This distinction between comparing measured segregation to a no-segregation ideal or 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). Of course to build up CW^R we also compute the isolation and exposure indexes that would be generated in the case of random allocation of workers, and report these as well.

$(1/4)\cdot 1 + (1/4)\cdot(1/3) + (1/2)\cdot(2/3) = 2/3$, $W_H = (1/4)\cdot 0 + (1/4)\cdot(2/3) + (1/2)\cdot(1/3) = 1/3$ and CW = 1/3, whereas if we exclude the worker we again get $H_H = 1/2$, $W_H = 1/2$, and CW = 0.

Although we just argued that in the case of random allocation Hispanics and whites should work with equal percentages of Hispanic co-workers on average (so that CW^R is zero), this result may not hold in parts of our analysis for two reasons. First, this is a large-sample result, and although the baseline sample size in our data set is large, the actual samples that we use to calculate some of our segregation measures are not always large, or at least not necessarily large enough to generate this asymptotic result. Second, some of our segregation measures are calculated conditional on geography (in particular, MSA/PMSA of residence), for reasons explained below. When we condition on geography, we calculate the extent of segregation that would be expected if workers were randomly allocated across establishments within a geographic area. If Hispanics and whites are not evenly distributed across geographic borders, random allocation of workers within geography still will yield the result that Hispanics are more likely to have Hispanic co-workers than are white workers, because for example, more Hispanics will come from the areas where both whites and Hispanics work with a high share of Hispanic workers. For that reason, in all cases, in order to determine how much segregation would occur randomly, we conduct Monte Carlo simulations of the extent of segregation that would occur with random allocation of workers.

There are, of course, other possible segregation measures, such as the traditional Duncan index (Duncan and Duncan, 1955) or the Gini coefficient. We prefer the co-worker segregation measure (CW) to these other measures for two reasons. First, the Duncan and Gini measures are scale invariant, meaning that they are insensitive to the proportions of each group in the workforce. For example, if the number of Hispanics doubles, but they are allocated to establishments in the same proportion as the original distribution, the Duncan and Gini indexes are unchanged. However, except in establishments that are perfectly segregated, the doubling of Hispanics leads each Hispanic worker in the sample to work with a larger percentage of Hispanic co-workers, and also each white worker to work with more Hispanics. In general, this implies that both the isolation and exposure indexes (H_H and W_H, respectively), will increase. But the isolation index will increase by more, since establishments with more Hispanics to begin with will have larger increases in the number of Hispanic workers, and hence CW will increase.[13] In our view, this kind of increase in the number of Hispanic workers *should* be characterized as an increase in segregation. Second, these alternative segregation measures have the same issue outlined above with respect to sensitivity to the number of matched workers in an establishment, and because they are measures that are calculated at only the establishment-level – unlike the co-worker segregation measure we use – there is no conceptual parallel to excluding the own worker from the calculation.[14]

We present some "unconditional" nationwide segregation measures, as well as "conditional" measures that first condition on metropolitan area (MSA/PMSA) of residence. In the first, the simulations randomly assign workers to establishments anywhere in the country; not surprisingly, in these simulations the random segregation measures are zero or virtually indistinguishable from zero. For comparability, when we construct these unconditional segregation measures we use only the workers included in the MSA/PMSA sample used for the conditional analysis.[15] The unconditional estimates provide the simplest measures of the extent of integration by skill, race, ethnicity, or sex, in the

[13] To see a simple example, suppose we take the economy with two-worker establishments and begin with random allocation, so we have one HH establishment, two HW establishments, and one WW establishment, with H_H and W_H equal to 1/2 and CW = 0. Doubling the number of Hispanics and allocating them proportionally yields one HHHH establishment, two HHW establishments, and still one WW establishment. Clearly Hispanics are more isolated, because those working in mixed establishments now work with a higher share Hispanic; H_H rises to 3/4. In this simple case, whites are not more exposed because excluding the own worker, 1/2 of whites still work in establishments with the share Hispanic equal to 100. Thus, CW rises to 1/4.

More generally, W_H will also increase, but not by as much as H_H, and CW will therefore rise. For perhaps the simplest such case, start with four establishments as follows: one HHH, one HHW, one HWW, and one WWW. In this case H_H = 2/3, W_H = 1/3, and CW = 1/3. Doubling the number of Hispanics and allocating them proportionally, we get the following four establishments: HHHHHH, HHHHHW, WWHH, and WWW: In this case H_H rises to 29/36 (increasing by 5/36), W_H rises to 14/36 (increasing by 2/36), and CW rises to 15/36 (increasing by 3/36).

[14] We believe this explains why, in Carrington and Troske (1998a, Table 3), where there are small samples of workers within establishments, the estimated (We need to check—is this for random allocation too?) Gini indexes are often extremely high..

workplace. However, they reflect the distribution of workers both across cities as well as across establishments within cities. As such, the unconditional measures may tell us less about forces operating in the labor market to create segregation, whereas the conditional measures – which can be interpreted as taking residential segregation by city as given – may tell us more about this. Because we use the same samples for the conditional and unconditional analyses, for these analyses the observed segregation measures are identical. Only the simulations differ, but these differences of course imply differences in the effective segregation measures.

For the Monte Carlo simulations that generate measures of random segregation, we first define the unit within which we are considering workers to be randomly allocated. We use U.S. Census Bureau MSA/PMSA designations, because these are defined to some extent based on areas within which substantial commuting to work occurs.[16] We then calculate for each metropolitan area 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. We do this simulation 100 times, and compute the random segregation measures as the means over these 100 simulations. Not surprisingly, the random segregation measures are very precise; in all cases the standard deviations were trivially small.

IV. Changes in Segregation

[15] The results in this paper are generally robust to measuring segregation at the level of the MSA/CMSA metropolitan area (rather than the MSA/PMSA level), as well as measuring unconditional segregation by including all workers in the United States whether or not they live or work in a metropolitan area. For the within MSA/CMSA analysis, results are very similar to the within MSA/PMSA analysis, with the only difference that the increase in black-white segregation is a about one-quarter smaller in the first case. For the national analysis using the full DEED's vs. the MSA/PMSA sample, the changes in segregation are always in the same direction and qualitatively similar, although the estimated percentage changes are a bit more moderate.

[16] See U.S. Census Bureau, http://www.census.gov/geo/lv4help/cengeoglos.html (viewed April 18, 2005). This is not to say that residential segregation at a level below that of MSA's and PMSA's 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 establishments within a metropolitan area in which a worker could be employed.

With the preceding technical material out of the way, the empirical results can be presented quite concisely.

Segregation by Education

The findings for changes in segregation by education are reported in Table 3. We begin by computing segregation between those with at least some college education and those with at most a high school education. The observed segregation measure for 1990 indicates that on average low education workers are in workplaces in which 54.2 percent of their co-workers are low education, while high education workers are in workplaces in which only 34.5 percent are low education, for a difference of 19.7. This is also the effective segregation measure for the national sample because random allocation of workers to establishments anywhere in the country leads to a random coworker segregation measure of zero. When we look within MSA's/PMSA's, randomness generates a fairly small amount of segregation, so the effective segregation measure declines only a little, to 17.3.

In the 2000 data, observed segregation is 1.4 percentage points higher (21.1), while random segregation is a shade lower. In combination, then, looking within MSA's/PMSA's, effective segregation by education rises to 19.2, or by 11.2 percent, from 1990 to 2000. In the national data, the increase is smaller, from 19.7 to 21.1 percent, or 7.1 percent.[17] One of the mechanisms for this increase in segregation by education is the decline over the decade in the fraction of workers in the sample with a at most a high school degree, from 42.9 percent in 1990 to 35.8 percent in 2000. This suggests that we should be cautious in inferring that the increase in segregation by education is attributable to increased returns to skill as in the Kremer and Maskin model (1996), as it is also possible that segregation by unobserved skill is unchanged but more workers with high unobserved skills have higher education in the 2000 data.

The next two panels of Table 3 report results for two alternative, education cutoffs: high school dropouts vs. at least a high school degree; and less than a bachelor's degree vs. at least a bachelor's degree. In both cases, the increase in segregation from 1990 to 2000 is somewhat larger. For the high

[17] We remind that reader that when we say "national," we refer to the MSA/PMSA sample.

school dropout vs. at least a high school degree breakdown, the overall national figures indicate that educational segregation increased by 11.7 percent, and by 14 percent within MSA's/PMSA's. When we instead classify workers by whether or not they have a bachelor's degree, the increases in segregation are from 15.7 to 16.4 percent.

These figures strike us as fairly notable increases in segregation by education. The direction of change is consistent with the conjecture of Kremer and Maskin (1996), and it is possible that the previous decade might have experienced a greater increase in segregation by education, given the sharper increase in schooling-related earnings differentials in that period, although the workforce adjustments may occur relatively slowly.

Segregation by Race

Evidence on changes in segregation by race are reported in Table 4. In 1990, the observed segregation measures indicate that blacks on average worked with workforces that were 23.7 percent black, whereas the comparable figure for whites was only 5.8 percent, for an observed segregation measure of 17.8. This rose between 1990 and 2000 to 21.8, driven mainly by a sharp increase in the average share black in workplaces were blacks were employed. Nationally, black-white segregation rises from 17.8 to 21.8, an increase of 22.5 percent. Within MSA's/PMSA's, the increase is slightly smaller, at 20.5 percent. We interpret these magnitudes as indicating a sharp increase in workplace segregation by race from 1990 to 2000.[18]

Hispanic-White Segregation

Next, Table 5 reports results for Hispanic-white segregation.[19] Observed Hispanic-white segregation is pronounced. In 1990, Hispanic workers on average worked in establishments with workforces that were 39.4 percent Hispanic, compared with a 4.5 percent figure for whites. Both of these

[18] In Hellerstein and Neumark (2005), we report bootstrapped standard errors for differences in estimates of effective segregation. Differences considerably smaller than the types of increases we find in this paper were strongly significant.

[19] Using the 1990 data only, Hellerstein and Neumark (2005) go into considerable detail regarding Hispanic-white segregation, finding that differences in English language skills account for about one-third of this segregation.

numbers increased slightly as of 2000, to 40.7 percent and 6 percent, respectively, so that the observed segregation measure remained roughly constant – 34.9 percent in 1990 and 34.7 percent in 2000.

Because of relatively sharp differences in the Hispanic composition of urban areas across the United States, randomness generates a considerable amount of Hispanic-white segregation. This is indicated in the table, where random segregation equals 18.8 in 1990 and 18.0 in 2000. However, again the changes are small, so that the change in effective Hispanic-white segregation appears to be relatively minor. In the national data segregation declines by 0.2 percentage points, or by less than 1 percent. And within urban areas, segregation increases slightly, from 19.8 to 20.4, or by only 2.7 percent. Overall, then, both the small magnitudes and the differences in results across and within urban areas lead us to conclude that little changed with respect to Hispanic-white workplace segregation between 1990 and 2000.

Sex Segregation

Finally, we turn to segregation by sex. A priori, we might expect to find substantial declines in this form of segregation, because of the declining differences in the types of jobs done by men and women (Wells, 1998) , As Table 6 reports, in 1990 women on average worked in establishments with workforces that were 59.9 percent female, as compared with establishments in which men worked, which were 36.2 percent female. Thus, observed segregation was 23.7. As of 2000, the increase in the share female with which men work increased relatively sharply, from 36.2 to 40.2, and as a result observed segregation fell to 20.4. Random segregation by sex is relatively trivial, because neither men nor women constitute a very small share of the workforce. As a result, the change in effective segregation is close to the change in observed segregation. In particular, effective segregation by sex declined from 23.7 to 20.4, or 13.9 percent, on a national basis. And the same 13.9 percent decline is estimated within urban areas because, of course, the distributions of men and women across cities are similar. We view this evidence on changes in sex segregation as suggesting a substantive decline over the decade.

A natural question is the extent to which this overall decline in sex segregation reflects declining differences in the occupational distribution of men and women, as compared with reductions in segregation across workplaces even for men and women in the same occupation. To address this question, following the methods in Hellerstein and Neumark (2005), we construct "conditional" random segregation measures, where we simulate segregation holding the distribution of workers by occupation fixed across workplaces. So, for example, if an establishment in our sample is observed to have three workers in occupation A, then three workers in occupation A will be randomly allocated to that establishment. As before, we compute the average (across the simulations) simulated fraction of co-workers who are female for females, denoting this F_F^C, and the average (across the simulations) simulated fraction of co-workers who are female for males, denoting this M_F^C. The difference between these two is denoted CW^C, and we define the extent of "effective conditional segregation" to be:

$$[\{CW^0 - CW^C\}/\{100 - CW^R\}]\times 100,$$

where CW^R is the measure of random segregation obtained when not conditioning on occupation. A conditional effective segregation measure of zero would imply that all of the effective segregation between women and men can be attributed to differences in the occupations employed by various establishments ("occupational segregation"), coupled with differences in the occupational distribution between women and men. Conversely, a conditional effective segregation measure equal to that of the (unconditional) effective segregation measure would imply that none of the effective segregation between women and men can be attributed to occupational segregation across workplaces. Columns (5) and (6) of Table 6 report the results of doing this calculation based on a classification of six 1-digit Census occupations.[20] We do this only for the within MSA/PMSA sample, because central to this analysis is the ability to randomly distribute workers to different establishments, and it only makes sense to do this within the urban areas in which workers commute. The estimates for 1990, in column (5), indicate that about one-sixth (16.4 percent) of the effective segregation of women from men is attributable to differences in the occupational distribution; conditional on occupation, effective segregation by sex

[20] The occupations are: managerial and professional specialty; technical, sales, and administrative support; service; farming, forestry, and fishery; precision production, craft, and repair; and operators, fabricators, and laborers.

falls from 23.3 (column (2)) to 19.5. In the 2000 data, reported in column (6), the effect of occupation is a little bit more pronounced, accounting for 19.5 percent of effective segregation.

This is surprising, if we think the major change from convergence in labor market outcomes for men and women is greater similarity of occupations, which might have led us to believe that occupational segregation would play a smaller role in workplace segregation. Reflecting this, between 1990 and 2000 the estimated reduction in sex segregation is larger when we condition on occupation, falling by 17.1 percent, versus 13.9 percent unconditionally. As shown in the last two columns of Table 6, the result regarding the contribution of occupational segregation to changes in segregation by sex is the same if we use 3-digit Census occupations. Although in each year occupational segregation accounts for much more of the amount of sex segregation (around 62 percent versus 18 percent for 1-digit occupations), it is still the case that the decline in segregation (24.3 percent) is larger conditional on occupation than is the case for conditioning on 1-digit occupation (17.1 percent) or unconditionally (13.9 percent).

These results imply that occupational segregation acted to hold back declines in workplace segregation in this decade, if anything, as workplaces would have become less segregated by sex had men and women held the same occupational distribution in each year. Note that this does not mean that sex segregation by occupation went up from 1990 to 2000 – just that because of the joint distribution of men and women across occupations and establishments, sex segregation by occupation contributed more to workplace segregation in 2000.

V. Conclusions

We present evidence on changes in workplace segregation by education, race, ethnicity, and sex. For this analysis, we use the newly-constructed 2000 Decennial Employer-Employee Dataset (DEED). The 2000 DEED, like the 1990 DEED, provides new opportunities to study workplace segregation at the establishment level. More significantly, by pairing the two we are able to present what we believe are the first estimates of changes in workplace segregation based on 2000 Census data. These estimates provide evidence that is complementary to that on changes in residential segregation in the decade between the Censuses. Moreover, we believe that evidence on workplace segregation and how it has changed is likely to be more informative about social interactions between groups (with reference to race, ethnicity, and sex), and directly informative about hypotheses regarding changes in workplace segregation by skill.

The evidence indicates that racial and ethnic segregation at the workplace level remains quite pervasive. For example, if we compare black and white workers, the difference in the share black among the workforce at the establishments where they work is around 22 percentage points. If we compare Hispanics and whites, the difference is about 50 percent larger. At the same time, there is fairly substantial segregation by skill, as measured by education. In other work (Hellerstein and Neumark (2005), using only the 1990 DEED) we explore the extent to which racial and ethnic segregation are attributable to skill differences between blacks and whites or Hispanics and whites; in the latter case we focus on language skills. Only for the latter is there evidence that skill differences play a substantial role, explaining about one-third of Hispanic-white segregation.

More significantly, putting together the 1990 and 2000 data, we find *no* evidence of declines in workplace segregation by race and ethnicity. Hispanic-white segregation was largely unchanged, while black-white segregation increased quite strongly – depending on which estimate we use, by about 20 to 22 percent. Over this decade, segregation by education also increased, by about 7 to 16 percent, which is consistent with conjectures that rising returns to skill might generate more segregation by skill, although it could also be attributable to rising education levels among workers with more unobserved skills, with a given pattern of segregation based on these skills.[21]

To the extent that declines in segregation are positive developments, the one bright spot is the decline in workplace segregation by sex, which fell about 14 percent from 1990 to 2000. One might think that this reflects declining differences in the occupational distribution of men and women, but it turns out that changes in occupational segregation did not contribute to declining sex segregation across workplaces, and that reductions in workplace segregation conditional on occupational segregation were if anything larger.

[21] Coupled with the earlier findings suggesting that black-white segregation is largely unrelated to education differences, this likely has little if anything to do with the increase in workplace segregation by race.

References

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2005. “Trends in U.S. Wage Inequality: Re-Assessing the Revisionists.” NBER Working Paper No. 11627.

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 2003. “New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data.” *Journal of Labor Economics*, Vol. 21, No. 4, October, pp. 887-922.

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.

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.

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 H., and Kenneth R. Troske. 1998a. “Interfirm Racial Segregation and the Black/White Wage Gap.” *Journal of Labor Economics*, Vol. 16, No. 2, April, pp. 231-60

Carrington, William J. And Kenneth Troske. 1998b. “Sex Segregation in U.S. Manufacturing.” *Industrial and Labor Relations Review*, Vol. 51, April, pp. 445-464.

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.

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

Echenique, Frederico, and Roland Fryer. 2005. “On the Measurement of Segregation.” NBER Working Paper No. 11258.

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

Glaeser, Edward L., and Jacob L. Vigdor. 2001. “Racial Segregation in the 2000 Census: Promising News.” The Brookings Institution Survey Series, April.

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.

Hellerstein, Judith, and David Neumark. In progress. “Space vs. Race: New Evidence on Spatial Mismatch and the Racial Composition of Jobs.”

Hellerstein, Judith, and David Neumark. 2005. "Workplace Segregation in the United States: Race, Ethnicity, and Skill." NBER Working Paper No. 11599.

Hellerstein, Judith, and David Neumark. 2004. "Production Function and Wage Equation Estimation with Heterogeneous Labor: Evidence from a New Matched Employer-Employee Data Set." NBER Working Paper No. 10325.

Hellerstein, Judith, and David Neumark. 2003. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." *Annales d'Economie et de Statistique*, Vol. 71-72, July-December, pp. 19-78.

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

Iceland, John, and Daniel H. Weinberg. 2002. "Racial and Ethnic Segregation in the United States: 1980-2000." U.S. Census Bureau, Census 2000 Special Reports. Available at http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf (viewed October 3, 2005).

James, Daniel R., and Karl E. Taeuber. 1985. "Measures of Segregation." In Sociological Methodology, ed. 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, December, pp. 802-25.

McConville, Shannon, and Paul Ong. 2001. "Examining Residential Segregation Patterns." Discussion Paper, The Ralph & Goldy Lewis Center for Regional Policy Studies, UCLA.

U.S. Census Bureau. "Census Geographic Glossary." http://www.census.gov/geo/lv4help/cengeoglos.html (viewed April 2005 2003).

U.S. Census Bureau, "Census Tracts and Block Numbering Areas." http://www.census.gov/geo/www/GARM/Ch10GARM.pdf (viewed April 2005).

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.

Wells, Thomas. 1998. "Change in Occupational Sex Segregation During the 1980s and 1990s." Center for Demography and Ecology Working Paper No. 98-14, University of Wisconsin, Madison.

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

Table 1: Means for Workers

	1990 SEDF	1990 Full DEED	1990 Restricted DEED	2000 SEDF	2000 Full DEED	2000 Restricted DEED
	(1)	(2)	(3)	(4)	(5)	(6)
Age	37.08	37.51	37.53	39.15	39.57	39.53
	(12.78)	(12.23)	(12.13)	(13.03)	(12.51)	(12.33)
Female	0.46	0.47	0.47	0.46	0.50	0.51
Married	0.60	0.65	0.63	0.58	0.62	0.60
White	0.82	0.86	0.84	0.78	0.83	0.79
Hispanic	0.07	0.05	0.06	0.09	0.07	0.08
Black	0.08	0.05	0.06	0.09	0.06	0.08
Fulltime	0.77	0.83	0.84	0.78	0.82	0.83
Number of Kids (if female)	0.75	0.73	0.69	0.78	0.76	0.74
	(1.04)	(1.01)	(0.99)	(1.07)	(1.04)	(1.03)
High School Diploma	0.34	0.33	0.30	0.31	0.29	0.25
Some College	0.30	0.32	0.33	0.33	0.35	0.35
BA	0.13	0.16	0.18	0.15	0.18	0.20
Advanced Degree	0.05	0.05	0.06	0.06	0.08	0.09
Ln(Hourly Wage)	2.21	2.30	2.37	2.55	2.63	2.70
	(0.70)	(0.65)	(0.65)	(0.73)	(0.70)	(0.70)
Hourly Wage	12.10	12.89	13.68	17.91	18.83	20.19
	(82.19)	(37.07)	(27.41)	(137.20)	(63.61)	(64.05)
Hours Worked in previous year	39.51	40.42	40.55	40.22	40.72	40.90
	(11.44)	(10.37)	(10.10)	(11.74)	(11.09)	(10.85)
Weeks Worked in previous year	46.67	48.21	48.46	47.23	48.38	48.56
	(11.05)	(9.34)	(9.05)	(10.58)	(9.27)	(9.05)
Earnings in previous year	22,575	25,581	27,478	33,521	37,244	40,272
	(26,760)	(29,475)	(30,887)	(42,977)	(47,237)	(50,406)
Industry:						
Mining	0.01	0.01	0.01	0.01	0.00	0.00
Construction	0.07	0.04	0.03	0.08	0.05	0.04
Manufacturing	0.25	0.34	0.35	0.21	0.26	0.26
Transportation	0.08	0.05	0.05	0.07	0.05	0.05
Wholesale	0.05	0.07	0.08	0.05	0.05	0.05
Retail	0.20	0.17	0.15	0.21	0.21	0.20
FIRE	0.08	0.08	0.09	0.07	0.07	0.07
Services	0.26	0.24	0.24	0.31	0.31	0.32
N	12,143,183	3,291,213	1,828,020	14,057,121	4,089,098	2,209,908

Table 2: Means for Establishments

	1990 BR	1990 Full DEED	1990 Restricted DEED	2000 BR	2000 Full DEED	2000 Restricted DEED
	(1)	(2)	(3)	(4)	(5)	(6)
Total Employment	17.57	52.68	104.67	18.77	48.74	95.54
	(253.75)	(577.39)	(996.52)	(138.11)	(232.05)	(371.18)
Establishment Size:						
1-25	0.88	0.65	0.39	0.87	0.66	0.41
26-50	0.06	0.15	0.22	0.06	0.15	0.21
51-100	0.03	0.10	0.19	0.03	0.09	0.17
101+	0.03	0.10	0.21	0.03	0.09	0.20
Industry:						
Mining	0.00	0.01	0.01	0.00	0.00	0.00
Construction	0.09	0.07	0.06	0.11	0.08	0.07
Manufacturing	0.06	0.13	0.23	0.06	0.13	0.18
Transportation	0.04	0.05	0.05	0.04	0.05	0.05
Wholesale	0.08	0.11	0.10	0.07	0.07	0.07
Retail	0.25	0.24	0.23	0.25	0.29	0.27
FIRE	0.09	0.10	0.11	0.09	0.08	0.07
Services	0.28	0.26	0.21	0.35	0.30	0.27
In MSA	0.81	0.82	1	0.81	0.79	1
Census Region:						
North East	0.06	0.06	0.05	0.06	0.05	0.04
Mid Atlantic	0.16	0.15	0.16	0.15	0.14	0.14
East North Central	0.16	0.20	0.21	0.16	0.20	0.21
West North Central	0.07	0.08	0.07	0.08	0.09	0.08
South Atlantic	0.18	0.16	0.15	0.18	0.16	0.16
East South Central	0.05	0.05	0.04	0.06	0.05	0.04
West South Central	0.10	0.10	0.09	0.10	0.10	0.10
Mountain	0.06	0.05	0.05	0.07	0.06	0.06
Pacific	0.16	0.15	0.17	0.16	0.15	0.17
Payroll ($1000)	397	1,358	2,910	694.44	1,993	4,421
	(5,064)	(10,329)	(16,601)	(69,383)	(115,076)	(198,414)
Payroll/Total Employment	21.02	24.24	26.70	33.74	35.91	42.27
	(1,385.12)	(111.79)	(181.48)	(70,72.29)	(1,834.40)	(1,877.29)
Share of Employees Matched	--	0.17	0.16	--	0.16	0.14
Multi-Unit Establishment	0.23	0.42	0.53	0.26	0.40	0.50
N	5,237,592	972,436	317,112	5,651,680	1,279,999	411,300

Table 3: Segregation by Education

	1990 U.S. MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample	2000 Within MSA/PMSA Sample
	% Low Ed	% Low Ed	% Low Ed	% Low Ed
	(1)	(2)	(3)	(4)
Co-Worker Segregation				
High school degree or less vs. more than high school				
Observed Segregation				
Low Education Workers	54.2	54.2	49.3	49.3
High Education Workers	34.5	34.5	28.2	28.2
Difference	19.7	19.7	21.1	21.1
Random Segregation				
Low Education Workers	42.9	44.6	35.8	37.3
High Education Workers	42.9	41.7	35.8	35.0
Difference	0	2.9	0	2.3
Effective Segregation	19.70	17.30	21.1	19.2
Percent Change, 1990-2000			**7.1**	**11.2**
Less than high school vs. high school degree or more				
Observed Segregation				
Low Education Workers	26.0	26.0	25.5	25.5
High Education Workers	10.8	10.8	8.6	8.6
Difference	15.2	15.2	16.9	16.9
Random Segregation				
Low Education Workers	12.7	13.9	10.3	11.3
High Education Workers	12.7	12.6	10.4	10.3
Difference	0	1.3	-0.1	1.0
Effective Segregation	15.20	14.08	17.0	16.1
Percent Change, 1990-2000			**11.7**	**14.0**
Less than bachelor's degree vs. bachelors' degree or more				
Observed Segregation				
Low Education Workers	80.7	80.7	77.7	77.7
High Education Workers	60.6	60.6	54.3	54.3
Difference	20.1	20.1	23.4	23.4
Random Segregation				
Low Education Workers	75.9	76.6	70.8	72.0
High Education Workers	75.9	73.5	70.8	68.1
Difference	0	3.1	0	3.9
Effective Segregation	20.10	17.54	23.4	20.3
Percent Change, 1990-2000			**16.4**	**15.7**
Number of Workers	1,828,020	1,828,020	2,209,908	2,209,908
Number of Establishments	317,112	317,112	411,300	411,300
"Pseudo" Median Workers Matched	9		8	
"Pseudo" Median Share of Workforce Matched	8.2		8.1	

Table 4: Black-White Segregation

	1990 U.S. MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample	2000 Within MSA/PMSA Sample
	% Black	% Black	% Black	% Black
	(1)	(2)	(3)	(4)
Co-Worker Segregation				
Observed Segregation				
Black Workers	23.7	23.7	28.7	28.7
White Workers	5.8	5.8	6.9	6.9
Difference	17.8	17.8	21.8	21.8
Random Segregation				
Black Workers	7.1	11.2	8.8	14.2
White Workers	7.1	6.8	8.8	8.3
Difference	0	4.4	0	5.9
Effective Segregation	17.8	14.02	21.8	16.90
Percent Change, 1990-2000			**22.5**	**20.5**
Number of Workers	1,618,876	1,618,876	1,893,034	1,893,034
Number of Establishments	285,988	285,988	360,072	360,072
“Pseudo” Median Workers Matched	9		8	
“Pseudo” Median Share of Workforce Matched	7.9		7.7	

Table 5: Hispanic-White Segregation

	1990 U.S. MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample	2000 Within MSA/PMSA Sample
	% Hispanic	% Hispanic	% Hispanic	% Hispanic
	(1)	(2)	(3)	(4)
Co-Worker Segregation				
Observed Segregation				
Hispanic Workers	39.4	39.4	40.7	40.7
White Workers	4.5	4.5	6	6
Difference	34.9	34.9	34.7	34.7
Random Segregation				
Hispanic Workers	6.9	24.4	9.2	25.5
White Workers	6.9	5.6	9.2	7.5
Difference	0	18.8	0	18.0
Effective Segregation	34.9	19.83	34.7	20.37
Percent Change, 1990-2000			**-0.6**	**2.7**
Number of Workers	1,625,953	1,625,953	1,906,878	1,906,878
Number of Establishments	293,989	293,989	373,006	373,006
“Pseudo” Median Workers Matched	8		7	
“Pseudo” Median Share of Workforce Matched	7.8		7.6	

Table 6: Segregation by Sex

	Unconditional				Conditional on 1-Digit Occupation		Conditional on 3-Digit Occupation	
	1990 U.S. MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample	2000 Within MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample	1990 Within MSA/PMSA Sample	2000 U.S. MSA/PMSA Sample
	% Female	% Female	% Female	% Female	% Female	% Female	% Female	% Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Co-Worker Segregation								
Observed Segregation								
Female Workers	59.9	59.9	60.6	60.6	59.9	60.6	59.9	60.6
Male Workers	36.2	36.2	40.2	40.2	36.2	40.2	36.2	40.2
Difference	23.7	23.7	20.4	20.4	23.7	20.4	23.7	20.4
Random Segregation								
Female Workers	47.4	47.7	50.5	50.7	49.7	52.6	54.9	57.1
Male Workers	47.4	47.2	50.5	50.3	45.4	48.3	40.7	43.9
Difference	0	0.5	0	0.4	4.3	4.3	14.2	13.2
Effective Segregation	23.70	23.32	20.40	20.08	19.50	16.16	9.55	7.23
Percent Change, 1990-2000			**-13.9**	**-13.9**		**-17.1**		**-24.3**
Fraction of sex segregation accounted for by occupation					16.38	19.50	59.05	64.00
Number of Workers	1,828,020	1,828,020	2,209,908	2,209,908	1,828,020	2,209,908	1,828,020	2,209,908
Number of Establishments	317,112	317,112	411,300	411,300	317,112	411,300	317,112	411,300
"Pseudo" Median Workers Matched	9		8					
"Pseudo" Median Share of Workforce Matched	8.2		8.1					