

**EXCERPT FROM *DOCUMENTING DESEGREGATION*, KEVIN STAINBACK AND
DONALD TOMASKOVIC-DEVEY (2012, RUSSELL SAGE)**

METHODOLOGICAL APPENDIX: USE OF EEO-1 DATA

This appendix provides a deeper introduction to the EEO-1 data that support this book and the analytic choices we have made. The most complete description of the data can be found in Robinson et. al (2005). The first few sections of this Appendix draw heavily on that article.

EEO-1 reports are a partial antidote to the near absence of available time-series workplace data. Because there are now millions of observations, these data allow for disaggregation of segregation trends to the community and industry level as well. In addition, these data have the potential to follow segregation trends in specific workplaces over time, nested in their corporate, industrial, and community context. We use the data in this way in Chapters Six, Seven and Eight.

Description of EEO-1 Data

Title VII of the Civil Rights Act of 1964 and its amendments mandate that both public and private employers submit reports on the sex and race/ethnic makeup of their employees to the Equal Employment Opportunity Commission. Reports include Employer Information Reports (EEO-1), Apprenticeship Information Reports (EEO-2 and 2E), Local Union Reports (EEO-3), State and Local Government Information Reports (EEO-4), Elementary-Secondary Staff Information Reports (EEO-5), and Higher Education Staff Information Reports (EEO-6) (Equal Employment Opportunity Commission 1981). In the past these data have not been available to the scientific community. We focus in this book only on EEO-1 reports of private sector firms. The EEOC publishes aggregate statistics based on these data and use them for

regulatory purposes. Workplace reports are treated by the EEOC as confidential. Access to confidential EEO-1 reports was gained through the use of an Intergovernmental Personnel Act agreement under which Tomaskovic-Devey became an unpaid employee of the EEOC for the purposes of conducting this research. As part of that agreement we are prohibited from identifying individual workplaces or firms.

EEO-1 reports contain establishment employment counts of sex by five race/ethnic (White, Black, Hispanic, Asian/ Pacific Islander, American Indian/ Alaskan Native) groups for nine occupational categories—officials and managers, professionals, technicians, sales workers, office and clerical workers, craft workers, operatives, laborers, and service workers. Firms determine the methods used to tally and classify employees within occupations. The preferred method up until 2007 was a “visual identification,” Currently the preferred method is self-identification by the employee. EEO-1 reports also include information on the establishment’s parent company, industry, and geographic location. Finally, each record states whether or not the firm is a federal contractor. There is no information on other characteristics of people, such as education and tenure, or of jobs such as training or earnings.

Coverage is limited to all private firms with 50 or more employees if federal contractors and 100 or more employees if non-contractors. Prior to 1983, separate reports were required for contractor firms with 25 or more employees and non-contractor firms with fifty or more employees. Firms are instructed that employees do not include temporary or casual employees, but do include leased employees as well as both part-time and full-time employees.

In accordance with guidelines established in Title VII of the Civil Rights Act of 1964, employers use various forms of EEOC reporting. Reporting is contingent on the layout and structure of the firm. A single establishment firm with 100 or more employees (or 50 or more

employees and a federal contract) files a single establishment report. In the year 2000 there were 14,065 single establishment reports. A multi-establishment firm with 100 or more employees (50 or more employees and a federal contract) is required to file a company-wide consolidated report, a headquarters unit report, and individual establishment reports for each establishment with 50 or more employees. It is the case, however, that many firms file establishment reports for workplaces with less than fifty employees and these observations appear among establishment reports. There were 25,410 such firm reports in the year 2000, with individual establishment reports for 158,250 unique workplaces (for year 2000 data, N=224471).

Headquarter and individual establishment reports contain geographic, industry, and employment information for each unit with 100 or more employees or 50 or more employees and a federal contract in the firm. There is also a category referred to as special reports in the EEO-1 files. These special reports include the employment counts for multiple small establishments (under fifty employees) that are part of a larger firm. There were 1059 special reports filed in the year 2000. We do not analyze special reports as they are not workplaces and their geography and industry are not clearly defined.

Coverage Rates

The EEOC estimated in 1966 that 75% of all employers with 100 or more employees who were required to, actually filed EEO-1 reports (Equal Employment Opportunity Commission, 1967). We have no idea how that response rate was calculated and nobody at the EEOC remembers, fifty years on who was responsible. Later EEOC reports do not include response rate estimates. Across all sectors, total private sector employment coverage drops from over 50% in 1966 to just over 40% in 1999. All of the drop in coverage takes place between 1982 and 1985 and is a function of changes in EEO firm size reporting requirements in 1983.

With the exception of the dip in 1983, EEO employment coverage has remained relatively stable over time. While a smaller proportion of all private sector establishments are captured post-1983, coverage rates gradually increase for most industries after 1983.

EEO-1 coverage is higher in sectors with larger firms. Manufacturing has the highest coverage rates (65-75% over time) and construction the lowest (10-20% over time). EEO coverage in services is roughly one-third of that reported by the Bureau of Labor Statistics. Industries characterized by small firms and small workplaces (i.e., construction, retail trade, services, and agriculture) are not well represented. It is also worth reiterating that the EEOC's definition of "employee" excludes temporary and casual workers "hired for a specified period of time or for the duration of a specified job" which may further explain coverage issues for certain industries (Equal Employment Opportunity Commission 1981:vii).

Data Quality Issues

EEO-1 data for years 1966 through 2005 contain little missing data or out of range values across the 5 million plus records examined. Original reliability of firm reports is unknown. There is likely to be measurement error associated with the method and quality of reporting used by firms (Becker 1980; Smith and Welch 1984). The EEO-1 survey instrument collects information on methods used by firms to assess employment composition, (i.e., visual inspection vs. payroll records). However, this variable is not in the current data files.¹ The EEOC does not perform external validity checks on reporting. The only suggestion we can find in the literature of systematic misreporting is provided by Smith and Welch (1984). They indirectly infer from aggregate comparisons with Current Population Survey data that, in the early years of EEO-1

¹ Based on our discussions with representatives at the Equal Employment Opportunity Commission, no one recalls ever using this item therefore analyses files were standardized such that this item was deleted from all years of data.

reports, some firms may have reclassified professionals to managers in order to appear more sex and ethnically integrated in their managerial ranks. This analysis was not conclusive, focusing on aggregate CPS and EEO-1 data comparison and ignoring alternative explanations such as possible changes in occupational composition resulting from secular shifts in industrial structure or firm composition over time.

We found a few problems in the EEO-1 data worth noting for researchers. In a very few cases (less than 0.1%), establishments lacked an assigned Standard Industrial Classification [SIC] code. By comparison, 3.6% of cases in US Current Population Survey have insufficient industry and occupation data (US Census Bureau 2002). A few establishments reported suspiciously high levels of employment although these cases were so rare as to be inconsequential in aggregate analyses. We investigated some of these cases by comparing the name and location of firms to corporate website information on employment. We concluded that, occasionally, consolidated firm reports are miscoded as headquarter or individual establishment reports. This type of miscoding, while rare, might be consequential for more focused studies of specific industries or communities. We ignore this potential small source of error in our analyses.

EEOC data are easy to use and have improved over time. Despite changes in computing technology and political support EEO-1 reports contain almost no missing data, and contain few cases with extreme or unusual values. Data quality seems to be as high or higher than academic and government surveys of individuals or firms. Response rates are probably comparable or superior to surveys of individuals or organizations, although this conclusion is based on the 1966 reported response rate and the relatively stability of employment coverage over time.

Industry Codes

United States Standard Industrial Categories (SIC) codes have evolved over time. SIC codes changed in years 1959, 1972, 1977, 1987. The North American Industrial Classification System [NAICS] has currently replaced the SIC system. This is reflected in the data. Although firms are responsible for specifically describing their “major activity,” the EEOC assigns the SIC codes in the data. This suggests that in most cases firms provide enough information for the EEOC to assign industry codes.

Prior to 1974 the EEO-1 reports contain the 1957 SIC scheme. Between 1974 and 1991 the 1972 SIC scheme is used. After 1991 the 1987 SIC scheme is used.² In the late 1990s the NAICS scheme is also present in the data files. We have created a conversion convention for our own work which identifies changes in SIC codes and standardizes to the 1987 coding scheme. The vast majority of codes do not change between the 1957 and 1987 SIC schemes. Four types of code changes occur.

1. New codes are created for new industries (e.g. 484 Cable TV).
2. New codes are created when new industries split out of old (e.g. 124 coal mining services split out of both 111 anthracite mining and 121 bituminous mining).
3. Old codes are discontinued and merged into new more global codes.
4. Old codes are discontinued and broken up into two new more specific codes.

We treat newly created industries as simple changes in the economy and make no adjustments in industry codes for them. It is likely however, that when 484 Cable TV was created as a new code, some establishments were reclassified from Television to Cable

² In some years the EEO-1 files contain multiple codes. Below we identify the variables in each EEO-1 reporting year file that corresponds to these basic codes.

Television. When a code change is not the creation of a new industry we use the following transformation matrix, always standardizing to the more recent coding scheme.

SIC Standardization rules			
1957 Code	1987 Code	1957 industry name	1987 industry name
012	017	Fruit, Tree Nut and Vegetable Farms	Same
014	019	General Farms	Same
073	078	Horticultural Services	Landscape Counseling & Planning
151	152	General Building Contractors	Residential Building Contractors
151	153		[Non-residential Building]
398	399	Miscellaneous Manufacturing	Manufacturing, NEC
402	401	Railroad Sleeping and Passenger	Railroad
404	401	Railway Express Agency, Inc	
522	507	Heating & Plumbing Equipment	Same
532	596	Mail Order Houses	Non-store Retailers
534	596	Merchandise Vending Machines	
535	596	Direct Selling Organizations	
595	594	Sporting Goods Stores	Miscellaneous Shopping Goods Stores
597	594	Jewelry Stores	
604	609	Trust Companies, No Deposits	Functions Related to Deposit Banking
612	603	Savings & Loans	Savings Institutions
613	615	Agricultural Credit Institutions	Business Credit Institutions
656	153	Operative Builders, Real Estate	Operative Builders, Construction
739	738	Business Services, NEC	Miscellaneous Business Services
1972 Code	1987 Code	1972 industry name	1987 industry name
082	083	Forest Nurseries & Tree Seeds	Forest Nurseries & Gathering of Food
084	083	Gathering of Food Products	
111	123	Anthracite Mining	Same
121	122	Bituminous Mining	Same
None	124	Was included in 111 and 121	Coal Mining Services
264	267	Converted Paper & Paperboard	Converted Paper & Paperboard
266	267	Building Paper & Paperboard	
303	306	Reclaimed Rubber	Fabricated Rubber, NEC
304	305	Rubber and Plastic Hose & Belting	Gaskets, Packing and Sealing Devices and Rubber
307	308	Misc Plastic Products	Plastic Products, NEC
383	382	Optical Instruments & Lenses	Laboratory Apparatus and Analytical, Optical
445	448	Local Water Transport	Water Transport of Passengers
446	449	Services Incidental to Water Transport	Same
471	473	Freight Forwarding	Arrangement of Transportation of Freight and Cargo

Measuring Race

The categories used by the EEOC do not distinguish between race and ethnicity. The data collection instrument instructs respondents to report all Hispanics, regardless of race, as Hispanic. Thus, it is not possible to treat race and ethnicity as conceptually distinct or to examine variation associated with how individual's self-categorize. This reporting convention may be a source of measurement error, at least under some conceptualizations of race and ethnicity. On the other hand, reported ethnic distinctions are from the point of view of the firm and probably match the socially constructed conceptions of race/ethnicity in these workplaces. In any case, researchers have no choice but to proceed under the assumption that these race/ethnic categories are mutually exclusive and socially meaningful.

Prior to 1990, EEO-1 survey forms did not contain separate counts for white employment. For these years, data analysts must estimate the white counts by totaling the number of non-whites in each occupation (i.e., Asian-American/ Pacific Islanders, African-Americans, American Indians/ Alaskan Natives, and Hispanics) and then subtracting this number from grand totals in each occupational category. Prior to 1990 a small proportion (less than 0.1%) of individual establishments contained erroneous grand totals. This could be a result of reporting errors or errors in the data entry process. After 1993, the data were virtually free of this problem. This suggests that EEOC data cleaning procedures improved in the early 1990s. In our analyses we constrain any negative counts to zero. Beginning in 2007 further distinctions between Hawaiian/Pacific Islanders and other Asians were introduced. Our analyses end in 2005.

Measuring Occupations

The EEOC provides the following occupational definitions for employers filling out the EEO-1 surveys. They also provide detailed census based coding, but since these are not widely used by

employers they probably play a smaller roll than these descriptions in helping employers fill out the surveys. Beginning in 2007 further distinctions between executives and lower level managers were introduced. Our analyses end in 2005.

EEO-1 Survey Occupational Definitions

<i>Managers</i>	Individuals who plan, direct and formulate policies, set strategy and provide the overall direction of enterprises/organizations for the development and delivery of products or services. Individuals who oversee and direct the delivery of products, services or functions at group, regional or divisional levels of organizations. Individuals who serve at functional, line of business segment or branch levels and are responsible for directing and executing the day-to-day operational objectives of enterprises/organizations, conveying the directions of higher level officials and managers to subordinate personnel and, in some instances, directly supervising the activities of exempt and non-exempt personnel. This category includes all managers from first line managers to chief executives. It does not include supervisors that lack the responsibilities detailed above. Supervisors are to be counted with the occupation they supervise.
<i>Professionals</i>	Most jobs in this category require bachelor and graduate degrees, and/or professional certification. In some instances, comparable experience may establish a person's qualifications. Examples of these kinds of positions include: accountants and auditors; airplane pilots and flight engineers; architects; artists; chemists; computer programmers; designers; dietitians; editors; engineers; lawyers; librarians; mathematical scientists; natural scientists; registered nurses; physical scientists; physicians and surgeons; social scientists; teachers; and surveyors.
<i>Technicians</i>	Jobs in this category include activities that require applied scientific skills, usually obtained by post secondary education of varying lengths, depending on the particular occupation, recognizing that in some instances additional training, certification, or comparable experience is required. Examples of these types of positions include: drafters; emergency medical technicians; chemical technicians; and broadcast and sound engineering technicians.
<i>Sales Workers</i>	These jobs include non-managerial activities that wholly and primarily involve direct sales. Examples of these types of positions include: advertising sales agents; insurance sales agents; real estate brokers and sales agents; wholesale sales representatives; securities, commodities, and financial services sales agents; telemarketers; demonstrators; retail salespersons; counter and rental clerks; and cashiers.
<i>Clerical, Administrative Support Workers</i>	These jobs involve non-managerial tasks providing administrative and support assistance, primarily in office settings. Examples of these types of positions include: office and administrative support workers; bookkeeping; accounting and auditing clerks; cargo and freight agents; dispatchers; couriers; data entry keyers; computer operators; shipping, receiving and traffic clerks; word processors and typists; proofreaders; desktop publishers; and general office clerks.

<i>Craft Workers</i>	<p>Most jobs in this category includes higher skilled occupations in construction (building trades craft workers and their formal apprentices) and natural resource extraction workers. Examples of these types of positions include: boilermakers; brick and stone masons; carpenters; electricians; painters (both construction and maintenance); glaziers; pipe layers, plumbers, pipefitters and steamfitters; plasterers; roofers; elevator installers; earth drillers; derrick operators; oil and gas rotary drill operators; and blasters and explosive workers. This category also includes occupations related to the installation, maintenance and part replacement of equipment, machines and tools, such as: automotive mechanics; aircraft mechanics; and electric and electronic equipment repairers. This category also includes some production occupations that are distinguished by the high degree of skill and precision required to perform them, based on clearly defined task specifications, such as: millwrights; etchers and engravers; tool and die makers; and pattern makers.</p>
<i>Operatives</i>	<p>Most jobs in this category include intermediate skilled occupations and include workers who operate machines or factory-related processing equipment. Most of these occupations do not usually require more than several months of training. Examples include: textile machine workers; laundry and dry cleaning workers; photographic process workers; weaving machine operators; electrical and electronic equipment assemblers; semiconductor processors; testers, graders and sorters; bakers; and butchers and other meat, poultry and fish processing workers. This category also includes occupations of generally intermediate skill levels that are concerned with operating and controlling equipment to facilitate the movement of people or materials, such as: bridge and lock tenders; truck, bus or taxi drivers; industrial truck and tractor (forklift) operators; parking lot attendants; sailors; conveyor operators; and hand packers and packagers.</p>
<i>Laborers and Helpers</i>	<p>Jobs in this category include workers with more limited skills who require only brief training to perform tasks that require little or no independent judgment. Examples include: production and construction worker helpers; vehicle and equipment cleaners; laborers; freight, stock and material movers; service station attendants; construction laborers; refuse and recyclable materials collectors; septic tank servicers; and sewer pipe cleaners.</p>
<i>Service Workers</i>	<p>Jobs in this category include food service, cleaning service, personal service, and protective service activities. Skill may be acquired through formal training, job-related training or direct experience. Examples of food service positions include: cooks; bartenders; and other food service workers. Examples of personal service positions include: medical assistants and other healthcare support positions; hairdressers; ushers; and transportation attendants. Examples of cleaning service positions include: cleaners; janitors; and porters. Examples of protective service positions include: transit and railroad police and fire fighters; guards; private detectives and investigators.</p>

A clear limitation of these data is that sex/race distributions are aggregated into occupational groups, not the actual jobs used to organize work and its rewards. There is a large substantive difference between a plant manager and an acting foreman. Not surprisingly, comparisons to available job within firm data show that when using EEO-1 estimates we underestimate segregation relative to its true level (Robinson et. al 2005). These analyses show that these underestimates are not systematically related to the true score, nor do they distort associations with important covariates like firm size or status composition. On the other hand, we see large residual variation across industry in estimates of segregation based on job-title versus occupational aggregations within establishments, reflecting variations in industry based divisions of labor. This suggests that these data are more powerful for segregation trend than level comparisons across industries and to segregation level comparisons within industries.

Adjusting Segregation Measures for Occupational-Job Mismatch

Following our prior work (Tomaskovic-Devey et. al 2006) we correct for this source of measurement error in three ways. First, since job titles proliferate with organizational size, all analyses of segregation control for the log of employment size. Second, since the source of error is a mismatch between actual job distinctions and occupational distinctions we make use of information on occupational heterogeneity, the dispersion of cases across the nine occupational categories, to adjust segregation measures upward (Gibbs and Martin 1962).³

When employees are found in fewer occupations we reason that this represents a larger mismatch between job titles and occupational titles. Establishments with low occupational

³ We use the Gibbs-Martin index of heterogeneity: $H = 100 * [1 - ((\sum X_{01-9}^2) / (\sum X_{01-9})^2)]$
Where $(\sum X_{01-9}^2)$ is establishment employment in each occupation squared and then summed across all nine occupations and $(\sum X_{01-9})^2$ is total establishment employment squared.

heterogeneity will also have low segregation because of an increased disjunction between the EEOC occupational categories and actual divisions of labor. An occupationally heterogeneous workplace might have substantial employment in all nine occupational categories. As such, it has an increased chance of displaying high levels of segregation because there are more positions to distribute people across. A firm with all employment in only one occupational category will have no observed segregation in the EEO-1 data. In the real world, however, this firm might make numerous job distinctions within that one occupational category and so have high segregation in practice. We directly adjust our measures of segregation to account for this source of measurement error.

We regress group specific (e.g. white male-white female) measures of segregation on the Gibbs-Martin index of occupational heterogeneity and find as expected that more occupationally heterogeneous workplaces have higher measured segregation. Since the maximum value on the Gibbs-Martin index is 89, we took the difference between the observed workplace level and 89 multiplied by the regression coefficient to adjust segregation measures upward to what they would have been if we had observed maximum heterogeneity. This still produces estimates of segregation that are lower than those observed with job title data. On the other hand, it reduces measurement error and makes between industry comparisons possible.

Still, our third approach to measurement error resulting from the use of occupational data is to remember that we are still underestimating actual job level workplace segregation with these data. For analyses of specific occupations, managers, professionals and craft, it means we are missing some distinctions (e.g. plant manager vs. foreman) that may be sources of gender or race based distinction. So here again, our occupational analyses understate real inequality.