Volume Title: The Analysis of Firms and Employees: Quantitative and Qualitative Approaches

Volume Author/Editor: Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 978-0-226-04287-9; 0-226-04287-1

Volume URL: http://www.nber.org/books/bend08-1

Conference Date: September 29-30, 2006

Publication Date: October 2008

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

Chapter Author: Judith Hellerstein, David Neumark, Melissa McInerney

Chapter URL: http://www.nber.org/chapters/c9115

Chapter pages in book: (163 - 195)

5

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

Judith Hellerstein, David Neumark,
and Melissa McInerney

5.1 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 2007), 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, forthcoming).

We just recently completed the construction of the 2000 Beta-DEED

Judith Hellerstein is an associate professor of economics at the University of Maryland, and a research associate of the National Bureau of Economic Research. David Neumark is a professor of economics at the University of California, Irvine, a research fellow of the Institute for the Study of Labor, and a research associate of the National Bureau of Economic Research. Melissa McInerney is a statisician at the U.S. Bureau of the Census, Center for Economic Studies, and a PhD candidate at the University of Maryland, Department of Economics.

This research was funded by National Institute of Child Health & Human Development (NICHD) grant R01HD042806. We also thank the Alfred P. Sloan Foundation for its generous support. We are grateful to Ron Jarmin, Julia Lane, and an anonymous reviewer for helpful comments. 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, the U.S. Bureau of the Census, or the Sloan Foundation. 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.

(based on the 2000 Census of Population).[1] In this paper, we use the 1990 and 2000 DEEDs to measure changes in establishment-level workplace segregation over the intervening decade, an analysis for which the DEEDs 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 and Ong 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 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, and so on. 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, and so on (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

1. The 2000 Beta-DEED is an internal U.S. Census Bureau data set 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.

2. See the U.S. Census Bureau, http://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 in the following.

1991).[4] There has generally been less attention paid to segregation by education, but in our earlier work (Hellerstein and Neumark, forthcoming), we documented rather extensive segregation by education (as well as language, which we do not consider in the present paper) in the 1990 DEED.

Measuring changes in workplace segregation along these lines is of interest for a number of reasons. 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 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 by race, ethnicity, and sex provides indicators of the success of these efforts. Finally, increases in the productivity (and pay) of more-educated workers relative to less-educated workers may have led to 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, Katz, and Kearney 2005).

We measure changes in segregation using the 1990 and 2000 Decennial Employer-Employee Databases (DEEDs). 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, and so on. Thus, data from the 1990 and 2000 DEEDs provides unparalleled opportunities to study changes in workplace segregation by skill, race, ethnicity, and sex.[6]

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 (1996) 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.

6. Carrington and Troske (1998a, b) use data sets much more limited in scope than the ones 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.

5.2 The 1990 and 2000 DEEDs

The analysis in this paper is based on the 1990 and 2000 DEEDs, 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 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 data sets.

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 List (SSEL); the BR is a database containing information for most business establishments operating in the United States in each year, which is continuously updated (see Jarmin and Miranda 2002).

Households receiving the 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 a list of most 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 Standard Industrial Classification (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 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, Haltiwanger, and Krizan 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 supplemented 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 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

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.

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 err on the side of avoiding false matches. We choose matching algorithms based on substantial experimentation and visual inspection of many thousands of records.

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 5.1, 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 elig-

8. One might also considering trying to impute matches where this strategy fails by matching based on imputed place of work instead of information in the Write-In file. 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.

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 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 (whether or not 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.)

Table 5.1 Means for workers

	1990			2000		
	SEDF (1)	Full DEED (2)	Restricted DEED (3)	SEDF (4)	Full DEED (5)	Restricted DEED (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
Full-time	0.77	0.83	0.84	0.78	0.82	0.83
No. 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

ible 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. 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. In addition, the percent female in the 2000 data is 46 in the SEDF, but 50 in the DEED; this is different than the discrepancy in 1990 where the percent female is 46 in the SEDF and only a slightly higher 47 percent in the DEED.

Part of the explanation for differences in racial and ethnic representation that result from the matching process 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 reflect weaker labor market attachment among minorities, suggesting that the segregation results we obtain might best be interpreted as measuring 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 when manufacturing comprised a larger fraction of workers to begin with in the SEDF. The reasons for this are given in the following 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 5.2 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 De-

Table 5.2 Means for establishments

	1990			2000		
	BR (1)	Full DEED (2)	Restricted DEED (3)	BR (4)	Full DEED (5)	Restricted DEED (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 ($1,000)	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)	(772.29)	(1,834.40)	(1,877.29)
Share of employees matched		0.17	0.16		0.16	0.14
Multiunit 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

cennial Census Long Forms, 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, where, for example, the DEEDs overrepresent workers in manufacturing establishments.) On average, establishments in the BRs have eighteen to nineteen employees, while the average in the DEEDs is forty-nine to fifty-three 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 DEEDs, corresponding to the sample of workers in columns (3) and (6) of table 5.1. 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, however, the DEED samples are far more representative than previous detailed matched data sets for the United States constructed using just the SEDF and the BR (see Hellerstein and Neumark 2003).[10]

Because the DEED captures larger establishments and because our sample restrictions 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 forty-one workers in 1990 and thirty-six in 2000, whereas the first quartile of the employment-weighted size distribution of all establishments in the BR for each year is nineteen in 1990 and twenty-one in 2001.[11] 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; Equal Employment Opportunity Commission (EEOC) laws cover employers with fifteen or more workers, and affirmative action rules for federal contractors cover employers with fifty or more workers. Because policy has been directed at larger establishments, examining the extent of and changes in workplace segregation in larger establishments is important.

5.3 Methods

We focus our analysis on a measure of segregation that is based on the percentages of workers in an individual's establishment, or workplace, in different demographic groups. Consider for clarity measuring segregation between white and Hispanic workers. For each white or Hispanic worker in our sample, we compute the percentage of Hispanic workers with which that worker works, excluding the worker him- or herself. Because we exclude

10. These earlier matched data sets—the Worker-Establishment Characteristics Database (WECD), which covers manufacturing only, and the New Worker-Establishment Characteristics Database (NWECD), which covers all industries—were smaller and less representative because the matching algorithm used could only be applied to establishments that were unique in a cell defined by detailed geographic information and industry classification. Thus, for example, manufacturing establishments were much more likely to occupy their own industry-location cell than were retail establishments.

11. 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.

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 coworkers in Hispanic workers' establishments who are Hispanic, denoted H_H, is called the "isolation index," and the average percentage of coworkers in white workers' establishments who are Hispanic, denoted W_H, is called the "exposure index." We focus more on a third measure, the difference between these, or

$$\mathrm{CW} = H_H - W_H,$$

as a measure of "coworker segregation." The variable 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.[12]

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 coworker segregation measure—that is, 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—that is, 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 similarly for the isolation and exposure indexes) and then focus on the difference ($\mathrm{CW}^O - \mathrm{CW}^R$), which measures segregation above and beyond that which occurs randomly.[13] Although theoretically one can have $\mathrm{CW}^O < \mathrm{CW}^R$ (that is, there is *less* segregation than would be generated randomly) or $\mathrm{CW}^O > \mathrm{CW}^R$, only the latter occurs in practice in our data. Again following Carrington and Troske, we scale this difference by the maximum

12. 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 $\mathrm{CW}' = W_W - H_W$.

13. 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, Falk, and Cohen 1976; Winship 1977; Boisso et al. 1994; and Carrington and Troske 1997). Of course, to build CW^R we also compute the isolation and exposure indexes that would be generated in the case of random allocation of workers, and we report these as well.

segregation that can occur, or (100 – CW^R), we refer to this measure as "effective segregation." Thus, the effective segregation measure is

$$\left(\frac{CW^O - CW^R}{100 - CW^R}\right) \times 100,$$

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, $H_H{}^R = (1/2) \cdot 1 + (1/2) \cdot 0 = 1/2$, $W_H{}^R = (1/2) \cdot 1 + (1/2) \cdot 0 = 1/2$, and $CW^R = 0$.[14] If we count the individual, then $H_H{}^R = (1/2) \cdot 1 + (1/2) \cdot (1/2) = 3/4$, $W_H{}^R = (1/2) \cdot (1/2) + (1/2) \cdot 0 = 1/4$, and $CW^R = 1/2$. 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 in the preceding, if we include the worker, then $H_H{}^R = (1/4) \cdot 1 + (1/4) \cdot (1/3) + (1/2) \cdot (2/3) = 2/3$, $W_H{}^R = (1/4) \cdot 0 + (1/4) \cdot (2/3) + (1/2) \cdot (1/3) = 1/3$ and $CW^R = 1/3$, whereas if we exclude the worker we again get $H_H{}^R = 1/2$, $W_H{}^R = 1/2$, and $CW^R = 0$.

Although we just argued that in the case of random allocation Hispanics and whites should work with equal percentages of Hispanic coworkers 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 samples that we use to calculate some of our segregation measures are not necessarily large enough to generate this asymptotic result. Second, some of our segregation

14. For the first calculation, for example, 1/2 of hispanic workers are in HH establishments, for which the share hispanic is 1, and 1/2 are in WH establishments, for which the share Hispanic (excluding the worker) is 0.

measures are calculated conditional on geography (in particular, MSA/PMSA of residence), for reasons explained in the following. 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 geographical areas still will yield the result that Hispanics are more likely to have Hispanic coworkers than are white workers because, for example, more Hispanics will come from areas where both whites and Hispanics work with a high share of Hispanic workers. For these reasons, in order to determine how much segregation would occur randomly, in all cases 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 coworker 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 coworkers 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 because establishments with more Hispanics to begin with will have larger increases in the number of Hispanic workers, and, hence, CW will increase.[15] 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 are also sensitive to the number of matched workers in an establishment (the same issue outlined in the preceding), and because they are measures that are calculated at only the establishment level—unlike the coworker segregation measure we use—there is no conceptual parallel to excluding the own worker from the calculation.[16]

15. 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, HHHHW, 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).

16. We believe this explains why, in Carrington and Troske (1998a, table 3), where there are small samples of workers within establishments, the random Gini indexes are often extremely high.

At the same time, because calculated changes in segregation between 1990 and 2000 based on our coworker segregation index are sensitive to the overall proportions of each group in the workforce, changes over the decade in the proportions of particular demographic groups that are matched to establishments can generate changes in measured segregation. So, for example, the fact that the fraction of workers who are Hispanic grew from 1990 to 2000 should yield a small increase in measured coworker segregation by ethnicity over the decade (even if Hispanics and whites are distributed across establishments in the same proportion in each year). We could avoid this problem by using scale-invariant segregation measures, but then we would fail to capture changes in segregation due to actual changes in workforce composition. That is, the fact that Hispanics make up a growing fraction of the workforce is an important phenomenon to capture.[17] Nonetheless, although we emphasize the coworker segregation measure throughout, we also report our key results based on the Duncan index to see how robust the conclusions are.

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.[18] The unconditional estimates provide the simplest measures of the extent of integration by skill, race, ethnicity, or sex in the workplace. However, they reflect the distribution of workers both across cities and 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 these forces. 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.

17. Some measured changes in the sample composition of workers over time may reflect changes in the match rates of various kinds of workers to establishments rather than a change in the underlying population composition. This is obviously a limitation of matched data sets like ours, one that exists to a much smaller extent in administrative data sets that come closer to capturing fully the universe of workers.

18. The results in this paper are generally robust to measuring unconditional segregation by including all workers in the United States whether they live and work in a metropolitan area. For the unconditional analysis using the full DEEDs versus 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 than those reported in the following.

For the Monte Carlo simulations that generate measures of random segregation, we need to first define the unit within which we are considering workers to be randomly allocated. This requires a specification of the relevant labor market. 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.[19] An MSA is a set of one or more counties that contains a population center and the adjacent densely-settled counties, with additional counties included if the share of residents commuting to the population core exceeds a certain threshold.[20] In the case of particularly large MSAs, such as Washington, DC-Baltimore, MD, the entire region meets the criteria to be a MSA, and two or more subsets of the region also meet the MSA definition. In cases such as these, we consider the smaller subsets of counties, called PMSAs. In the Washington, DC-Baltimore, MD example, the larger area (called a Consolidated Metropolitan Statistical Area, or CMSA) is comprised of three PMSAs: Baltimore, MD; Hagerstown, MD; and Washington, DC. Thus, the metropolitan areas on which we focus should be relatively well-defined labor markets, rather than huge areas covering many cities.[21] For example, the 10th percentile of the distribution of MSA/PMSA populations is comprised of smaller metropolitan areas such as Sheboygan, WI, with approximately 100,000 residents, and the 90th percentile is Sacramento, CA, having roughly 1.6 million residents.[22] At the same time, we are certainly not claiming that residential segregation at a level below that of the MSA/PMSA does 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.

Returning to the simulation procedure, we calculate for each MSA/

19. See the U.S. Census Bureau, http://www.census.gov/geo/lv4help/cengeoglos.html (viewed April 18, 2005).

20. See the Geographic Areas Reference Manual, http://www.census.gov/geo/www/GARM/Ch13GARM.pdf (viewed June 12, 2007). There are a handful of MSAs or PMSAs for which the constituent counties change between 1990 and 2000 or an MSA was abolished or created. The following tables report results using the MSAs/PMSAs present in each year. We constructed a restricted sample that for the most part held MSA/PMSA boundaries fixed by using only counties that were in the same MSA/PMSA in each of the two years; the estimated levels of and changes in segregation were almost identical.

21. Nonetheless, 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. The only difference is that the increase in black-white segregation is about one-quarter smaller in the first case than in the estimates reported in the following. In addition, we examined our main results for cities disaggregated by quartiles of the population-weighted size distribution, and there was no systematic relationship between city size and changes in segregation along the dimensions we study.

22. These are calculated from Summary File 1 for the 2000 Decennial Census. The population-weighted totals reflect slightly larger MSA/PMSAs. The population weighted 10th percentile is Galveston, TX, with approximately 250,000 residents, and the 90th percentile is Chicago, IL, with approximately 8.3 million residents.

PMSA 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.

5.4 Changes in Segregation

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

5.4.1 Segregation by Education

The findings for changes in segregation by education are reported in table 5.3. 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 coworkers 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 MSAs/PMSAs, 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 lower. In combination, then, looking within MSAs/PMSAs, effective segregation by education rises two percentage points, or by 11.3 percent, from 1990 to 2000. In the national data, the increase is smaller, from 19.7 to 21.1 percent, or 7.0 percent.[23] The next two panels of table 5.3 report results for two alternative education cutoffs: high school dropouts versus at least a high school degree; and less than a bachelor's degree versus at least a bachelor's degree. For the high school dropouts versus at least a high school degree breakdown, the overall national figures indicate an increase in segregation similar to that seen in the first panel of the table; educational segregation increased by 1.7 percentage points (11.1 percent nationally) and by 1.9 percentage points (13.6 percent) within MSAs/PMSAs. When we instead classify workers by whether

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

Table 5.3 Segregation by education (% low education)

	1990 U.S. MSA/PMSA sample (1)	1990 Within MSA/PMSA sample (2)	2000 U.S. MSA/PMSA sample (3)	2000 Within MSA/PMSA sample (4)
Coworker 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	19.7	17.3	21.1	19.2
Percentage point (percent) change, 1990–2000			1.4 (7.0)	2.0 (11.3)
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	13.8	10.4	11.3
High-education workers	12.7	12.6	10.4	10.3
Difference	0	1.3	0	1.0
Effective segregation	15.2	14.1	16.9	16.0
Percentage point (percent) change, 1990–2000			1.7 (11.1)	1.9 (13.6)
Less than bachelor's degree vs. bachelor's 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	20.2	20.2	23.4	23.4
Random segregation				
Low-education workers	75.9	76.6	70.8	71.9
High-education workers	75.9	73.5	70.8	68.2
Difference	0	3.1	0	3.8
Effective segregation	20.2	17.6	23.4	20.4
Percentage point (percent) change, 1990–2000			3.3 (16.2)	2.8 (16.0)
No. of workers	1,828,020	1,828,020	2,209,908	2,209,908
No. of establishments	317,112	317,112	411,300	411,300

they have a bachelor's degree, the increases in segregation are somewhat larger, between 2.8 and 3.3 percentage points, or 16 to 16.2 percent.[24]

These figures strike us as modest but measurable 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 decade of the 1980s might have experienced even 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. Nonetheless, we may want to be cautious in inferring that the increase in segregation by education is attributable to increased returns to skill. 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 low levels of education—for example, the fraction with at most a high school degree drops from 42.9 percent in 1990 to 35.8 percent in 2000. It is also possible, then, that segregation by skill (rather than measured education) is actually unchanged, but more workers with high unobserved skills have higher education in the 2000 data.

5.4.2 Segregation by Race

Evidence on changes in segregation by race is reported in table 5.4. 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 an increase in the average share black in workplaces where blacks were employed. Nationally, black-white segregation rose 4 percentage points, from 17.8 to 21.8, or an increase of 22.3 percent. Within MSAs/PMSAs, the increase is slightly smaller, at 2.8 percentage points, or 20.3 percent. We interpret these magnitudes as indicating a relatively large increase in workplace segregation by race from 1990 to 2000.

5.4.3 Hispanic-White Segregation

Next, table 5.5 reports results for Hispanic-white segregation.[25] 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 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.

24. In Hellerstein and Neumark (forthcoming), 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.

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

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. Segregation declines in the national

Table 5.4 **Black-White segregation (% Black)**

	1990 U.S. MSA/PMSA sample (1)	1990 Within MSA/PMSA sample (2)	2000 U.S. MSA/PMSA sample (3)	2000 Within MSA/PMSA sample (4)
Coworker 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	14.0	21.8	16.8
Percentage point (percent) change, 1990–2000			4.0 (22.3)	2.8 (20.3)
No. of workers	1,618,876	1,618,876	1,893,034	1,893,034
No. of establishments	285,988	285,988	360,072	360,072

Table 5.5 **Hispanic-White segregation (% Hispanic)**

	1990 U.S. MSA/PMSA sample (1)	1990 Within MSA/PMSA sample (2)	2000 U.S. MSA/PMSA sample (3)	2000 Within MSA/PMSA sample (4)
Coworker 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	19.8	34.7	20.4
Percentage point (percent) change, 1990–2000			–0.2 (–0.4)	0.6 (3.0)
No. of workers	1,625,953	1,625,953	1,906,878	1,906,878
No. of establishments	293,989	293,989	373,006	373,006

data by 0.2 percentage point, or by less than 1 percent. And within urban areas, segregation increases slightly, from 19.8 to 20.4, or by only 3 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.

5.4.4 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 5.6 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 23.6. 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 23.6 to 20.4, or 13.7 percent, on a national basis. And virtually the same decline, 3.2 percentage points or 13.6 percent, is estimated within urban areas because, of course, the distributions of men and women across cities are similar. We view the magnitude of these changes in sex segregation as suggesting a substantive decline over the decade.

One possible explanation for the overall decline in sex segregation is convergence in the occupational distributions of men and women, rather than a reduction in segregation across workplaces even for men and women in the same occupation. To address this possibility, following the methods in Hellerstein and Neumark (forthcoming), 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 coworkers who are female for females, denoting this F_F^C, and the average (across the simulations) simulated fraction of coworkers 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

$$\frac{CW^O - CW^C}{100 - CW^R} \times 100,$$

where CW^R is the measure of random segregation obtained when not conditioning on occupation. A conditional effective segregation measure of

Table 5.6 Segregation by sex (% female)

	Unconditional				Conditional on 3-digit occupation	
	1990 U.S. MSA/PMSA sample (1)	1990 Within MSA/PMSA sample (2)	2000 U.S. MSA/PMSA sample (3)	2000 Within MSA/PMSA sample (4)	1990 Within MSA/PMSA sample (5)	2000 Within MSA/PMSA sample (6)
Coworker segregation						
Observed segregation						
Female workers	59.9	59.9	60.6	60.6	59.9	60.6
Male workers	36.2	36.2	40.2	40.2	36.2	40.2
Difference	23.6	23.6	20.4	20.4	23.6	20.4
Random segregation						
Female workers	47.4	47.7	50.5	50.7	54.4	56.8
Male workers	47.4	47.2	50.5	50.3	41.1	44.1
Difference	0	0.5	0	0.4	13.3	12.6
Effective segregation	23.6	23.3	20.4	20.1	10.4	7.8
Percentage point (percent) change, 1990–2000			–3.2 (–13.7)	–3.2 (–13.6)		–2.6 (–24.8)
Fraction of sex segregation accounted for by occupation					55.4	61.2
No. of workers	1,828,020	1,828,020	2,209,908	2,209,908	1,828,020	2,209,908
No. of establishments	317,112	317,112	411,300	411,300	317,112	411,300

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 distributions of 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 5.6 report the results of doing this calculation based on a consistent occupation classification across 1990 and 2000, as developed in Meyer and Osborne (2005), which is approximately at the three-digit level.[26] 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 makes more sense to do this within the urban areas in which workers commute. The estimates for 1990, in column (5), indicate that a substantial fraction (nearly 50 percent) of the effective segregation of women from men is attributable to differences in the occupational distribution; conditional on occupation, effective segregation by sex falls from 23.3 (column [2]) to 10.4. In the 2000 data, reported in column (6), the effect of occupation is a little bit more pronounced, accounting for 61.2 percent of effective segregation. Finally, conditional on occupation, sex segregation within MSAs/PMSAs declines over time by 2.6 percentage points (from 10.4 to 7.8); in absolute terms, this is similar to the decline in unconditional segregation, but because effective segregation conditional on occupation (in 1990) was only about 45 percent as large as the unconditional effective segregation measure, the decline in conditional segregation between columns (5) and (6) represents a much larger percentage decline—24.8 percent. Altogether, these results suggest that the decline in sex segregation over the decade is not being driven by the increased propensity of women to work in the same occupations as men.

5.5 The Impact of Changing Establishment and Industry Composition

Changes in segregation can arise due to a multitude of factors, some of them compositional, such as the changing occupational distribution of women as discussed in the previous section. In this section, we explore the robustness of our full-sample results to two other types of potentially important compositional changes. First, we explore whether the changes in segregation are due to the changing composition of establishments by recalculating our segregation indexes for only the sample of establishments

26. There are nontrivial differences in occupation codes at the three-digit level between 1990 and 2000. The structure of occupation codes at the one-digit level changed even more dramatically between 1990 and 2000, so we do not attempt a concordance at this higher level of aggregation.

that exist in both the 1990 and 2000 Restricted DEED samples (corresponding to columns [3] and [6] of table 5.1).[27] Ideally, we would like to isolate the separate roles of establishment entry and exit—that is, births of new establishments and deaths of existing ones. However, given that we only match some establishments, we cannot necessarily distinguish births and deaths from matches and nonmatches. But assuming that matching is random with respect to segregation, focusing on the set of establishments that are in both samples is informative about the combined roles of establishment entry and exit.

Second, we explore the robustness of our changes in segregation to changes in the industry mix of employment over the decade by reweighting the segregation indexes for 2000 to reflect the industrial composition of employment at the one-digit level that exists in our 1990 data. This is a little more complicated. First, because we are interested in calculating within-MSA indexes, it is actually the within-MSA industry composition that we need to hold fixed at 1990 levels. As a result, we include in the sample only MSAs that exist in both years. Second, we exclude mining because mining makes up such a trivial proportion of employment that there are some MSAs that have matched workers in mining in 1990 but not in 2000.

To understand how we construct changes in segregation over the decade while holding the distribution of employment across industries within MSAs fixed at 1990 levels, consider again the example of ethnic segregation we discussed in section 5.3. Obviously, we compute H_H (the isolation index) and W_H (the exposure index) for 1990 in the same way we did previously because no adjustment needs to be made when accounting for the 1990 industry composition. In order to compute H_H for 2000 with industry composition fixed as of 1990, we compute the isolation index separately for each industry/MSA pair in 2000.[28] We then take a weighted average across industries of these isolation indexes, where the weight is the product of two components: the fraction of total Hispanic employment (in this example) that works in that industry/MSA pair in 2000, and the ratio of the employment share in the industry/MSA pair in 1990 relative to 2000. The fraction of Hispanic employment serves to aggregate up the industry/MSA-specific isolation indexes to the full-sample isolation index (and, if used alone to weight up the industry/MSA-specific indexes would yield the 2000 unadjusted isolation index), while the ratio of the employment shares adjusts the data appropriately to reflect the composition of employment in 1990 across industries. For the exposure index, W_H, we do the same thing, calculating a

27. By restricting the sample to establishments that exist in the Restricted DEED samples in both years, we drop some very small MSAs from some of the samples we used to calculate segregation indexes in earlier tables, in cases where there are no matched workers for whom to calculate indexes across the two years.

28. For the random segregation indexes, the industry used is the random industry to which the worker is assigned.

separate exposure index for each industry/MSA pair and then weighting by the product of the industry employment share ratio times the fraction of white employment in that pair in 2000. Because the fraction Hispanic in an industry MSA/pair may differ from the fraction white in that same industry/MSA pair, the reweighting may have differential effects on the exposure and isolation indexes. As a consequence, adjusting for industry employment changes over the decade will have the largest impact on measured changes in segregation when there has been differential employment growth in industries with a large share Hispanic coupled with a large difference between the share of Hispanic and the share of white employment in the industry (or if there is a large difference between the isolation and exposure indexes).[29]

The results of these alternative computations are presented in condensed form in table 5.7, where we report only the within-MSA effective segregation measures in each year and the changes over the decade. In the first panel of table 5.7, we report results for coworker segregation by high school degree status. In column (1), we first report the within-MSA effective segregation measure in 1990 of 17.3 (from table 5.3). Following that number, we report the corresponding figure for the sample of establishments that existed both in 1990 and 2000, finding that coworker segregation by high school degree status in 1990 is somewhat lower, at 15.7. The fixed-industry-composition coworker segregation measure for 1990 is 17.3, identical to that for the full sample.[30] In column (2), we report the coworker segregation measures for 2000. For the fixed-establishment sample, coworker segregation by high school degree status is 17.0, 2.2 percentage points lower than for the full sample, and for the results holding industry composition fixed, the coworker measure is slightly higher, at 20.3. Overall, the change over the decade of 2 percentage points for the full sample is close to the 1.4 percentage point increase for the fixed-establishment sample, and the increase holding industry composition fixed is a bit larger, at 3.1 percentage points. In general, though, the observed increase in coworker segregation for the full sample over the decade is robust to the changing mix of establishments and industries.

In the second and third panels of table 5.7, we report the results for the alternative education cutoffs. The results again reflect some small differences across the sample of establishments and mix of industries, and the overall qualitative results again point to increases in segregation by education over the decade.

29. This turns out to be quite significant in our calculations for changes in sex segregation holding the industry composition of employment fixed, where the services industry grew rapidly and is also heavily female.

30. Because we exclude workers in mining and workers in MSAs that were not defined as such in 1990 and 2000, the results for 1990 can be slightly different than we report in the full sample in table 5.3.

Table 5.7 Alternative coworker segregation calculations

	1990 Within MSA/PMSA sample, effective segregation (1)	2000 Within MSA/PMSA sample, effective segregation (2)	Percentage point (percent) change, 1990–2000 (3)
Segregation by education			
High school degree or less vs. more than high school			
Full sample, table 5.3	17.3	19.2	2.0 (11.3)
Establishments present in 1990 and 2000	15.7	17.0	1.4 (8.9)
Fixed industry composition	17.3	20.3	3.1 (17.8)
Less than high school vs. high school degree or more			
Full sample, table 5.3	14.1	16.0	1.9 (13.6)
Establishments present in 1990 and 2000	11.4	12.7	1.2 (10.7)
Fixed industry composition	13.8	15.8	2.0 (14.3)
Less than bachelor's degree vs. bachelor's degree or more			
Full sample, table 5.3	17.6	20.4	2.8 (16.0)
Establishments present in 1990 and 2000	15.4	17.4	2.0 (12.8)
Fixed industry composition	17.6	21.8	4.2 (24.0)
Black-White segregation			
Full sample, table 5.4	14.0	16.8	2.8 (20.3)
Establishments present in 1990 and 2000	11.2	12.6	1.4 (12.7)
Fixed industry composition	14.1	14.7	0.6 (4.6)
Hispanic-White segregation			
Full sample, table 5.5	19.8	20.4	0.6 (3.0)
Establishments present in 1990 and 2000	16.5	15.6	–0.9 (–5.6)
Fixed industry composition	19.1	22.0	2.9 (15.3)
Segregation by Sex			
Unconditional			
Full sample, table 5.6	23.3	20.1	–3.2 (–13.6)
Establishments present in 1990 and 2000	25.2	23.0	–2.3 (–8.9)
Fixed industry composition	23.4	14.4	–9.0 (–38.3)

Note: Mining is excluded for "Full sample" and "fixed industry composition."

Racial segregation increased over the decade for the full sample by 2.8 percentage points (20.3 percent), but increased by only about half that much for the sample of establishments that exist in both years. This means that new establishments in 2000 are characterized by more racial segregation than establishments that existed in 1990. Moreover, holding the industry composition of employment fixed at 1990 levels, racial segregation increased by a much smaller amount over the decade—0.6 percentage points (4.6 percent). The fact that newer establishments and the industries that are gaining in employment over the decade are also more segregated by race in 2000 than older establishments and declining industries could portend continuing increases in racial segregation.

Interestingly, the results are somewhat different for Hispanic-white segregation, as we report in the fifth panel of table 5.7. In the overall sample, coworker segregation increased relatively little over the decade, by only 0.6 percentage points (3 percent). For the sample of establishments that exist in both years, coworker segregation actually fell a little, from 16.5 to 15.6, whereas for the fixed-industry-composition, the coworker segregation measure rose over the decade from 19.1 to 22.0. While none of these results point to major differences, it appears that the changing industry mix served to decrease Hispanic-white segregation, while the entry and exit of establishments seems to have worked in the opposite direction.

As reported in the last panel of table 5.7, the results are most notably different for sex segregation—in particular, with respect to the role of industry composition. First, for the full sample, coworker segregation fell by 3.2 percentage points over the decade, whereas for the sample of continuing establishments, it fell by 2.3 percentage points. Because the baseline coworker segregation measure in 1990 for the continuing establishments sample is slightly higher (25.2) than for the full sample (23.3), on a percentage basis segregation actually declined somewhat more for the full sample, but the difference is small. However, a much sharper difference arises when comparing the change over the decade to that obtained holding the distribution of employment across industries fixed at 1990 levels. In particular, coworker segregation in 2000 is 20.1 in the full sample, but only 14.4 in the fixed-industry-composition results. As a result, coworker segregation for the fixed-industry-composition calculation falls over the decade by a full 9 percentage points, or 38.3 percent.

Industry composition has such a strong influence on changes in measured sex segregation because there was very sharp employment growth in services, which is a highly sex segregated industry with a high share of female employment. Thus, absent the growth in services (which is what we mimic by holding the industry composition of employment fixed), sex segregation would have declined by considerably more. To see this, table 5.8 presents detailed information on isolation and exposure indexes and observed segregation by industry, as well as the distribution of employment of men and women across industries, and industry employment growth over the decade. The table shows, first, that with the exception of transportation, which is a relatively small industry, services is the most segregated industry in both years. The percentage point decline in observed segregation is relatively similar across industries, with the exception of wholesale. In addition, the services industry was the largest employer of women in both years, accounting for 35 percent of female employment in

Table 5.8 Observed sex segregation, by industry, within MSA/PMSA

	Observed segregation		Percentage point (percent) change, 1990–2000	Share of female or male employment in industry		Ratio of 1990 to 2000 industry employment
	1990 (1)	2000 (2)	(3)	1990 (4)	2000 (5)	(6)
Construction						
Female	23.7	21.6		0.02	0.02	0.92
Male	21.1	23.9		0.05	0.06	
Difference	2.5	–2.3	–4.9 (–191.4)			
Manufacturing						
Female	43.7	41.2		0.24	0.17	1.29
Male	28.2	29.8		0.44	0.35	
Difference	15.5	11.4	–4.1 (–26.3)			
Transportation						
Female	49.0	48.8		0.04	0.04	1.05
Male	29.4	32.2		0.07	0.07	
Difference	19.6	16.7	–3.0 (–15.1)			
Wholesale						
Female	40.2	38.6		0.06	0.03	1.50
Male	31.9	31.6		0.10	0.07	
Difference	8.3	6.9	–1.4 (–16.3)			
Retail						
Female	59.3	57.1		0.17	0.20	0.80
Male	42.9	44.7		0.14	0.20	
Difference	16.4	12.4	–4.0 (–24.3)			
FIRE						
Female	69.3	69.1		0.12	0.09	1.24
Male	61.3	63.7		0.05	0.05	
Difference	8.0	5.4	–2.6 (–32.7)			
Services						
Female	74.2	72.7		0.35	0.44	0.76
Male	55.1	58.4		0.15	0.21	
Difference	19.2	14.3	–4.8 (–25.1)			

1990 and 44 percent of female employment in 2000. So, for example, services alone accounts for half of the isolation index in 2000.[31] Services was also the second largest employer of men in both years, but well behind manufacturing. Employment of both men and women in services grew sharply over the decade. This is reflected in the distribution of men and women across industries by year (columns [4] and [5]), as well as in the ratio of overall employment in 1990 relative to 2000, as reported in column (6); services has the lowest ratio (0.76), corresponding to the sharpest growth.

31. This can be seen by multiplying the isolation index for services of 72.7 percent by the employment share of services in female employment of 44 percent, as reported in table 5.8, accounting for 53 percent of the overall isolation index in 2000 of 60.7 corresponding to the sample used in that table.

Table 5.9 **Sex segregation including and excluding services, within MSA/PMSA**

	1990	2000	2000, fixed industry composition
Including services			
Observed segregation			
Female	59.9	60.7	55.1
Male	36.2	40.2	40.3
Difference	23.7	20.5	14.7
Random segregation			
Female	47.7	50.7	50.7
Male	47.2	50.4	50.4
Difference	0.5	0.4	0.4
Effective segregation	23.4	20.2	14.4
Percentage point (percent) change, 1990–2000		–3.2 (–13.5)	–9.0 (–38.3)
No. of workers	1,739,063	2,151,566	2,151,566
No. of establishments	301,029	398,958	398,958
Excluding services			
Observed segregation			
Female	52.1	51.2	48.9
Male	32.9	35.4	35.1
Difference	19.2	15.9	13.8
Random segregation			
Female	41.0	42.2	42.2
Male	40.5	41.9	41.9
Difference	0.5	0.3	0.3
Effective segregation	18.8	15.7	13.6
Percentage point (percent) change, 1990–2000		–3.1 (–16.5)	–5.2 (–27.7)
No. of workers	1,310,125	1,450,311	1,450,311
No. of establishments	236,412	289,206	289,206

Note: Mining is excluded.

As a result of the fact that services is a relatively highly sex segregated industry, coupled with the fact that it is a heavily female industry that grew tremendously over the decade, the services industry plays a large role in overall changes in sex segregation. To reinforce the importance of the growth in services employment in mitigating the decline in sex segregation, table 5.9 shows our calculations of effective sex segregation and how it changed over the decade, with and without holding the industry composition of employment fixed and with and without including services. The top panel shows calculations for all industries, echoing the earlier results.[32] In

32. As noted earlier, the estimates allowing the industry composition to change are slightly different from in table 5.6 because of changes in the sample to do the calculation holding industry composition fixed.

the bottom panel, however, services is simply dropped from the calculation, and the difference in the change in segregation from holding industry composition fixed is only about half as large.[33]

5.6 Robustness to Alternative Segregation Measures

Finally, table 5.10 compares the estimates for the key results using the coworker segregation measure and the Duncan index to see how robust the results are to alternative segregation measures. For segregation by education, the measure of effective segregation in each year and for each educational split is quite similar. The changes in segregation are also similar, with the only exception that the Duncan index points to a larger increase in segregation between those with less than a high school degree versus those with a high school degree or more. The results for black-white and Hispanic-white segregation all show increases over the decade, although the increase in the Duncan index relative to the coworker measure is a little smaller for black-white segregation and a little larger for Hispanic-white segregation. The estimates for sex segregation corresponding to table 5.6 (i.e., unconditional, and conditional on occupation), are very similar for the two measures, with both showing marked declines, and of similar magnitudes. However, holding industry composition fixed appears to have much less impact using the Duncan index. Given that the importance of industry composition using the coworker segregation measure is derived in part from the large share female in services coupled with the strong growth of that industry, it is not surprising that the effect of holding industry constant has less of an impact when using the Duncan index because this index is not sensitive to simple changes in a group's representation in the workforce that are distributed across establishments in proportion to their original distribution.

We argued earlier that we have some preference for the coworker measure over indexes like the Duncan index, most importantly because we think the variation in the coworker measure in response to simple changes in the share of a particular group in the workforce represents meaningful changes in workplace segregation, but also for the more technical reasons discussed in section 5.3. In general, though, the conclusions that can be drawn from the two segregation measures are qualitatively similar and, in particular, the directions of the changes across the decade are always the same. Given the differing properties of the two measures, however, the quantitative answers obviously differ somewhat. Nonetheless, as a summary measure of the comparability of the estimates, the last row of the table shows that the estimated percentage point and percent changes are highly correlated across the two indexes (0.78 and 0.83, respectively), computed across all of the estimates reported in the tables.

33. Note that there was also strong growth in retail, another industry that is relatively sex-segregated.

Table 5.10 **Comparisons of results for coworker segregation and Duncan Indexes, effective segregation, within MSA/PMSA**

	Effective segregation 1990 (1)	Effective segregation 2000 (2)	Percentage point change, 1990–2000 (3)	Percent change, 1990–2000 (4)
Table 5.3				
High school degree or less vs. more than high school				
Coworker	17.3	19.2	2.0	11.3
Duncan	25.3	28.6	3.3	13.0
Less than high school vs. high school degree or more				
Coworker	14.1	16.0	1.9	13.6
Duncan	29.8	37.5	7.7	25.7
Less than bachelor's degree vs. bachelor's degree or more				
Coworker	17.6	20.4	2.8	16.0
Duncan	26.3	28.5	2.2	8.5
Table 5.4				
Black-white				
Coworker	14.0	16.8	2.8	20.3
Duncan	18.4	20.9	2.5	13.5
Table 5.5				
Hispanic-white				
Coworker	19.8	20.4	0.6	3.0
Duncan	19.8	23.2	3.5	17.5
Table 5.6				
Male-female				
Unconditional				
Coworker	23.3	20.1	–3.2	–13.6
Duncan	31.6	28.4	–3.2	–10.2
Conditional on 3-digit occupation				
Coworker	10.4	7.8	–2.6	–24.8
Duncan	11.3	8.5	–2.8	–24.7
Table 5.7				
Male-female				
Unconditional, fixed industry composition				
Coworker	23.4	14.4	–9.0	–38.3
Duncan	31.7	28.5	–3.2	–11.3
Correlation between indexes			.78	.83

Note: See notes to corresponding tables.

5.7 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 [forthcoming], using only the 1990 DEED), we explore the extent to which racial and ethnic segregation is 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 by about 3 to 4 percentage points, or about 20 percent. This increase in racial segregation was reinforced by the entry and exit of establishments and by the changing industry composition of employment, suggesting that there may be forces at work that will lead to the persistence of or even increases in racial segregation. Over this decade, segregation by education also increased, by about 2 to 3 percentage points, or 11 to 16 percent. This increase 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 but with an unchanging pattern of segregation based on these skills.[34]

34. 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.

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 3 percentage points, or 14 percent, from 1990 to 2000. Changes in the occupational distribution of men and women did not play a major role in this decline in segregation. If we hold the distribution of men and women across three-digit occupations fixed, the absolute decline in segregation is roughly the same, although it is larger in percentage terms. On the other hand, shifts in the industry composition of employment worked *against* the decline in sex segregation, as the fastest-growing industry was services, which is also one of the most sex-segregated and most heavily female industries. What this implies for future changes in sex segregation depends on whether the shifts in industrial composition continue as in the recent past or change course, on changes in the distribution of women across industries, and on changes in sex segregation across establishments within industries.

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. Cambridge, MA: National Bureau of Economic Research.

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,* ed. J. Haltiwanger, J. Lane, J. R. Spletzer, J. Theeuwes, and K. Troske, 175–203. Amsterdam: Elsevier Science.

———. 2003. New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of Labor Economics* 21 (4): 887–922.

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* 61 (1): 161–71.

Carrington, William J., and Kenneth Troske. 1997. On measuring segregation in samples with small units. *Journal of Business and Economic Statistics* 15 (4): 402–9.

———. 1998a. Interfirm racial segregation and the black/white wage gap. *Journal of Labor Economics* 16 (2): 231–60.

———. 1998b. Sex segregation in U.S. manufacturing. *Industrial and Labor Relations Review* 51:445–64.

Cortese, Charles F., R. Frank Falk, and Jack K. Cohen. 1976. Further considerations on the methodological analysis of segregation indices. *American Sociological Review* 51 (4): 630–7.

Duncan, Otis D., and Beverly Duncan. 1955. A methodological analysis of segregation indices. *American Sociological Review* 20 (2): 210–7.

Echenique, Federico, and Roland Fryer. 2005. On the measurement of segregation. NBER Working Paper no. 11258. Cambridge, MA: National Bureau of Economic Research.

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. Cambridge, MA: National Bureau of Economic Research.
Glaeser, Edward L., and Jacob L. Vigdor. 2001. Racial segregation in the 2000 Census: Promising news. The Brookings Institution Survey Series. Washington, DC: Brookings Institution, 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* 26 (3): 457–72.
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* 71–72:19–78.
———. 2007. Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. In *Hard to Measure Goods and Services: Essays in Honor of Zvi Griliches*, eds. E. R. Berndt and C. R. Holten, 31–71. Chicago: University of Chicago Press.
———. Forthcoming. Workplace segregation in the United States: Race, ethnicity, and skill. *Review of Economics and Statistics.*
Higgs, Robert. 1977. Firm-specific evidence on racial wage differentials and workforce segregation. *American Economic Review* 67 (2): 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. http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf.
James, Daniel R., and Karl E. Taeuber. 1985. Measures of segregation. In *Sociological methodology,* ed. N. B. Tuma, 1–32. San Francisco: Jossey-Bass.
Jarmin, Ron S., and Javier Miranda. 2002. The longitudinal business database. CES Working Paper no. CES-WP-02-17. Washington, DC: Center for Economic Studies.
King, Mary C. 1992. Occupational segregation by race and sex, 1940–1988. *Monthly Labor Review* 115 (4): 30–7.
Kremer, Michael, and Eric Maskin. 1996. Wage inequality and segregation by skill. NBER Working Paper no. 5718. Cambridge, MA: National Bureau of Economic Research.
McConville, Shannon, and Paul Ong. 2001. Examining residential segregation patterns. The Ralph and Goldy Lewis Center for Regional Policy Studies, Discussion Paper, The University of California at Los Angeles.
Meyer, Peter B., and Anastasiya M. Osborne. 2005. Proposed category system for 1960–2000 Census occupations. BLS Working Paper no. 383. Washington, DC: Bureau of Labor Statistics, September.
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* 27 (4): 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. Madison: University of Wisconsin.
Winship, Christopher. 1977. A revaluation of indexes of residential segregation. *Social Forces* 55 (4): 1058–66.