

Online appendix for ‘Firm turnover and the return of racial establishment segregation’

John-Paul Ferguson
Stanford University
jpferg@stanford.edu

Rembrand Koning
Harvard Business School
rem@hbs.edu

A Correction to equations in the main text

After this article went to print, a colleague pointed out that we had made an error in several of our equations. Specifically, in equations 2, 4, and 5 in the main text, we write that the Theil statistic can be decomposed into $K + 1$ terms: a term for segregation between K mutually exclusive and exhaustive sub-units, and K terms for segregation within those sub-units. Specifically, in equation 2, we write that segregation between US establishments can be decomposed into segregation between areas plus segregation within areas, as follows:

$$\begin{aligned} H_{aj} &= \underbrace{\sum_a \frac{w_a}{w} \frac{E - E_a}{E}}_{\text{Between area}} + \underbrace{\sum_a \frac{w_a}{w} \left(\sum_{j \in a} \frac{w_j}{w_a} \frac{(E_a - E_{aj})}{E_a} \right)}_{\text{Within area, between establishment}} \\ &= H_a + \sum_a \frac{w_a}{w} (H_j^a) \\ &= H_a + \bar{H}_j^a \end{aligned}$$

This equation is missing a term. We depict the within area, between establishment segregation as the weighted sum of segregation between establishments within each area, where the weights are the areas relative sizes, $\frac{w_a}{w}$. However, as Reardon et al. (2000) and others point out, the weights in this sum should be the areas relative sizes *and their relative diversities*. Equation 2 should have read like this:

$$\begin{aligned} H_{aj} &= \underbrace{\sum_a \frac{w_a}{w} \frac{E - E_a}{E}}_{\text{Between area}} + \underbrace{\sum_a \frac{w_a}{w} \frac{\textcolor{red}{E}_a}{\textcolor{red}{E}} \left(\sum_{j \in a} \frac{w_j}{w_a} \frac{(E_a - E_{aj})}{E_a} \right)}_{\text{Within area, between establishment}} \\ &= H_a + \sum_a \frac{w_a}{w} \frac{\textcolor{red}{E}_a}{\textcolor{red}{E}} (H_j^a) \\ &= H_a + \bar{H}_j^a \end{aligned}$$

For the same reason, equation 4 should read that, for any nested levels $x \supset y$,

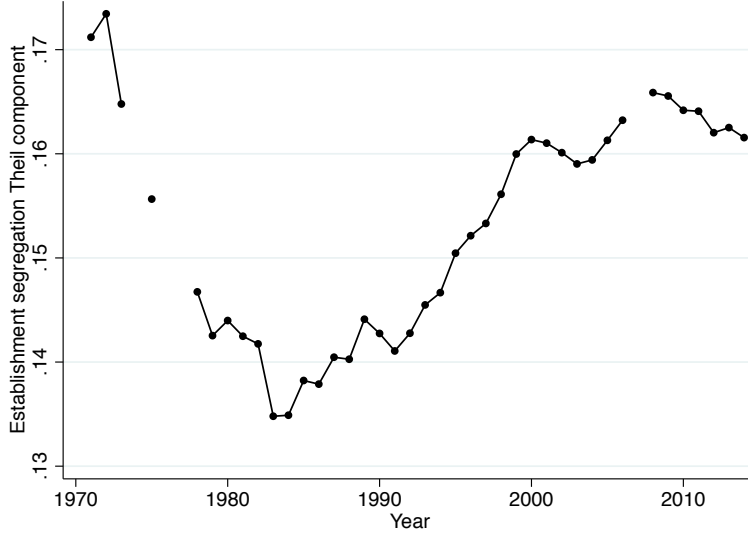


Figure A1: Within-area, between-establishment segregation calculated from the Theil statistic, 1971 to 2014. This figure reproduces Figure 4 from the main text, where the missing term discussed above was not included.

$$H_y^x = \sum_{y \in x} \frac{w_y}{w_x} \frac{E_x - E_y}{E_x}$$

$$\bar{H}_y^x = \sum_x \frac{w_x}{w} \frac{\bar{E}_x}{\bar{E}} H_y^x$$

And similarly, equation 5 should look as follows:

$$H_{ajo} = H_a + \bar{H}_j^a + \bar{H}_o^{aj}$$

$$= \underbrace{\sum_a \frac{w_a}{w} \frac{E - E_a}{E}}_{\text{Between-area}} + \overbrace{\sum_a \frac{w_a}{w} \frac{\bar{E}_a}{\bar{E}} \left(\sum_j \frac{w_{aj}}{w_a} \frac{E_a - E_{aj}}{E_a} \right)}^{\text{Within-area, between-establishment}} + \underbrace{\sum_j \frac{w_{aj}}{w_a} \frac{\bar{E}_{aj}}{\bar{E}_a} \left(\sum_o \frac{w_{ajo}}{w_{aj}} \frac{E_{aj} - E_{ajo}}{E_{aj}} \right)}_{\text{Within-establishment, between-occupation}}$$

We overlooked the term when writing down the equations and then performed the analyses without it. Fortunately, this term is a weight for a sub-unit's relative diversity, and so the sum of them across sub-units should approximate unity.

To get a sense of how our results change when we include this term, compare Figures A1 and A2. Figure A1 replicates Figure 4 from the main text, which is our key result: within-area, between establishment segregation by race has increased over the last generation. That figure was calculated without the $\frac{E_a}{E}$ weight. Figure A2 is the same time series, this time calculated with the term included.

The key difference in Figure A2 is an intercept shift. The entire time series is shifted down somewhat. Why does this happen? Recall that $\frac{E_a}{E}$ is a weight for how diverse a sub-unit is, relative to the total. Intuitively, if we have two sub-units that are the same size and that have the same

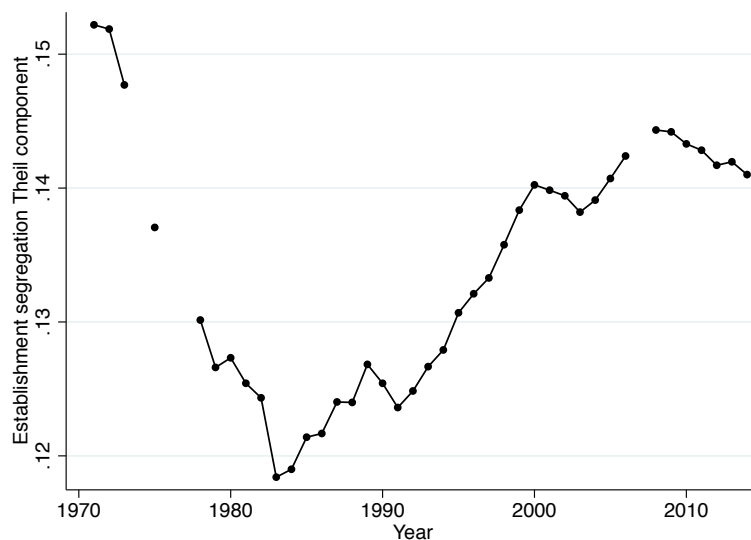


Figure A2: Within-area, between-establishment segregation calculated from the Theil statistic, 1971 to 2014. Compare this to Figure A1. Weighting each county by its relative diversity shifts the entire time series downward slightly, but has no qualitative impact on the overall trend.

(assume it is complete) segregation between workplaces, but one sub-unit is 1% minority and the other is 49% minority, we would think that segregation is “worse” in the second, more diverse sub-unit. The probability that we could get complete racial segregation by chance when only 1% of the workforce is different is far lower than if the workforce is evenly divided.

In practical terms, U.S. counties vary greatly in their diversity. Where there is less diversity in a county (less entropy, in the language of the Theil statistic), we want to down-weight that county’s contribution to the total. In the rarer cases where a county’s racial diversity is much greater than the U.S. as a whole (such as Queens County, New York, or San Francisco County, California), we want to up-weight it. Since there are far more relatively homogeneous counties in the United States than relatively diverse ones, the net effect of including the $\frac{E_a}{E}$ term is to move the time series downward.

However, including the relative diversity term does nothing to alter the pattern of segregation *over time*. We see the same macro-level trends, like a decline in the “regulatory decade” of 1972–1982, followed by a rise through the financial crisis, and slight drift downward afterward. We also see similar micro-level wobbles from year to year. Reproducing all of the results with the relative diversity term included yields similar results. Those are available on request.

We chose to update this appendix for two reasons. The first is because we are doing further work with the Theil statistic, focusing specifically on its decomposition and the contributions of its various terms, and it would appear inconsistent to present a different decomposition in later papers without noting the error we’d made here. The second is because we think science is an open-ended process, and that post-publication comment and suggestion should be taken seriously. We are very grateful to our colleague, whom I think we can just name as Sean Reardon, for noting what he thought was just a typo, and thus saving us from propagating this error through future work.

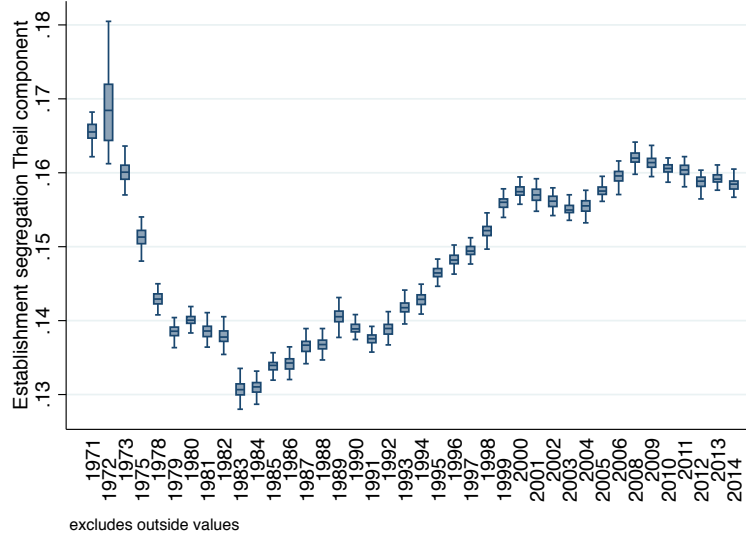


Figure B1: Within-area, between-establishment racial segregation with bootstrapped confidence intervals. Compare with figure 4 in the main text.

B Uncertainty and the Theil statistic

The Theil statistic is not, at root, an inferential statistic like a regression coefficient. Instead it was developed as a descriptive statistic for use in situations where it was presumed full knowledge of the population was available. It does incorporate a measure of uncertainty, specifically, the extent to which knowledge of a sub-unit provides information about properties of a member of the whole. But because it is defined on a population, concepts like variation in the Theil or its components are not defined.

That said, we need not assume that the “true” population is completely captured by such a statistic. As we discuss below, while all establishments with 100 or more employees are required to file EEO-1 surveys with the EEOC, compliance is not universal. Furthermore, firms can move in and out of observation when they are close to the size threshold. Such factors make it useful to ask how much variability there is in a statistic like the Theil, in part so that we can get a sense of how likely a given year-to-year movement in the statistic is to be due to fluctuations in the population, rather than true change.

Following Efron and Tibshirani (1986), we can treat the EEO-1 survey data as arising from a theoretical super-sample wherein what we observe in each year and county is a “random” draw of establishments from the larger, unobserved super-population. In this view, uncertainty reflects variation in the distribution of establishments in that super-population. Concretely, we can use bootstrap methods to sample establishments with replacement within counties and years. We do so 100 times. This allows us to estimate 100 different Theil-statistic components for each year and to use the variation in these estimates to construct confidence intervals.

We do so in figures B1, B2, and B3. These correspond to figures 4, 5, and 10 in the main text. In virtually all years, the bootstrapped confidence intervals are extremely tight. While the fine structure of year-to-year changes might reflect some inherent variability in the underlying population of establishments, the general movements in these Theil components with which we concern ourselves are far larger than what we might attribute to noise.

The exception that proves the rule here is 1972. In that year and only that year, bootstrapping

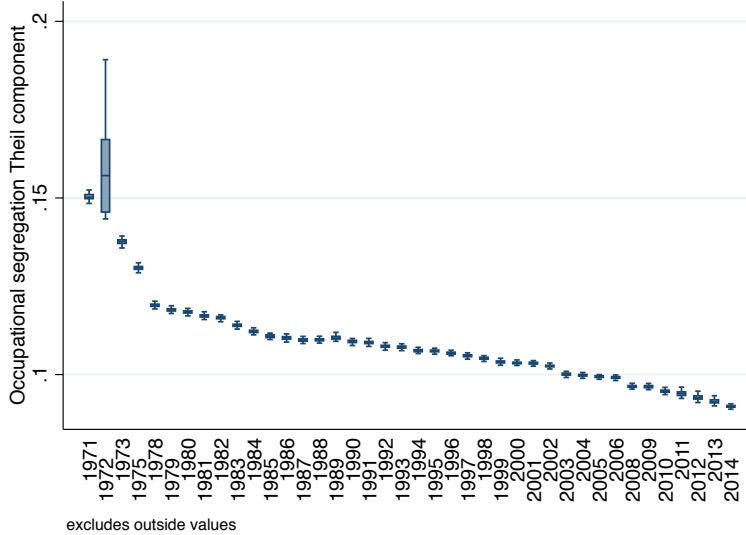


Figure B2: Within-area, within-establishment, between-occupation racial segregation with bootstrapped confidence intervals. Compare with figure 5 in the main text.

produces a much wider confidence interval, suggesting that the final results are far more sensitive to precisely which establishments are used to calculate the statistic. This jibes with previous work’s having raised concerns about the quality of the 1972 data (Stainback and Tomaskovic-Devey, 2012) and with the anomalous *increase* in measured segregation that year, in the midst of a decade of otherwise sharp decrease. These bootstrapped confidence intervals lead us to suspect that the increase in 1972 is indeed a statistical artifact rather than a true increase. By comparison, the chances that increases after the early 1980s are really noise are incredibly small.

C Changes to EEO-1 reporting in 1983 and 2007

We document trends in segregation from the early 1970s through 2014. As Stainback and Tomaskovic-Devey (2012) discuss in their methodological appendix, over this period EEO-1 reporting remained consistent, with two exceptions. In 1983, the reporting threshold was raised from 50 to 100 employees (from 25 to 50 for federal contractors). In 2007, the EEOC temporarily increased reporting rates and permanently changed the EEO-1 survey form.

The 1983 change does not impact our results. Figure C1 excludes all sub-100-employee establishments (including post-1983 federal contractors), in order to keep reporting requirements consistent pre- and post-1983. The decline and rise in between-establishment segregation shown in our main results remains. As we discuss in more detail below, the changes to the EEO-1 survey design in 2007 also do not affect our findings. However, the 2007 reporting expansion does make comparing trends before and after 2007 tricky.

In 2007, the EEOC used Dun & Bradstreet data on the universe of US firms to identify non-reporting establishments and deliberately targeted them for inclusion (Stainback and Tomaskovic-Devey, 2012). This effort was not continued in subsequent years. Clearly, if the additional establishments targeted in 2007 differ from those that regularly report then changes in our segregation measures in 2007 may be due to changes in the reporting sample rather than to changes in establishment segregation.

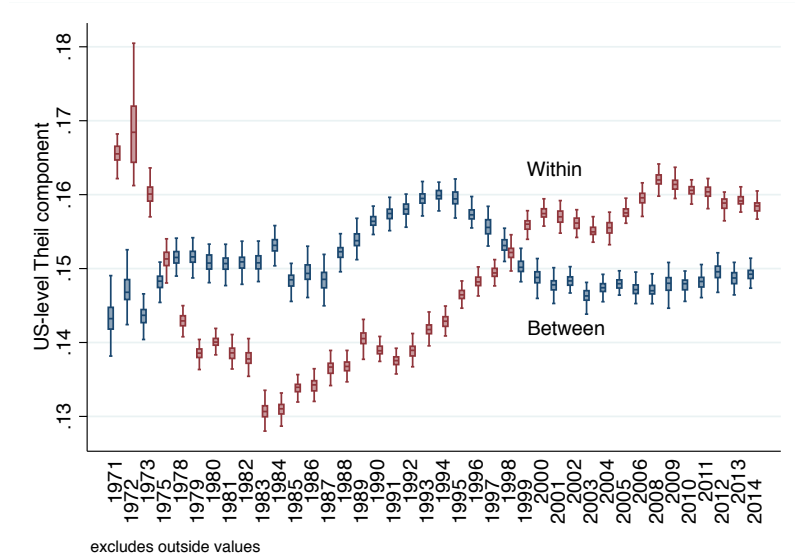


Figure B3: Between-area and Within-area, between-establishment racial segregation with bootstrapped confidence intervals. Compare with figure 10 in the main text.

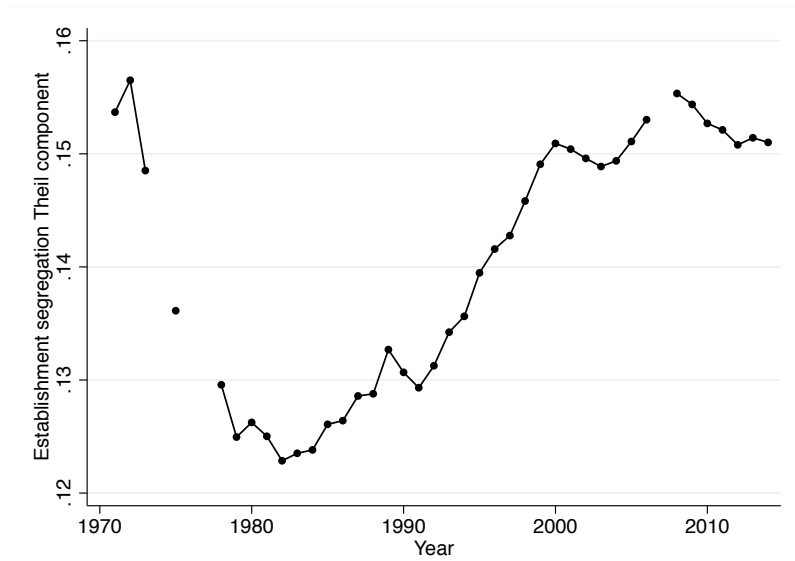


Figure C1: Within-area, between-establishment segregation calculated while excluding all firms with fewer than 100 employees (irrespective of federal contracting status). See the text for information on the post-2007 adjustment. Compare with Figure 4 in the main text.

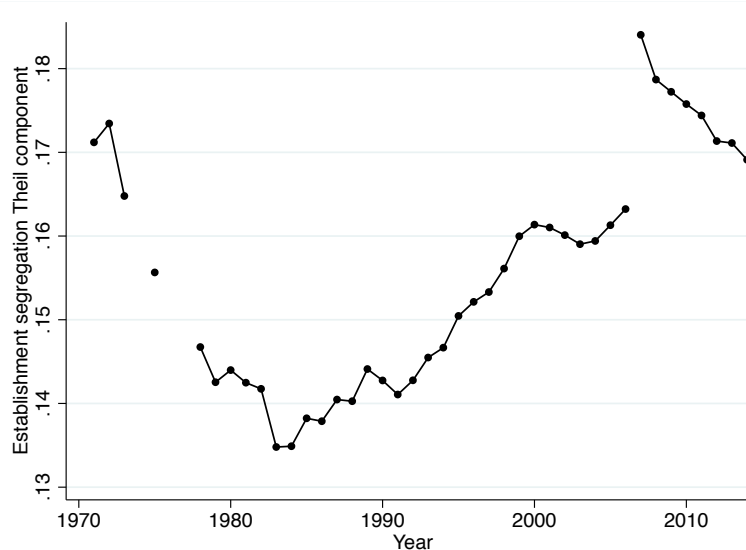


Figure C2: Within-area, between-establishment segregation calculated from the Theil statistic, including 2007 and the 2007 cohort in subsequent years. Compare with Figure 4 in the main text.

Figure C2 shows between-establishment segregation. Unlike the figures in the paper, figure C2 includes data from 2007 and, in subsequent years, retains the cohort of firms that reported for the first time in 2007. Compared to figure 4, figure C2 shows a large jump in the level of segregation and then a steep decline. What about the 2007 cohort explains this jump and decline?

The jump occurs because there is a dramatic increase in first-time reporting by single-establishment firms, firms that also tend to be much more segregated. Table C1 shows the number of new firms and establishments by year, where “new” means a first-time reporter in the EEOC’s data. In 2006 there were 7,053 new establishments from new firms; in 2007 there were over four times as many. There was no comparable jump in recorded establishments for firms that were already in the EEOC’s data: said establishments grew by only 322, to 20,776.

Table C1: New firms and establishments reported by new (first-time-reporting) versus existing firms, 2005–2009

	New firms	New establishments		
		From new firms	From existing firms	Total
2005	3,575	6,836	19,013	25,849
2006	3,947	7,053	20,454	27,507
2007	22,424	30,077	20,776	50,853
2008	4,741	8,864	22,677	31,541
2009	3,405	6,143	16,567	22,710

Table C2 displays summary statistics for new establishments from first-time-reporting firms for the years 2005 through 2009. Establishments in the 2007 sample are much more likely to be from single-establishment firms: 59% in 2007 as against 26% in 2006. The 2007 new firm cohort is also more segregated: 23.5% of establishments are majority-minority in 2007 as against 16.3% in 2006. The average between-establishment Theil contribution (multiplied by 1,000) is 2.29 in 2007, compared to 1.152 in 2006.

The Theil contribution is larger because new single-establishment firms are more segregated

Table C2: Summary statistics (means) for new establishments from first-time-reporting firms, 2005–2009.

	2005	2006	2007	2008	2009
Number of establishments	6,836	7,053	30,077	8,864	6,143
Number of employees	148.5	146.6	177.7	134.4	164.5
Single-establishment firm	0.224	0.264	0.591	0.253	0.274
Percent white	0.681	0.676	0.633	0.675	0.657
Percent black	0.129	0.127	0.125	0.126	0.129
Percent hispanic	0.136	0.143	0.199	0.147	0.156
Percent other	0.054	0.054	0.044	0.052	0.058
Majority minority	0.148	0.163	0.235	0.160	0.185
Between-establishment H	0.963	1.152	2.292	1.576	1.367
contribution: $\frac{w_j}{w} \frac{E-E_j}{E} \times 1,000$					

Table C3: Summary statistics (means) for first-time-reporting, single-establishment firms, 2005–2009

	2005	2006	2007	2008	2009
Number of firms	1,510	1,819	17,653	2,213	1,634
Number of employees	177.7	180.7	211.3	183.1	178.1
Percent white	0.686	0.668	0.609	0.668	0.628
Percent black	0.129	0.121	0.128	0.116	0.135
Percent hispanic	0.131	0.149	0.220	0.156	0.173
Percent other	0.055	0.062	0.043	0.060	0.064
Majority minority	0.151	0.168	0.269	0.172	0.209
Between-establishment H	1.234	1.667	2.724	1.608	0.691
contribution: $\frac{w_j}{w} \frac{E-E_j}{E} \times 1,000$					

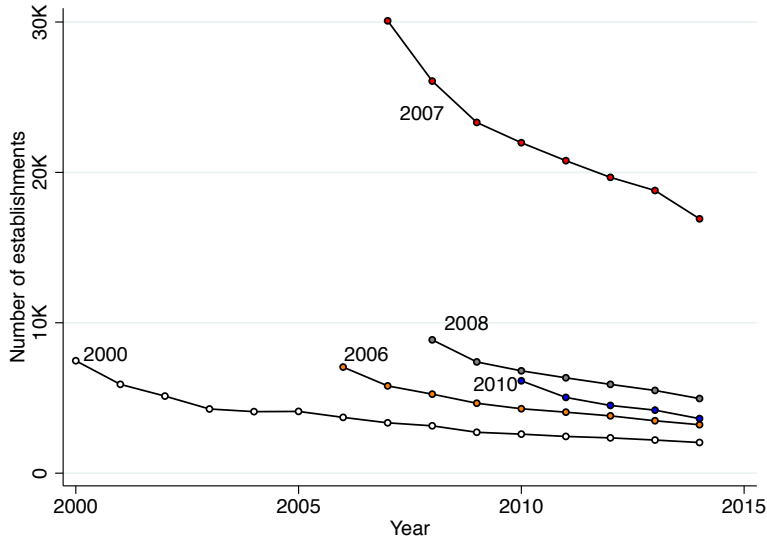


Figure C3: Number of new establishments that remain from firms that report for the first time in the 2000, 2006, 2007, 2008, and 2010.

in all years, and the 2007 single-establishment firms are especially so. Table C3 replicates Table C2 but only includes single-establishment firms. The average between-establishment contribution is higher for single-establishment firms than for multi-establishment ones in all years except 2009. Thus, even if the establishments in the 2007 cohort were as segregated as similar establishments in other years, because the 2007 cohort contains so many more single-establishment firms, we would still see a large increase in between-establishment segregation. The jump occurs because the 2007 reporting expansion netted a much higher proportion of much more segregated single-establishment firms.

The post-2007 downward trend in Figure C2 is also an artifact of this one-time expansion. Figure C3 shows cohort survival curves for establishments from first-time-reporting firms.¹ We show selected years between 2000 and 2010; the 2007 cohort is clearly an outlier in both level and changes. Because new establishments from new firms drove the jump in 2007, as these establishments exit the data we see correspondingly large declines. As we discuss in the main text, year-on-year changes in a cohort's contribution to total segregation can be analyzed into two parts: the changes in the “fundamental segregation” of the component, and changes in the relative size of the component. Formally, $H_2 - H_1 = (S_2 - S_1) \times \frac{W_2}{W_1}$. Because the 2007 cohort is so large relative to the others, the decreases in its size due to establishment turnover are large relative to changes in the other cohorts. By comparison, fundamental segregation in the 2007 cohort changes at rates similar to the other cohorts. Thus the post-2007 decline seen in figure C2 overwhelmingly reflects changes in size in the non-representative 2007 cohort.

This decline also highlights why we cannot just exclude 2007 from analyses and hope to correct for the reporting expansion. To fully correct for the bias it introduces, we must also exclude these “expansion” establishments in all subsequent years. Otherwise, the downward trend in figure C2 will influence calculations post-2007.

Given these patterns, we chose to drop 2007 from our analyses *and* to drop any establishments from firms that first enter the data in 2007. Doing so is conservative, as the 2007 segregation spike

¹Figure 8 in the paper shows survival curves for all new establishments, irrespective of whether the parent firm is a first-time reporter. This is why the counts differ between the figures.

Section D—EMPLOYMENT DATA

Employment at this establishment—Report all permanent full-time and part-time employees including apprentices and on-the-job trainees unless specifically excluded as set forth in the instructions. Enter the appropriate figures on all lines and in all columns. Blank spaces will be considered as zeros.

JOB CATEGORIES		NUMBER OF EMPLOYEES										
		OVERALL TOTALS (SUM OF COL. B THRU K)	MALE						FEMALE			
			WHITE (NOT OF HISPANIC ORIGIN)	BLACK (NOT OF HISPANIC ORIGIN)	HISPANIC	ASIAN OR PACIFIC ISLANDER	AMERICAN INDIAN OR ALASKAN NATIVE	WHITE (NOT OF HISPANIC ORIGIN)	BLACK (NOT OF HISPANIC ORIGIN)	HISPANIC	ASIAN OR PACIFIC ISLANDER	AMERICAN INDIAN OR ALASKAN NATIVE
		A	B	C	D	E	F	G	H	I	J	K
Officials and Managers	1											
Professionals	2											
Technicians	3											
Sales Workers	4											
Office and Clerical	5											
Craft Workers (Skilled)	6											
Operatives (Semi-Skilled)	7											
Laborers (Unskilled)	8											
Service Workers	9											
TOTAL	10											
Total employment reported in previous EEO-1 report	11											

Figure C4: The EEO-1 matrix used in 1997 to report race-gender-occupation counts.

implies that firms which normally report are more integrated than the universe of firms. The EEO-1 data likely places a lower bound on the amount of segregation observed across US workplaces. Consistent with past research using the EEO-1 data, outside of 2007, we find scant evidence that reporting rates changed dramatically (Robinson et al., 2005; Stainback and Tomaskovic-Devey, 2012). Table C1 reveals consistent reporting rates from 2005 through 2009, a pattern that holds when looking at the entire series from 1971 through 2014. We recommend that all analysts use these 2007 corrections when studying pre- and post-2007 trends in the EEO-1 data.

Finally, while the reporting expansion potentially affects our results, the contemporaneous changes to the EEO-1 survey design in 2007 do not. In 2007, “Hispanic” was separated out as an ethnicity (compare the reporting matrices in figures C4 and C5) and reporting establishments were asked to move from relying on “visual inspection” to “self identification” of race and ethnicity. Table C4 shows the percent White, Hispanic, Black, and Other for 2004-2012, while excluding 2007 and the 2007 new-firm cohort from subsequent years. Trends appear consistent over time. The 2008 percentages are consistent with what we would expect given the 2004-2006 percentages. Minorities do appear to increase more slowly from 2009-2012, a plateau likely caused by the great recession. However, by 2013 and 2014 growth appears to be back in line with pre-2008 levels. We find little evidence that the survey redesign led to measurable changes in worker classification and identification.

D Benchmarking EEO-1 workforces against county demographics

We formalize segregation as deviations in composition of sub-units from larger units. In our case the sub-units are establishments, and the unit is the “labor market.” As we discuss in the main text, we define the labor market as the sum of employees in large establishments in the county where

SF 100 - Page 2

Section D-EMPLOYMENT DATA

Employment at this establishment - Report all permanent full- and part-time employees including apprentices and on-the-job trainees unless specifically excluded as set forth in the instructions. Enter the appropriate figures on all lines and in all columns. Blank spaces will be considered as zeros.

Job Categories	Number of Employees (Report employees in only one category)															Total Col A - N
	Race/Ethnicity															
	Hispanic or Latino		Not-Hispanic or Latino													
	Male	Female	White	Black or African American	Native Hawaiian or Other Pacific Islander	Asian	American Indian or Alaska Native	Two or more races	White	Black or African American	Native Hawaiian or Other Pacific Islander	Asian	American Indian or Alaska Native	Two or more races		
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
Executive/Senior Level Officials and Managers 1.1																
First/Mid-Level Officials and Managers 1.2																
Professionals 2																
Technicians 3																
Sales Workers 4																
Administrative Support Workers 5																
Craft Workers 6																
Operatives 7																
Laborers and Helpers 8																
Service Workers 9																
TOTAL 10																
PREVIOUS YEAR TOTAL 11																

Figure C5: The EEO-1 matrix used in 2008 to report race-gender-occupation counts.

Table C4: Percentage of White, Black, Hispanic and Other workers in the EEO-1 data excluding establishments from firms that report for the first time in 2007.

	Percent White	Percent Hispanic	Percent Black	Percent Other
2004	0.691	0.115	0.140	0.054
2005	0.685	0.119	0.141	0.056
2006	0.677	0.123	0.142	0.058
2008	0.668	0.128	0.142	0.061
2009	0.671	0.127	0.140	0.062
2010	0.670	0.127	0.139	0.064
2011	0.667	0.129	0.139	0.065
2012	0.662	0.130	0.141	0.067
2013	0.656	0.133	0.144	0.068
2014	0.647	0.136	0.148	0.069

the focal establishment is located. There are risks to doing so. If there are substantial differences between the workforces of large establishments and the overall workforce in an area, then proxying for the total workforce as we do here can bias our calculations of segregation.

At the same time, there are arguments for using the large-establishment workforce as the reference population. Black men and women are disproportionately likely to work in large establishments, and white men are disproportionately less likely to do so (Dobbin, 2009; Stainback and Tomaskovic-Devey, 2012; Rosenfeld, 2014). Thus finding black workers over-represented in these establishments, relative to the total working-age population in an area, is problematic evidence for segregation insofar as we do not know whether this represents large-establishment hiring practices or rather some prior social process that sorts workers of different races into different types of firms. Black men are also incarcerated at much higher rates than whites. This can bias measurement of the true labor force in an area.

Ultimately, whether it makes sense to use the realized workforce in large establishments to benchmark segregation is an empirical question. The higher the correlation between the large-establishment workforce and the working-age population across counties, the more confident we can be in our choice. With that in mind, we benchmarked our measurements in relevant years against data on county populations from the decennial census. We think that the census years are the best comparisons because the Census Bureau’s intercensal datasets are based on projections, while the EEO-1 surveys are as close to a complete census as the EEOC is able to conduct. We want to be sure that differences between the two are not driven by the Census Bureau’s estimation techniques.

Some care must be taken when comparing estimates of racial composition from the Census and the EEOC. On the census, respondents can choose a racial classification and separately choose whether they have a hispanic background or ethnicity. On the EEO-1 surveys, employers were traditionally asked to count employees according to a racial classification scheme that included hispanic as a separate race. Since 2007, the commission has identified hispanic as an ethnicity, but the procedure for completing the EEO-1 survey asks employers to first determine whether an employee has hispanic ancestry and, if yes, then mark them as hispanic without indicating further racial information. In practice, this means that the EEOC has treated hispanic ancestry as mutually exclusive with other racial classifications (the EEO-1 form has recently added an option for “two or more races,” but very few workers identify as such), while the census has treated it as combining with other racial classifications.

These different classification schemes have implications for how to count the population. The large majority of census respondents who indicate that they have hispanic backgrounds choose white as their race. This implies that many of the workers who are listed as hispanic on EEO-1 surveys would list their race as white on the census. We can test this by comparing the size of the white population on the census with the white plus the hispanic workforce share on the EEO-1 surveys.

Table D1 shows average racial population shares for county workforces compiled from the EEO-1 surveys (E) and racial shares of the working-age population from the decennial census (C). It also shows the differences between the two sources. For example, in 1980 the average large-establishment workforce was 9.9 percent black ($\mu_E^b = .099$), while the typical working-age county population was 9 percent black ($\mu_C^b = .090$). The difference between the two $\delta(b_E, b_C)$ was just under 1 percent, implying that black workers were over-represented in large-establishment workforces. This is exactly what prior work has found.

Table D1 also shows that white workers are substantially under-represented relative to their share of the population. In 1980 this under-representation amounted to 4.7 percentage points, or more than five times the discrepancy for black workers. While previous work has found that whites

Table D1: Reported average shares (μ) of large-establishment workforces (E) and county working-age populations (C) in census years, for white (w), black (b), and white-plus-Hispanic ($w+h$) workers

Year	μ_E^b	μ_C^b	$\delta(b_E, b_C)$	μ_E^w	μ_C^w	$\delta(w_E, w_C)$	μ_E^{w+h}	$\delta((w+h)_E, w_C)$
1980	.099	.090	.009	.849	.896	-.047	.888	-.008
1990	.109	.102	.006	.832	.857	-.025	.875	.019
2000	.114	.094	.019	.792	.873	-.081	.866	-.008
2010	.112	.095	.017	.772	.859	-.088	.862	.002

Large-establishment workforces (E) are compiled from annual EEO-1 surveys collected by the EEOC. County working-age populations (C) are compiled from the decennial census.

Hispanic men and women are separate, mutually exclusive category on the EEO-1 surveys. Hispanic ancestry is a separate question on the census that can vary independently of identified race. Historically, most respondents who claim Hispanic ancestry identify on the census as white.

The EEO-1 surveys used to construct these totals are the same used in the main analysis. Thus for example headquarters-composition surveys are excluded, and the 2007 cohort of entering firms are excluded when calculating workforce shares in 2010. See the main text for more information on the construction of the sample.

are less likely to work in large establishments, this discrepancy seems suspiciously large. It also tends to grow over time—by 2010, it had reached 8.8 percentage points.

The share of white workers in large establishments tracks much more closely to the share of census respondents who identify themselves as white, non-hispanic; the relevant figure for 2010 was 72 percent (Census, 2010). Meanwhile, 16 percent of census respondents listed themselves as having hispanic origin, and of these, more than 75 percent listed white as their race. If we add together the shares of white and hispanic workers from the EEO-1 surveys (μ_E^{w+h}), we get figures that are far closer to the white shares from the census. In 2010, for example, $\delta((w+h)_E, w_C)$ was only 0.2 percentage points.

Table D2 presents correlations between these two sources across counties. When it comes to total size, the two sources are correlated above .9. Tellingly, the correlation for black workers’ shares are also around .9. Quite small numbers of census respondents who identify as having hispanic origin also identify as black (far more choose “some other race”: Census (2010)), so the differences in the EEOC’s and Census Bureau’s classification schemes should have little effect on the correlation for black workers. For whites, the correlation is much weaker, around .67. Here again though we can see the effect of combining white and hispanic workers on the EEO-1 surveys for comparison with the census. The correlation of the *two* groups with the white workforce from the census is, again, close to .9 across counties.

These correlations will never be perfect, not least because the proportion of the total workforce employed in large establishments will vary by county. Nonetheless, the close matches through the years both with the overall composition of the workforce and its variation across counties gives us confidence that we are not systematically mis-characterizing patterns of segregation by focusing on the large-establishment workforce. Certainly given that this is the population of workplaces where enforcement is targeted, it is reasonable to focus on dynamics within it. But more generally we do not find substantial variance from the underlying population, beyond the under-representation of white workers that previous research has documented.

Table D2: Correlations between large-establishment workforce (E) and county working-age population (C) shares in census years

Year	$\rho(\text{Total}_E, \text{Total}_C)$	$\rho(\text{White}_E, \text{White}_C)$	$\rho(\text{Black}_E, \text{Black}_C)$	$\rho(\text{Other}_E, \text{Other}_C)$	$\rho((\text{White} + \text{Hispanic})_E, \text{White}_C)$
1980	.91	.68	.89	.60	.87
1990	.89	.66	.88	.63	.84
2000	.91	.67	.91	.82	.89
2010	.92	.64	.89	.87	.87

Large-establishment workforces (E) are compiled from annual EEO-1 surveys collected by the EEOC. County working-age populations (C) are compiled from the decennial census.

“Other” is defined for EEO-1 surveys as the total of Asian, Pacific Islander (or Native Hawai’ian and other Pacific Islander in 2010), American Indian, and Alaskan Native employees summed across occupations. Hispanic men and women—a separate, mutually exclusive category on the EEO-1 surveys—are not included. “Other” is defined for the census county-descriptive data as the population identifying as neither white nor black. Hispanic ancestry is a separate question on the census that can vary independently of identified race.

EEO-1 surveys used to construct these totals are the same used in the main analysis. Thus for example headquarters-composition surveys are excluded, and the 2007 cohort of entering firms are excluded when calculating workforce shares in 2010. See the main text for more information on the construction of the sample.

E Segregation trends across large counties

We argue that the pattern of between-establishment segregation that we observe, with declines through the early 1980s followed by an increase and then stabilization at relatively high levels after the early 2000s, is not simply a trend in the national, aggregated data. Rather, such trends can be observed within individual counties as well. We presented the trends in segregation in four counties—those containing Chicago, Cleveland, Houston, and Philadelphia—as examples. A reasonable concern is whether these four counties just happen to resemble the national pattern, and how they compare to the United States as a whole.

To give a sense of the variability across counties, we ranked counties in the United States by the size of their workforces employed in large establishments each year, then computed these counties’ contributions to the national establishment segregation component each year. The results are shown in figure E1. The range depicted in gray is the difference between the 5th- and 95th-percentile counties, ordered by their within-area, between-establishment segregation scores. As is apparent from that range, most large counties in the United States also experienced declines in establishment segregation through the early 1980s, followed by a resurgence. There are considerable differences in the *levels* of segregation across cities. Segregation is much lower in counties like Lancaster, Pennsylvania (the 95th-largest county for much of the early 1980s), largely because these counties have less racial diversity and thus less potential for segregation. And in this comparison, establishments in Cook County, Illinois consistently stand out as among the most segregated by racial employment in the country.

Taken as a group, the 100 largest counties in the United States comprise more than 40 percent of *all* large establishments observed by the EEOC in most years. In general, the evolution of establishment segregation across these counties is similar. The non-monotonic pattern we find in the national results is not an artifact of aggregation.

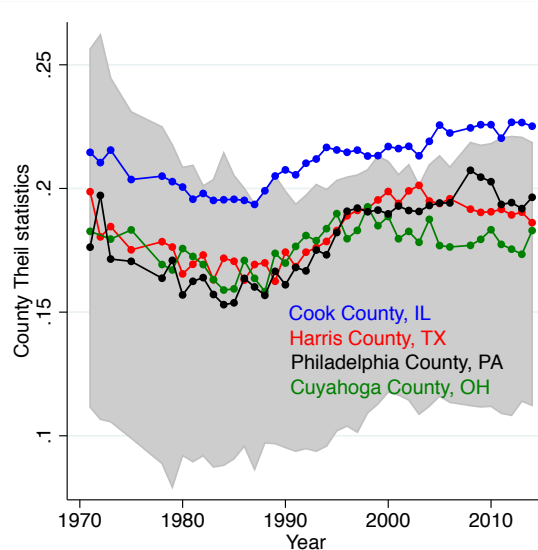


Figure E1: Within-area, between-establishment contributions to the national Theil statistic. For each year, we calculated county components for the 100 largest US counties, ordered by the size of their large-establishment workforce. The range in gray is the difference between the 5th and 95th counties.

F Segregation trends in single- versus multi-establishment firms

The EEOC reports workforce-composition data by establishment, but the ultimate target of anti-discrimination policy is the firm. Accordingly, figure 6 in the main text charts, for multi-establishment firms, the trends in segregation between firms versus the trends in segregation between the establishments within firms. While the level of segregation between establishments within firms is consistently higher, virtually all of the increase in segregation between the establishments of multi-establishment firms has come between the firms. (The higher level of between-establishment segregation within firms in part reflects how multi-establishment firms usually operate across different labor markets with different workforce compositions.)

Figure 6 is important for making the case that it is reasonable, given these data on *establishments*, to talk about a trend of increasing segregation between *firms*. However, there is a risk that it might suggest that increases in segregation are entirely or mostly a large, multi-establishment-firm phenomenon. This is not the case. Figure F1 shows the same within-area, between-establishment component as figure 4 in the main text, here calculated solely for single-establishment firms. A similar increase over time is observable in these establishments as well. Taken together, these two charts do not give direct evidence of what mechanisms may have produced the increase in establishment segregation, but they do suggest that whatever mechanisms have produced it are not dependent on the size or scale of the parent firms.

G Establishment segregation as a function of occupational segregation

When considering occupational segregation in the main text, we treated occupations as nested within establishments. That is, our hierarchical decomposition of total segregation first considers between-labor market segregation, then within-labor market, between-establishment, then finally within-labor market, within-establishment, between occupation. We chose this approach because

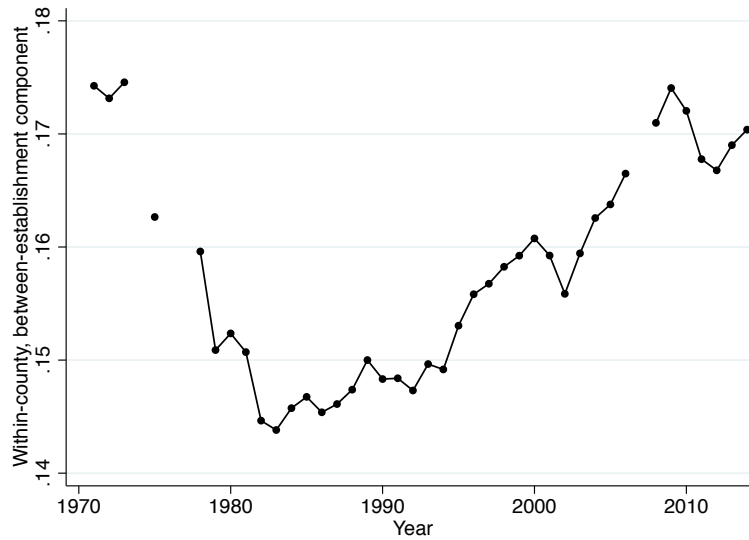


Figure F1: Segregation between establishments, calculated solely for single-establishment firms. The trend seen here is similar to the overall trend seen in figure 4 of the main text, as well as the trend between multi-establishment firms (as opposed to the trend between establishments within those firms) seen in figure 6 in the main text.

it gives results that are conceptually closer to the indices of dissimilarity that prior work, such as Tomaskovic-Devey et al. (2006), calculated. In that work occupational segregation is built up as a weighted average across establishments, meaning that it conditions on the overall composition of the establishment workforces. However, one could also split establishments so as to nest them within occupations. Doing so makes it possible to examine the overall trends in occupational segregation within distinct labor markets, without accounting for how those performing the occupation are allocated across workplaces. As we discuss in the main text, it is useful to consider this trend if and when we want to check whether establishment segregation can be “explained” as a mechanical byproduct of occupational segregation, combined with different allocation of occupations across establishments.

In that light, we also computed H_{aoj} , the hierarchical decomposition of segregation by labor market, then occupation, then establishments (technically the component of each establishment within each occupation). Figure G1 presents \bar{H}_o^a , the within-area, between-occupation component of H_{aoj} . The slowing down of occupational desegregation after the early 1980s is even more pronounced in figure G1 than it is in figure 5. Figure G2 meanwhile plots \bar{H}_j^{ao} , the within-area, within-occupation, between-establishment component. Here the increases depicted in figure 4 are still visible, which is the same as saying that overall changes in occupational segregation can explain very little of the increase in establishment segregation. When we consider the profile of figure G1, this makes sense: occupational segregation cannot be mechanically producing increases in establishment segregation if it is not itself rising.

References

- Census (2010). 2010 census shows America’s diversity. Technical report, United States Census Bureau. Release CB11-CN.125.
- Dobbin, F. (2009). *Inventing Equal Opportunity*. Princeton University Press, Princeton, NJ.

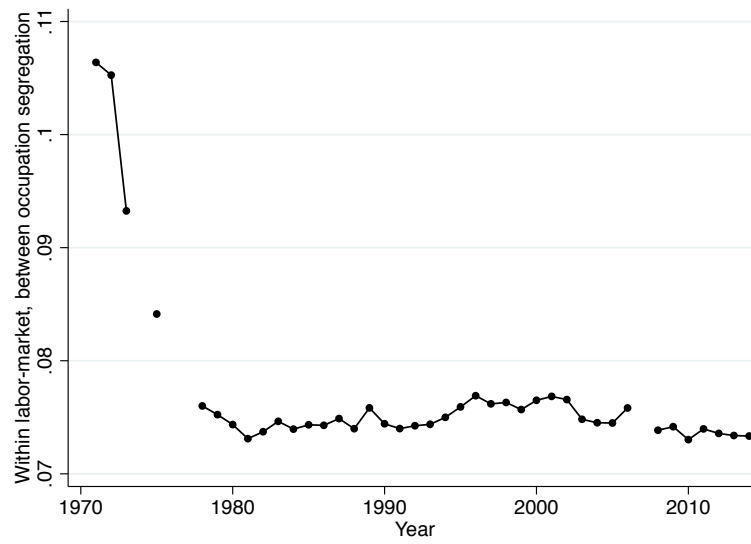


Figure G1: Segregation within labor markets and between occupations, 1971–2014. This figure treats establishments as nested within occupations, as compared to figure 5 in the main text, where occupations are nested within establishments. Thus this depicts overall segregation between occupations in a labor market, irrespective of how members of those occupations are allocated across workplaces.

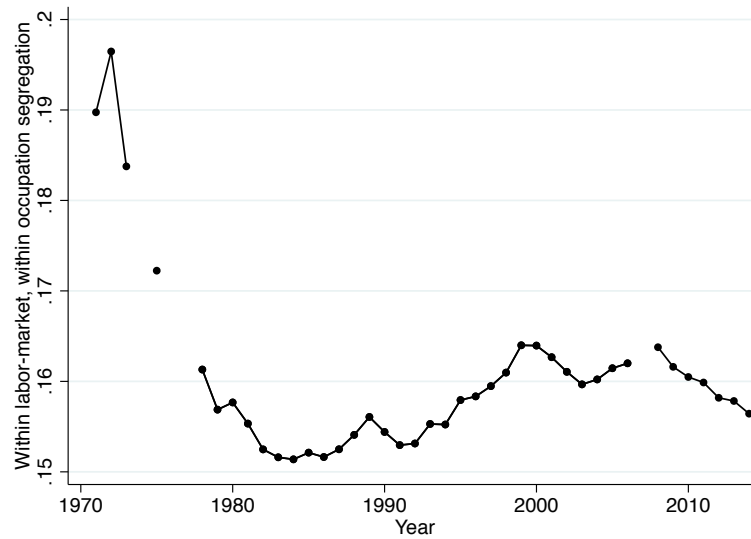


Figure G2: Segregation within labor markets and occupations, between establishments, 1971–2014. This figure treats establishments as nested within occupations, as compared to figure 5 in the main text, where occupations are nested within establishments. Thus while this figure depicts the same general trend as figure 4 in the main text, the magnitude of the decline in between-establishment segregation in the 1970s appears larger, since this figure has “baked in” the decreases in figure G1.

- Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1):54–77.
- Reardon, S. F., Yun, J. T., and Eitle, T. M. (2000). The changing structure of school segregation: Measurement and evidence of multiracial metropolitan-area school segregation, 1989–1995. *Demography*, 37(3):351–364.
- Robinson, C., Taylor, T., Tomaskovic-Devey, D., Zimmer, C., and Irvine Jr., M. W. (2005). Studying race/ethnic and sex segregation at the establishment level: Methodological issues and substantive opportunities using EEO-1 reports. *Work and Occupations*, 32:5–38.
- Rosenfeld, J. (2014). *What Unions No Longer Do*. Harvard University Press, Cambridge, MA.
- Stainback, K. and Tomaskovic-Devey, D. (2012). *Documenting Desegregation: Racial and Gender Segregation in Private-Sector Employment Since the Civil Rights Act*. Russel Sage, New York.
- Tomaskovic-Devey, D., Zimmer, C., Stainback, K., Robinson, C., Taylor, T., and McTague, T. (2006). Documenting desegregation: Segregation in American workplaces by race, ethnicity, and sex, 1966–2003. *American Sociological Review*, 71:565–588.