

**APPENDIX C**  
**TECHNICAL APPENDIX**

APPENDIX C TECHNICAL APPENDIX .....268

    §1. *Basic Methodology*.....268

    §2. *Standard Deviations*.....272

    §3. *Affected Workers* .....274

    §4. *Extrapolations*.....275

    §5. *Net Change from 1975-1999*.....275

    §6. *Limitations of the data and methodology*.....276

    §7. *Geographic Heterogeneity*.....276

    §8. *Industrial Heterogeneity* .....278

    §9. *Occupational Heterogeneity* .....278

    §10. *Other Limitations* .....278

    §11. *Sensitivity Tests for Assumptions* .....279

    §12. *Summary of Limitations* .....280

    §13. *Endnote* .....280

**§1. BASIC METHODOLOGY**

This report follows the rules for comparisons used in the methodology reported by Blumrosen, Bendick, Miller, and Blumrosen (1999).<sup>1</sup> In a nutshell, the methodology compares the utilization of minorities or women at one establishment to the utilization of minorities or women at peer establishments in the same occupation, industry, and metropolitan statistical area (MSA). An establishment is said to discriminate if its utilization of minorities or women is significantly below what we would expect if an establishment were to hire without regard to race, national origin, or gender from the labor pool. The labor pool in this study consists of all the workers employed in establishments in the EEO1 database in that particular occupation, industry, and MSA. Presumably, those who are employed have the skills, desire, and ability to hold the jobs in which they are incumbents; for this reason, the law uses the employment patterns of peer establishments to construct a benchmark for measuring discrimination.

To determine the likelihood that an establishment is hiring without regard to race and gender, we compare the number of minorities or women that an establishment actually employs to the number of minorities or women we would expect it to employ. We determine the expected number of employees of each race and gender based on the race and gender composition in the relevant labor market (occupation, industry, and MSA)

within the EEO1 database. We used the data available to the fullest extent possible to define labor markets that were homogeneous yet also large enough to generate enough valid statistical comparisons to perform the required statistical analysis.

In particular, we begin by restricting the relevant labor market to the Metropolitan Statistical Area, which is defined by the Census Bureau to be “a core area containing a large population nucleus, together with adjacent communities that have a high degree of social and economic integration with the core;” integration is determined largely based on commuting patterns and extent of urbanization. Then, we draw comparisons only within an occupational category and industry, where industries are defined at the 3-digit SIC level. While the occupational categories in the EEO1 data are broad (there are only 9 categories), it is likely that the type of work done and the skills required within each occupational category and industry are similar, since industries are fairly well-defined at the 3-digit SIC level.

To further ensure that we are making reliable comparisons the following conditions must be met: (1) the establishment being examined have at least 20 employees in the occupation group being examined; (2) the industry being examined must have at least two other establishments available for comparison, each having at least 20 employees in the occupational category; (3) the industry in the MSA must have at least 120 employees in the occupational group; (4) zero utilization in an occupational group within an establishment being examined would represent a statistically significant difference from employment at peer establishments which do employ minorities and women in that occupational group; (5) either the peer establishments employ at least 200 workers in the occupational group or the establishment being examined constituted less than 80% of all employment in that occupational group.

If members of all races and genders were chosen without regard to their race or gender from the relevant labor market, we would expect an establishment’s employment patterns to be consistent with the hypergeometric distribution. That is, the single most likely outcome is that the firm’s labor force will closely parallel the demographic characteristics of the relevant labor market, but other outcomes are also possible, with the likelihood of any particular outcome given by the hypergeometric distribution (sampling without replacement). The exact formula for the hypergeometric distribution is

---

$$P(x = X) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

where

$x$  = the number of workers of the (unspecified) minority group employed by the firm in this occupation

$n$  = the number of workers employed by the firm in this occupation

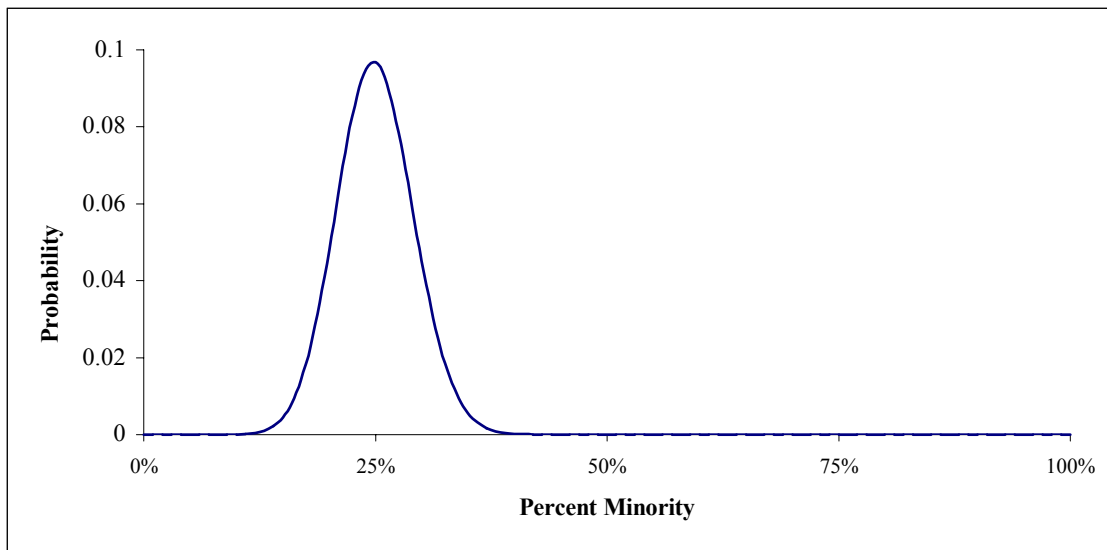
$M$  = the number of workers of the (unspecified) minority group in this occupation in the relevant labor market

$N$  = the number of workers in this occupation in the relevant labor market

Essentially, the hypergeometric distribution is a simple binomial distribution corrected for the existence of a finite population.

As an example, consider the distribution that arises from the following parameters: 100 workers at the establishment in this occupation, 1000 workers in this occupation in the relevant labor market, of which 250 are members of the considered minority group. The probability of different outcomes is shown below.

---



If this hypothetical establishment were hiring without regard to minority status from a labor market with 250 minority individuals and 750 non-minority individuals, there would be a 90% chance that the number of minorities employed by the establishment would fall in the range spanning from 18 to 32 individuals. Thus, there is a 10% chance that an establishment would fall outside this range, and hire fewer than 18 minorities. There is approximately a 5% chance that this establishment would hire 16 or fewer minorities if the establishment's hiring practices were race- and gender-blind.

## §2. STANDARD DEVIATIONS

A standard deviation measures the amount of variation in a random variable. For a normally distributed random variable, there is approximately a 68% chance that the value of the variable will fall within one standard deviation to either side of the mean, and there is approximately a 95% chance that the value will fall within two standard deviations to either side of the mean. Also, for a normally distributed random variable, there is a 90% chance that the variable's value will fall within 1.65 standard deviations to either side of the mean. For a hypergeometric distribution, these percentages are slightly different due to the discrete and sometimes asymmetric nature of the distribution, but in most cases the critical values are similar.

For each of their occupational categories, we classify establishments into the following categories based on their utilization of minorities and women compared to their peer establishments in the industry and MSA:

- No Discrimination: less than 1.65 standard deviations below average
- At Risk: at least 1.65 standard deviations below average. For this category, we do not calculate “affected workers.”
- Presumed: the probability that the establishment’s employment pattern occurred by chance alone is less than 5%. To be in this category, establishments must be at least 2 standard deviations below average, and statistically significant according to Fisher’s exact test.
- Clearly Visible: the probability that the establishment’s employment pattern occurred by chance alone is less than 1%. To be in this category, establishments must be at least 2.5 standard deviations below average, and statistically significant according to Fisher’s exact test.
- Hard Core: at least 2.5 standard deviations below average for 2 of the last 3 years and at least one year from 1991 through 1996; statistically significant according to Fisher’s exact test; never above average since 1991.

Table 1: Threshold Minority Numbers and Percentages Associated with Standard Deviations

Percent Minority in  
 Industry ,occupation  
 and MSA

25% equals Mean

Standard Deviations Below the Mean

Number of Workers in Occupation at Establishment	1.65		2		2.5	
	Percent	Workers	Percent	Workers	Percent	Workers
50	14.9%	7	12.8%	6	9.7%	5
100	17.9%	18	16.3%	16	14.2%	14
150	19.2%	29	17.9%	27	16.2%	24
250	20.5%	51	19.5%	49	18.2%	45
500	21.8%	109	21.1%	106	20.2%	101
1000	22.8%	228	22.3%	223	21.6%	216

Table 1 provides an illustration of the threshold percentage of minorities that would categorize an establishment as a discriminator according to the categories in the report. In this hypothetical illustration, the percentage of workers that are minorities in this industry and occupation is 25%. As Table 1 shows, an establishment that employs 100 workers in this occupational category would be 1.65 standard deviations below the mean (an at risk discriminator) if the percentage of its workers in this occupation that are minorities is less than 17.9% (that is, if it had fewer than 18 minority workers); a minority percentage of 16.3% or lower (fewer than 16 minority workers) classifies the establishment as a presumed discriminator (2 standard deviations). If the establishment has fewer than 14 minority workers it is classified as either Clearly Visible or Hard Core depending on its history.

The percentage of minority employees that categorizes an establishment varies by size of establishment. An establishment employing 1000 individuals in this occupation would be 1.65 standard deviations below the mean if it were 22.8% and 2 standard deviations below the mean if the percentage of its workers that were minorities is 22.3%. Thus, this establishment will be 2 standard deviations below the mean if its minority employment in an occupation is even 2.7 percentage points below the mean or 223 workers.

The number of people in an occupational category within an establishment from which we draw statistical comparisons ranges from 20 to over 10,000. The average number of employees per establishment in an occupational category ranges from roughly 100 to 150 depending on the occupation. If we weight the establishments by size, the average number of employees per establishment in an occupational category ranges from about 250 to 950 individuals. (The unweighted average uses the establishment as the unit of analysis; the weighted average uses the employee as the unit of analysis.)

We place establishments into categories for each of the occupations for which it had valid comparisons. Thus, an establishment could be At Risk in terms of its utilization of managers, Presumed in terms of its utilization of professionals and show no discrimination against office and clerical employees. For the tables in the introductory chapter, we place establishments into the worst category that they fall into across all of their occupations.

If an establishment employs individuals in multiple occupations, the likelihood that it will be a statistical outlier by chance alone increases. For example, if an establishment is defined as a statistical outlier if the likelihood that its utilization of minorities in a particular occupational category occurred by chance alone is 5% (2 standard deviations) then the likelihood that an establishment will be a statistical outlier in any of three occupational categories is 14%. The chance that an establishment will be a statistical outlier for either minorities or women is also higher than 5%. Likewise, when the probability that an establishment's utilization of minorities or women occurred by chance alone in a single occupational category is 1% (2.5 standard deviations-as were about two-thirds of the statistical outliers in this analysis), then the probability that its utilization of minorities or women would be that low in any one of three categories is less than 3%.

### **§3. AFFECTED WORKERS**

We calculate the number of affected minority and female workers at the establishments in the Presumed, Clearly Visible, and Hard Core categories. (No affected workers were assigned to the “at risk” category.) To do this, we calculate the difference between the average utilization of minorities in an occupational group and the utilization that the establishment

---

has. Where underutilization is not significantly significant, we set the number of affected workers equal to zero; likewise, establishments that have higher than average utilization are defined as having no affected workers.

The "Minority" analysis considers all minorities (Native Americans, Hispanic, Blacks, and Asians) as a group, whereas the analysis of Blacks, Hispanics, Native Americans, and Asians considers each minority group separately. Thus, an establishment may have a low utilization of Blacks, for example, but may have an average utilization of Minorities, if it has higher than average utilization of Asians, Hispanics and/or Native Americans. Thus, the affected number of Minority workers will not be the sum of Black, Hispanic, Asian, and Native American affected workers.

#### **§4. EXTRAPOLATIONS**

The number of establishments in the United States, by size of establishment, was determined from the U.S. Census' County Business Patterns. The number of non-reporting establishments in each size category was calculated by subtracting the number of establishments in the EEO-1 database from the number of establishments in the County Business Patterns database. In our extrapolations, we assumed that the percentage of establishments that met the legal definition of discrimination in each size category was identical for reporting establishments and non-reporting establishments. For example, we multiplied the percentage of establishments that were 1.65 standard deviations below the mean among reporting establishments of size 50-100 employees by the number of non-reporting establishments of size 50-100 employees to estimate the number of non-reporting establishments of size 50-100 employees that were 1.65 standard deviations below the mean.

#### **§5. NET CHANGE FROM 1975-1999**

The percentage of minorities in an occupational category in 1975 was multiplied by the total employment in that occupational category in 1999 to obtain the hypothetical number of minorities that there would be in 1999 if the racial distribution in 1999 were equivalent to the distribution in 1975. To determine the net change in the number of minorities, we subtracted the hypothetical number of minorities (based on the 1975 distribution) from the

---



actual number of minorities in 1999. We followed the same procedure to determine the net change in the number of women.

## **§6. LIMITATIONS OF THE DATA AND METHODOLOGY**

In general, statistical data cannot show “intent.” We do observe a great deal of heterogeneity in employment patterns. That is, we observe employment patterns that would be extremely unlikely to arise if all establishments within an industry and MSA were hiring without regard to race and gender from a single, homogeneous labor market. The methodology of the report presumes the existence of intentional discrimination on the part of the employer when an establishment’s utilization of minorities in an occupation is substantially lower than that of its peer establishments in the same industry and MSA. There are three basic assumptions we make about labor market homogeneity. We discuss these assumptions in detail and their implications below. In general, if any of these assumptions are invalid, we may misidentify peer establishments, and incorrectly assume that an establishment has engaged in intentional discrimination.

## **§7. GEOGRAPHIC HETEROGENEITY**

In this analysis, we consider all of the establishments within a Metropolitan Statistical Area (MSA) as members of the same labor market. Because employers may hire from outside the MSA, the labor market includes people who live outside the MSA but work inside the MSA.

Many MSA’s are sufficiently large so as to span regions of quite disparate racial characteristics. For example, the MSA that includes New York City consists of Bronx, Kings, New York, Queens, Richmond, Putnam, Rockland, and Westchester counties. If there is racial heterogeneity (differences in the ethnic composition of particular areas within an MSA), then our methodology must assume that an individual is indifferent between taking a job in his own neighborhood to one anywhere else in the MSA (as well as anywhere outside the MSA if the relevant labor market includes people who live outside the MSA that they work in.) This assumption is likely to be more plausible in some areas (e.g., smaller MSAs with good public transportation, or in areas where minorities are more evenly

---

distributed throughout the entire labor market) than others. To the extent that this assumption is inaccurate (that is, if labor markets are heterogeneous within an MSA), workers at some establishments may resemble the racial characteristics of the neighborhood rather than the whole MSA. When this happens, this analysis will overstate the amount of intentional discrimination in some establishments. For example, using our methodology, we found that 76 (42%) grocery stores in the New York MSA are presumed discriminators against African-Americans and that 67 (37%) grocery stores are presumed discriminators against Hispanics. Some of these establishments may be hiring without regard to race and gender within their local neighborhood for some jobs, but the analysis categorizes them as discriminators because the racial composition of their neighborhood is different from the racial composition of the MSA as a whole. This methodology classifies all occurrences of these differences as intentional discrimination with the understanding that this issue may be raised by individual establishments in legal proceedings.

While geographic areas tend not to be segregated by gender, many MSAs span regions of quite different average income and wealth profiles that may affect woman's participation in the labor market. Also, women of different races may not be equally likely to work; thus racial heterogeneity within an MSA may affect, to some extent, the differential employment of women within MSAs. (For example, black women are more likely to participate in the labor force than women of other races. Thus, geographic areas that have a high concentration of blacks may also have greater female participation in the labor market.) For these reasons, the labor market for women may be somewhat heterogeneous within an MSA

One way to test the extent that geographic heterogeneity is a problem would be to use the addresses of establishments to see if all (or most) of the establishments in an industry in a given MSA are actually located in a concentrated geographic area. If the establishments in given industry were located close to each other, we would expect them to be less affected by geographic heterogeneity, and would therefore expect them to have similar employment patterns. The data does not include the location information necessary to perform such an analysis. Another way to mitigate the problem of geographic heterogeneity would be to use a smaller geographic area as the unit of analysis, rather than the MSA (but such data is not available).

In summary, because the data do not contain precise locations of establishments, we use the MSA as the labor market, but note that the

implications of using the MSA as the labor market should be kept in mind when interpreting the results.

### **§8. INDUSTRIAL HETEROGENEITY**

This analysis assumes that all establishments within a 3-digit industry (SIC) code will be hiring from similar labor market pools. In general, 3-digit industries are fairly well-defined, so we would expect similar employment patterns within 3-digit industries. Examples of 3-digit SICs range from “Eating and Drinking Places,” to “Paperboard Containers and Boxes” and “Fabricated Structural Metal Products.”

### **§9. OCCUPATIONAL HETEROGENEITY**

This analysis assumes that the racial and gender distribution of workers will be homogeneous within each of the nine occupational categories in the relevant labor market. Because we are only comparing the employment patterns of establishments in each occupation within the same 3-digit SIC, it is likely that the type of work done and the skills required within each occupational category are similar. However, within some industries, it is possible that certain occupational groups might not be homogeneous. For example, the gender distribution for professionals in a doctor’s office specializing in obstetrics-gynecology might be different than the gender distribution of a doctor’s office specializing in the treatment of prostate disorders, simply because the underlying gender distribution of doctors in these two specialties is different. (Not only could there be differences in the gender distribution of the types of doctors that would be qualified to work in such offices, but there could also be differences in the gender distribution of the support staff that would choose to work in such offices). We understand that hiring decisions in the above examples must be made in a non-discriminatory manner.

### **§10. OTHER LIMITATIONS**

Even if the labor market is homogeneous, intentional discrimination on the part of employers is not the only reason one might observe heterogeneity in employment patterns. Employees also make choices (for example, whites wanting to work with whites, blacks wanting to work with

---

blacks, whites not wanting to work with blacks, blacks not wanting to work with whites) that may influence the gender and racial distribution of employment. Note, however, that such employee preferences do not constitute a justification for employment decisions, and that compliance by employers in sustaining such preferences may be illegal.

## §11. SENSITIVITY TESTS FOR ASSUMPTIONS

It is impossible to test how plausible the assumptions underlying the methodology are, especially within the EEO-1 database itself. We can, however, do a sensitivity test for the methodology by applying it to white males. We understand that there is no legal authority to use the EEO-1 data for this purpose, and we do not address any issues relating to affirmative action or reverse discrimination. However, because we expect that the likelihood of intentional employment discrimination against white males is relatively small, we might (at least partly) attribute the statistical outliers we find for white men to limitations in the methodology.

When we implement this methodology for white men, we find that there is less heterogeneity in the employment of white males than for other race/gender groups, as about 29 to 36 percent of comparisons are 2 standard deviations below the mean for race/gender minority groups but about 21% of comparisons for white males are 2 standard deviations below the mean. However, this is still a substantial number of statistical outliers for white men. One might suspect that some of the statistical outliers for white men are due to limitations in the methodology, such as heterogeneity, or the uneven breakdown of historical patterns of discrimination and surmise that similar limitations in the methodology could apply to minority groups.

It is also possible that the average employment in an industry, occupation, and MSA is affected by industrial practices. For example, if some establishments in an industry have strong affirmative action programs, these establishments may appear to be statistical outliers in terms of their employment of white men. If this is the case, the other establishments in the same industry (that is, those with less strong affirmative action programs) may appear to discriminate against minorities because the industry average is higher than it would otherwise be due to those with strong affirmative action programs. Similarly, if the percentage of minorities in an industry is low due to discrimination, then an establishment that is hiring more minorities than the industry average may appear to be a statistical outlier in

terms of its employment of white men (when it is simply deviating from the industry average which has a low percentage of minorities due to discrimination). One consequence of discrimination against minorities or women by an important part of an industry may be to lower the average against which all establishments are measured, thus obscuring discrimination on the part of other establishments.

## **§12. SUMMARY OF LIMITATIONS**

In summary, we used the data available to the fullest extent possible to define labor markets that were homogeneous yet also large enough to generate valid statistical comparisons. However, as described above, there may still be substantial heterogeneity within the labor markets we defined. To the extent that the labor market for the establishment under analysis is heterogeneous within an MSA, occupation, and 3-digit-SIC, respect to any of the above factors, this analysis may overestimate the amount of intentional discrimination. The statistics we report are consistent with widespread “intentional discrimination” on the part of employers, but there are other possible explanations for what we observe, and this dataset cannot distinguish among these explanations.

## **§13. ENDNOTE**

- 
1. Blumrosen, et al, EMPLOYMENT DISCRIMINATION AGAINST WOMEN AND MINORITIES IN GEORGIA (1999) Rutgers Law School.
-