People v. Castro (1989). 545 N.Y.S.2d 985.
Presseisen v. Swarthmore College (1978). 442 F.Supp. 593 (E.D.Pa.), *aff'd without opinion,* 582 F.2d 1275 (3d Cir. 1978).
Rajendar v. University of Minnesota (1984). 730 F.2d 1110 (8th Cir. 1984).
Roberts, H. V. (1979). Harris Trust and Savings Bank: An analysis of employee compensation. Report 7946, Dept. Economics, Graduate School of Business, Univ. Chicago.
Roberts, L. (1991). Fight erupts over DNA fingerprinting. *Science* **254** 1721–1723.
Scott, E. L. (1977). *Higher Education Salary Evaluation Kit.* American Association of University Professors, Washington, D.C.
Segar v. Civiletti (1981). 508 F.Supp. 690 (D.D.C. 1981), *aff'd sub nom.* Segar v. Smith, 738 F.2d 1249 (D.C.Cir. 1984), *cert. denied sub nom.* Meese v. Segar, 471 U.S. 1115 (1985).
Sharif v. New York State Education Department (1989). No. 88 Civ. 8435 (S.D.N.Y. 1989).
Smith v. University of North Carolina (1980). 632 F.2d 316 (4th Cir. 1980).
Sobel v. Yeshiva University (1988). 566 F.Supp. 1166 (S.D.N.Y. 1983), *rem.,* 707 F.2d 1478 (2d Cir. 1986), 656 F.Supp. 587 (S.D.N.Y. 1987), *rev'd and rem,* 839 F.2d 18 (1988).
Stacy, D. R. and Holland, C. L. Jr. (1984). Legal and statistical problems in litigating sex-discrimination claims in higher education. *Journal of College and University Law* **11** 107–177.
Tande, C. M. (1989). DNA typing: A new investigatory tool. *Duke Law Journal* 474–494.
Trout v. Hildago (1981). 517 F.Supp. 873 (D.D.C. 1981), *aff'd sub nom.* Trout v. Lehman, 702 F.2d 1094 (D.C.Cir. 1983), *vac. on other grds,* 460 U.S. 1056 (1983), 652 F.Supp. 144 (D.D.C. 1986).
United Airlines, Inc., v. Evans (1977). 431 U.S. 553.
Vuyanich v. Republic National Bank of Dallas (1981). 521 F.Supp. 656 (W.D.Tex. 1981), *vac. on other grds,* 723 F.2d 1195 (5th Cir.), *cert. denied,* 496 U.S. 1073 (1984).
Watson v. Fort Worth Bank and Trust (1988). 487 U.S. 977.
Ward's Cove Packing Co. v. Atonio (1989). 57 U.S.L.W. 4583.
Wolter, K. M. (1991). Accounting for America's uncounted and miscounted. *Science* **253** 12–15.
Wolter, K. M. and Causey, B.D. (1991). Evaluation of procedures for improving population estimates for small areas. *J. Amer. Statist. Assoc.* **86** 278–284.

Comment

Delores A. Conway

Gray's article addresses statistical problems and concerns prevalent in legal cases of employment discrimination. Although she focuses on universities and the academic environment, the statistical issues apply to more general employment settings. In particular, the treatment of outliers, omitted variables, measurement issues, selection of variables, delineation of the population and comparison across competing groups arise in most legal cases of employment discrimination (Finkelstein, 1980). These problems expose the heart of statistical evidence and determine its probative value in legal settings.

One of the strengths of the paper lies in the numerous citations to actual legal cases that illustrate the use of specific methods. Gray notes that similar statistical results may be probative in one case and completely dismissed in another. Tabulated results from two legal cases illustrate the interplay between the legal and statistical issues when assessing employment discrimination at universities.

I commend the author for a careful and comprehensive discussion of the legal and statistical issues. This paper should be especially useful to practitioners and provides a checklist of problems to be addressed in the development of statistical evidence. My comments attempt to clarify and extend the discussion, as well as provide an economic perspective. The multiple regression framework shows how the statistical issues are interrelated and how summaries change with different viewpoints of the data. Two examples from legal cases illustrate the statistical complexities in assessing discrimination across different job structures within an organization. We conclude with some additional comments on the role of statistical evidence in Title VII legal cases.

STATISTICAL MODELS AND JOB STRUCTURES

Gray lists many of the considerations in the use of statistical methods to measure discrimination. Although they are presented in a somewhat isolated fashion, many of the statistical difficulties are interrelated. Solutions to one set of problems often resolve or magnify others.

The development of appropriate statistical models for Title VII cases is not a simple matter, because of the lack of a clear, causal model of the employment process and of limitations from observational data.

Delores A. Conway is Associate Professor of Statistics, School of Business Administration, University of Southern California, Los Angeles, California 90089-1421.

Gray offers many helpful suggestions regarding the population definition, delineation of appropriate job groups, measurement issues, variable selection and use of transformations and interactions. Each of these steps requires considerable care so that the resulting comparisons are accurate and appropriate.

The multiple regression models typically used in studies of salary discrimination provide a useful framework to see the complexity of the issues and how they are interrelated. Let Y be a measure of income such as salary or log salary, $\boldsymbol{X}$ refer to the set of observed job-qualification variables and Z be an indicator variable that equals 1 for females and 0 otherwise. (The same conclusions apply when the indicator Z identifies other protected classes, such as blacks or hispanics.) The direct regression of Y on $\boldsymbol{X}$ and Z is given by

$$E(Y \mid \boldsymbol{X}, Z) = a + \boldsymbol{b}'\boldsymbol{X} + cZ, \tag{1}$$

where c estimates the mean difference in female salaries at the same level of observed qualifications. As Gray notes, the selection of which variables to include and the form of the model are complex decisions and directly affect estimates of c, a key parameter for measuring discrimination. However, these decisions affect more general assessments of employer behavior, especially the job structure of the organization.

To fix ideas, we consider data from an actual legal case alleging salary discrimination against females at a major retail company. The results are from a larger study investigating possible placement and salary discrimination against females employed by the company from 1974–1983 (Conway and Roberts, 1986b). We focus on the subset of 104 employees in the three lower job ranks of a large department who were hired during 1978–1979. For each employee, observed income and qualification measures include starting salary and job level, salary increases, changes in job level or title, age, years of formal education, advanced degrees and major, years of work experience prior to joining the company and seniority level computed from date of hire. For this group, the overall mean log salaries are 9.679 for males and 9.611 for females, indicating a 6.8 percent shortfall in female mean salaries.

Within Job Levels 1, 2 and 3, there are 32, 31 and 41 total white employees and 15, 16 and 14 females, respectively. Figure 1 graphs observed log salaries and educational level for the 104 employees in the three job levels. Separate graphs for males and females remove obscurities and isolate any potential sex interaction between salaries, education and job level.

The graphs of the pooled data are essentially flat, although there is a slight positive relationship within each job group. It is noteworthy that educational level does not account for the substantial variation in salaries across different job groups, for either males or females. In particular, education alone does not explain why employees in Job Level 3 are paid significantly more than those in Job Level 1. The estimated sex coefficient c from the direct regression of log salary on education and sex for the pooled data is -0.062,

FIG. 1. *Log salary versus educational level for* 104 *employees in three job groups at a retail firm.*

TABLE 1
Pooled regression relationship among salary, job qualifications and sex across 104 employees in three job levels at a retail firm

Variable	Estimated coefficient	Standard error	t value
	a. Direct regression results		
Constant	9.274	0.161	57.62
Sex	−0.0261	0.0255	−1.02
Education level	0.00587	0.00962	0.61
Age of employee	0.00047	0.00013	3.52
Work experience	0.00053	0.00083	0.64
Work experience squared	−0.000006	0.000003	−1.82
Seniority level	0.00423	0.00076	5.57
Seniority squared	−0.000013	0.000003	−4.22
	Adjusted $R^2 = 0.478$	Standard error = 0.1257	
	b. Reverse regression results		
Constant	4.891	0.4817	10.16
Sex	−0.0082	0.0174	−0.47
Log salary	0.495	0.0498	9.94
	Adjusted $R^2 = 0.500$	Standard error = 0.0862	

statistically significant and largely unchanged from the overall comparison. The multiple R squared for the model is 0.056, reflecting very low predictive ability.

The picture changes dramatically with the inclusion of additional job qualifications X. Table 1a gives the results for the direct regression of log salary on sex, education level, age, work experience and seniority level. The estimated sex coefficient shows an approximate 2.6 percent shortfall in female mean salaries that is not statistically significant. We define $Q = a + b'X$ as the estimated qualification index for the model from the direct regression coefficients in Table 1a. Figure 2 graphs log salary versus the qualification index and shows a much clearer separation of the three job levels.

Although salaries and education level in Figure 1 appear unrelated to the three job levels within the

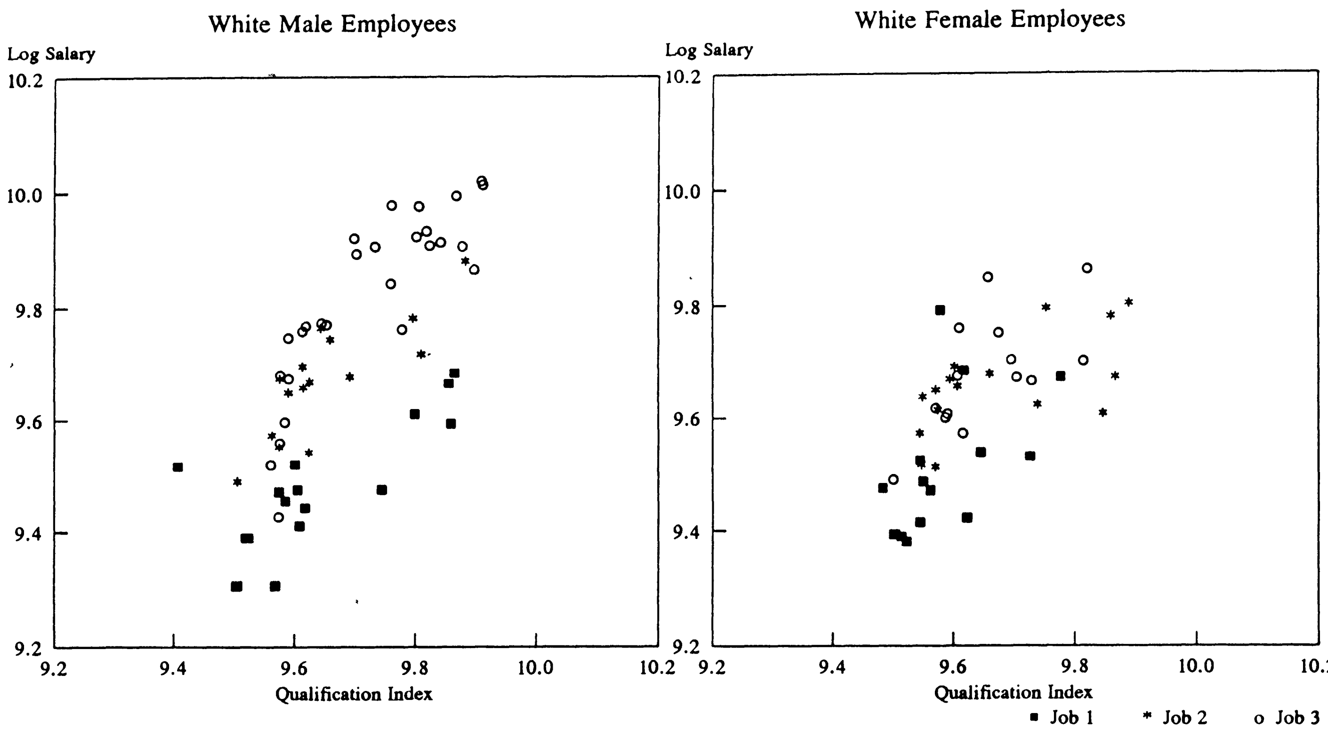

FIG. 2. *Log salary versus qualification index based on pooling the three job groups at a retail firm.*

department, Figure 2 shows a much clearer relationship among salaries, qualifications and the job structure. In particular, additional qualifications help to explain the lower salaries in Job Level 1 relative to the other two job groups. Table 2a summarizes the results when the same model is fit separately to each job level and confirms that the relationship between salary and job qualifications varies across the job groups. Work experience and seniority account for variations in log salary for all three groups. Education has a significant effect in Job Level 2, in contrast to the age of the employee, which is significant for Job Levels 1 and 3.

The graphs of log salary versus the qualification index estimated within each job level from the model appear in Figure 3. The relationship between salary and job qualifications is stronger within each job group, confirmed by the higher adjusted R^2 values and lower standard errors of the residuals in Table 2a. There is less overlap in salaries and qualification for employees in different job groups, and the three job levels are more distinct. Further refinements might consider additional qualifications and develop separate models for each job group.

As the statistician incorporates more information about the employment process, conclusions about potential discrimination often change across the different analyses. Mean overall salaries for females are 6.8 percent lower than for males, and this difference is significant. The direct regression results in Table 1a take into account differences in qualifications between the two groups and show approximate parity in mean salaries. Within each job group, the separate regression results for the same model show a significant 7.1 percent mean salary shortfall for females in Job Level 3 and approximate parity for Job Level 2. Although Job Level 1 shows an initial 8.5 percent mean female excess in salaries, removing the one female outlier in Figure 3 results in approximate parity.

The results appear contradictory and perplexing; nonetheless, they are typical of results in employment discrimination cases. The example is also interesting since the pooled regression estimates of c support the defendant, whereas the disaggregated estimates in Table 2a support the plaintiff. In many legal cases, it is the other way around, with the pooled regression results indicating significant salary shortfalls for females. This is in part due to the larger sample size and the tendency of the pooled regression results to estimate the combined effects of potential hiring and salary discrimination (Malkiel and Malkiel, 1973).

As noted by Gray, inclusion of additional variables and disaggregation of the data can change assessments of discrimination. In fact, more general assessments of employer behavior change because of these considerations. The job structure within the organization and relationship among salaries and job qualifications become progressively clearer in Figures 1 through 3. We see the importance of correct model identification to understand and evaluate employer behavior accurately.

A critical question facing the statistician is when to conclude the model development and present the re-

TABLE 2

Individual regression results within three job levels to describe the relationship among salary, job qualifications and sex for 104 retail employees

	Job level 1		Job level 2		Job level 3	
Variable	Coefficient	*t* value	Coefficient	*t* value	Coefficient	*t* value
			a. Direct regression results			
Constant	9.230	46.90	9.090	43.13	9.129	49.92
Sex	0.0847	2.99	−0.0252	−1.18	−0.0706	−2.06
Education level	0.0016	0.14	0.0295	2.74	0.0186	1.64
Age of employee	0.00042	3.00	−0.00013	−0.41	0.00051	3.51
Work experience	−0.0023	−2.00	0.0015	1.65	0.0011	1.16
Work experience squared	0.000008	1.93	−0.000006	−1.80	−0.000011	−2.10
Seniority level	0.0067	3.30	0.0038	5.04	0.0067	4.03
Seniority squared	−0.000032	−1.65	−0.000011	−4.48	−0.000015	−3.36
Adjusted R^2	0.602		0.631		0.674	
Standard error	0.0755		0.0567		0.0887	
			b. Reverse regression results			
Constant	2.973	3.64	2.760	3.34	3.210	4.30
Sex	−0.074	−3.66	0.020	1.29	0.026	1.05
Log salary	0.686	7.96	0.715	8.37	0.673	8.84
Adjusted R^2	0.687		0.695		0.680	
Standard error	0.0569		0.0435		0.0678	

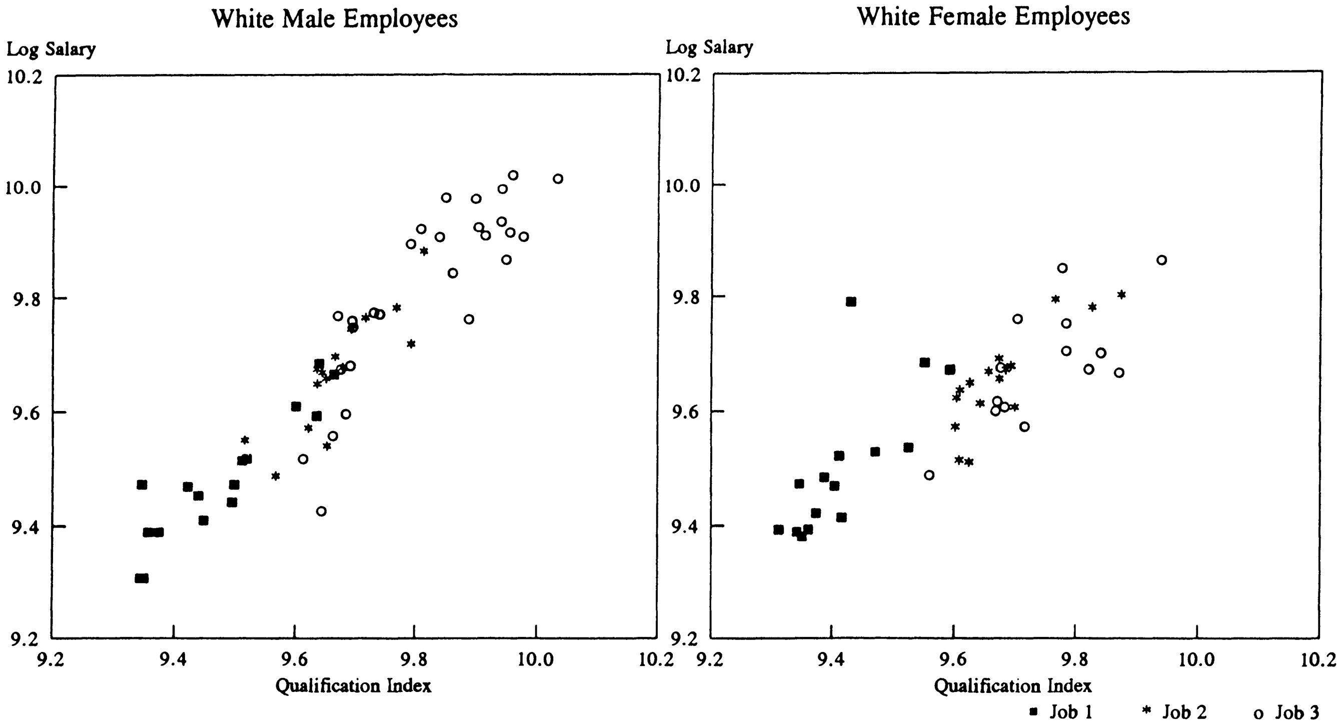

FIG. 3. *Log salary versus qualification index constructed within each job group for* 104 *employees at a retail firm.*

sults as an accurate representation of employer behavior. Although statistical measures of fit and diagnostic tests aid this decision, there is always the risk that the analysis is incomplete or the model not correctly identified. Gray does not address this specific question, and it would be helpful to obtain her views. A related question concerns the presentation and communication of complex statistical results to a court that is generally not educated in statistical methods. In my own work, graphs and tables help to clarify statistical summaries, especially those that illuminate the raw data. Gray's suggestions regarding presentation of results would also be beneficial.

DIFFERENT PERSPECTIVES OF EMPLOYMENT DATA

A major concern in legal cases of employment discrimination is an accurate assessment of potential discrimination. The adversarial nature of litigation encourages participants to provide evidence most favorable to their side. The statistical results for *Ottavani v. SUNY at New Paltz* in Tables 4 and 5, respectively, of the article show substantial disagreement between the plaintiff's and defendant's measures of discrimination. Gray explains that the differences result from using different measures of seniority.

In many legal cases, conflicting results are not so readily explained and may arise from distinct viewpoints of the data. These issues are more difficult to resolve, since both perspectives are often justified, and the observed data alone cannot resolve the disagreement. Here, underlying models of the employment process are critical. It is important for the statistician to view the data from many perspectives in order to obtain a more comprehensive understanding as well as to anticipate rebuttal testimony.

As an illustration, reverse regression provides a different perspective of fairness, namely, whether males and females have similar qualifications at given salary levels. Unfairness results as shunting with females having significantly higher mean qualifications than their male counterparts at the same salary. Defining $Q = a + \mathbf{b}'\mathbf{X}$ as the estimated qualification index in (1), we write the reverse regression model as,

$$E(Q \mid Y, Z) = a^* + b^*Y + c^*Z. \tag{2}$$

The coefficient c^* estimates the mean disparity in qualifications for females at the same salary level. By viewing fairness from different perspectives, direct and reverse regression each examine a different conditional relationship among salaries, qualifications and sex. When used together, they tend to give a fuller perspective of the observed employment data.

The direct regression estimate of potential salary discrimination against females from (1) is c, and the comparable reverse regression estimate from (2) is the ratio, $-c^*/b^*$. In actual applications, these two estimates may conflict and suggest different conclusions about discrimination. Conway and Roberts (1983, 1984), Dempster (1984, 1988) and Goldberger (1984, 1988) formulate different causal models of employer behavior where direct and reverse regression are appropriate for assessing discrimination in the underlying employment process.

For the retail employees, the reverse regression estimate $-c^*/b^*$ from the pooled data in Table 2b is +0.017, indicating a mean female salary excess of 1.7 percent. This estimate is different from the 2.6 percent shortfall estimated from direct regression, although both are statistically insignificant from zero. In contrast, the reverse regression estimates $-c^*/b^*$ from Table 2b are 0.123, −0.017 and −0.057, respectively, for Job Levels 1 through 3. These estimates are very close to the direct regression estimates of c as 0.099, −0.017 and −0.073, respectively. The standard errors of the estimates, computed from a result in Theil (1971, p. 383), are also approximately the same for direct and reverse regression. Again, the 5.7 percent mean salary shortfall for females in Job Level 3 is statistically significant, with approximate parity in the other two groups after removing the outlier.

Conway and Roberts (1984, p. 131) note that the direct and reverse regression estimates coincide whenever mean qualifications for males and females are the same, which is more likely to occur in homogeneous job groups. Conflicts in the direct and reverse regression overall assessments of salary discrimination highlight limitations of observational data and may result from the placement of employees into job groups. Separate analyses of the placement and salary decisions with reference to homogeneous job groups helps to isolate these effects. However, the analysis becomes more complex and requires a detailed understanding of the organization's job structure.

We illustrate this with data from a second legal case and the clerical group of 276 employees at Harris Bank considered by Conway and Roberts (1986a). Two distinct job groups are easily identified in the data base, with 85 females and 160 total employees in the first, and 81 females and 116 total employees in the second. Table 3 gives the results from the direct and reverse regression models relating salaries and relevant job qualifications, estimated for the pooled data and each job group separately. Figure 4 graphs the relationship between log salary and the qualification index for the pooled data. The graphs for the separate qualification indexes are omitted, since they are practically identical.

The overall estimate of possible salary discrimination against females is −0.145 and +0.006 from direct and reverse regression, respectively. These estimates change to −0.141 and −0.003, respectively, when the male outlier in Job Group 2 is eliminated from Figure 4. The fitted models within each job group give estimates of potential salary discrimination as −0.108 and −0.201 for Job Groups 1 and 2, respectively, from direct regression and +0.080 and −0.183 from reverse regression. When the male outlier in Figure 4 is removed, the direct regression estimate for Job Group

TABLE 3

Regression summaries that describe the relationship among salary, job qualifications and sex for 276 clerical employees at Harris Bank in two job groups

Variable	Job group 1 Coefficient	Job group 1 t value	Job group 2 Coefficient	Job group 2 t value	Combined jobs Coefficient	Combined jobs t value
			a. Direct regression results			
Constant	9.275	66.64	9.082	58.61	9.228	98.45
Sex	−0.1078	−2.92	−0.2015	−3.80	−0.1478	−5.52
Some college work	0.1123	3.08	0.1133	2.93	0.1146	4.34
College degree	0.4291	6.89	0.2635	3.97	0.3697	8.25
Graduate work	0.5795	4.38	0.4909	1.23	0.5070	4.84
Age of employee	−0.000086	−5.58	0.00041	1.05	−0.00066	−5.06
Work experience	0.0015	3.00	−0.00041	−0.31	0.00138	3.10
Work experience squared	−0.000003	−2.88	−0.000002	−0.43	−0.000003	−3.38
Seniority level	0.00413	2.84	0.00246	1.45	0.00403	3.74
Adjusted R^2	0.597		0.351		0.515	
Standard error	0.1791		0.1711		0.1773	
			b. Reverse regression results			
Constant	4.868	12.91	7.163	18.80	5.752	20.59
Sex	−0.0390	−1.74	0.0442	2.38	−0.0025	−0.16
Log salary	0.485	12.16	0.242	6.01	0.391	13.25
Adjusted R^2	0.591		0.229		0.453	
Standard error	0.1223		0.0819		0.1097	

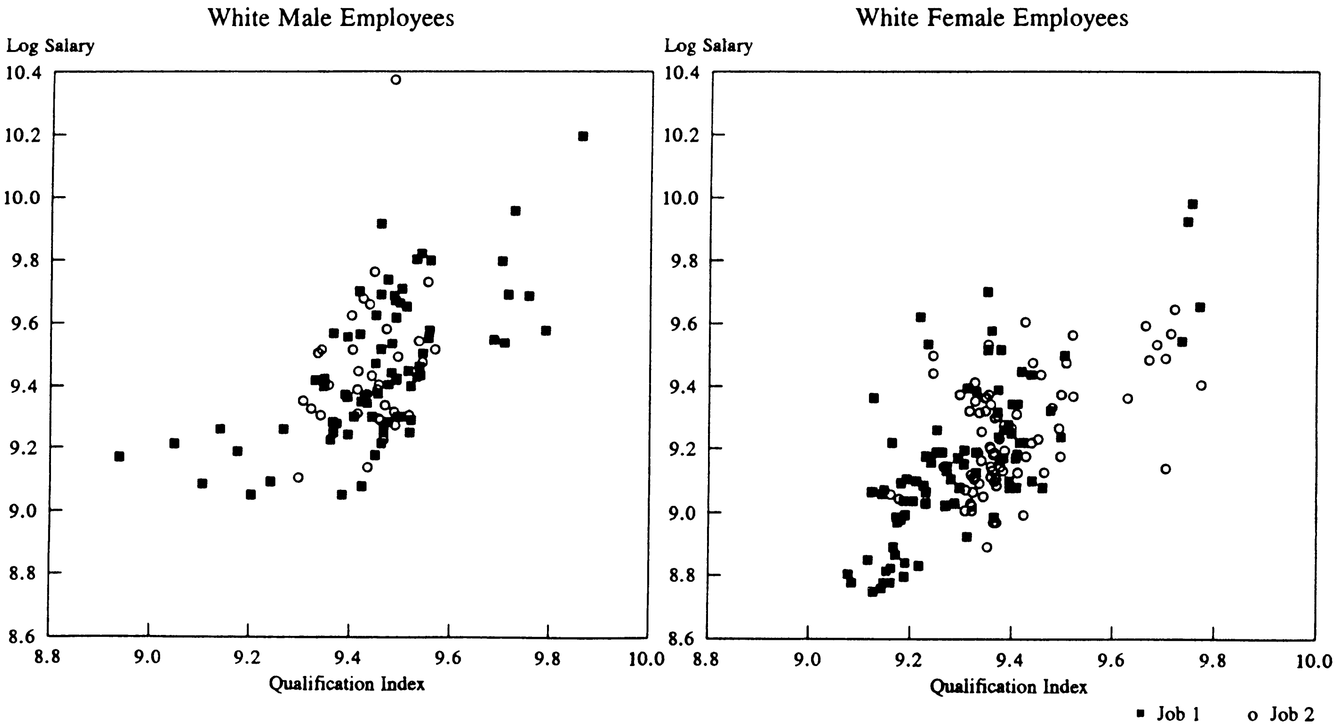

FIG. 4. *Log salary versus qualification index based on pooling* 276 *employees in two job groups at Harris Bank.*

2 is −0.165, whereas the reverse regression estimate is −0.122.

The direct and reverse regression estimates are significantly different for the pooled group and Job Group 1, with much closer agreement for Job Group 2. Figure 4 shows that salaries and qualifications of employees in Job Group 1 surround those of Job Group 2. The figure highlights two distinct subgroups with considerable leverage on the regression results for Job Group 1. Closer examination of the organization's job structure is needed and may suggest further disaggregation of Job Group 1. However, these considerations go beyond the observed data.

The difficult task confronting the statistician is to measure discriminatory departures from the employment process, while taking into account all the legitimate economic factors that may be operating. It is especially difficult to measure these departures when our economic understanding is still evolving about the employment process and the critical factors at work. The difficulties are compounded with the use of observational data, where relevant measures may be miscoded, incomplete, approximated by proxies, or simply not available.

The adversarial pressures in legal cases to obtain precise, numerical summaries can lead to oversimplification of the statistical results. Conclusions may be extended to a broader scope than what is warranted by the data from a scientific analysis. Meier (1986, p. 273) notes that in Title VII cases, "the statistician is strongly tempted to give the definitive rather than a qualified answer to the key questions." Fisher (1986, p. 277) also points out the strong temptation in legal cases to "seek certainty in quantification rather than to study all of the aspects of a truly complex problem." Dempster (1988, p. 192) cautions against the "tendency to rush into probability models without thinking through the scientific bases and contexts that give the models meaning, so that subsequent cranking out of manipulations with the models quickly loses touch with the real questions and required uncertain judgments under analysis."

ROLE OF STATISTICAL EVIDENCE

Throughout her paper, Gray points out the complex difficulties in the statistical assessment of discrimination within an organization and the need for careful analysis. Having directly confronted these issues in my own analysis of different employment data sets, I wholeheartedly endorse this viewpoint. Gray also feels that the courts do not seem to listen to the statistical results. However, when the statistical issues are complex, the results may not be definitive enough to warrant a straightforward judicial decision.

A strong assertion in the article concerns the historical reluctance of courts to find colleges and universities guilty of discrimination. Gray attributes this to judicial bias and advances the hypothesis that judges "identify strongly with the decision makers in colleges and universities . . . and find it hard to believe the evidence of a pattern and practice of discrimination presented by statistics." This is an hypothesis with merit, but one that presupposes the conclusiveness of statistical evidence. Having worked on a number of legal cases involving employment discrimination, I can see several reasons why other aspects of the case may outweigh the statistical evidence.

First, the nature of statistical evidence is largely supportive in legal cases and helps the judge frame a picture of the total evidence. Court opinions hinge on many considerations, including the judge, expertise of the attorneys, testimony of witnesses, documentation of employee policies and numerous legal restrictions, that may override statistical conclusions. I know of at least two cases where fraud and lack of disclosure were dominant concerns and outweighed all other evidence, including the statistical results.

A second reason concerns the adversarial nature of legal proceedings, which often creates tensions and may interfere with statistical conclusions. The court's objective is to get at the truth; but the issues are complex, time is limited and legal matters often restrict the analysis. The statistical conclusions from the opposing sides may be incomplete, exposing two different viewpoints of the data, rather than a more general consensus embracing both. Commenting on truncation bias effects on salary regressions in *Vuyanich v. Republic National Bank* (1981, p. 199), Judge Higginbotham remarked, "Because the controversy here appears to center on an issue on the frontier of econometrics, and there seems, at least to a court unschooled in the intricacies of econometrics, to be genuine conflict between the experts as to the proper approach, we do not decide the issue."

Some attorneys adopt a strategy where the opposing statistical experts simply "cancel each other out." Although this may be appropriate for the legal case, it hardly advances our scientific knowledge about the employment process. But it is a reality of litigation and one that confronts statistical experts for both plaintiffs and defendants.

Finally, there are formidable difficulties in statistical assessments of discrimination from employment data. We do not have a well-developed economic model of the employment process that yields definitive measures of discriminatory departures. Bloom and Killingsworth (1982), Ehrenberg and Smith (1985) and Cain (1986) document the current level of knowledge and evolving economic theory. The intent of Title VII has been to prevent unequal treatment of individuals but also not to infringe on differential employment decisions that arise from job relatedness or "business necessity" (Fiss, 1979). What constitutes "business necessity" is not a simple matter and varies considerably across different organizations. The most recent legislation for the 1991 Civil Rights Act affirmed the importance of "business necessity," while deliberately leaving it vaguely defined and for individual cases to decide.

There is a great need for a well-developed model of the employment process and accurate measures of relevant income, job qualifications and market-related factors. Although we have made substantial progress, the appropriate models and data bases are still evolving. The work to improve inadequate data sources is hardly glamorous and requires painstaking effort. Nonetheless, only with accurate data bases and more complete structural models of the employment process can we hope to achieve better statistical measures of discrimination. In the meantime, we must strive to maintain scientific objectivity and a balanced viewpoint, summarizing the data as best we can and acknowledging areas of uncertainty, to promote informed policy decisions.

ACKNOWLEDGMENTS

This work was partially supported by the School of Business Administration Research Fund, University of Southern California. The author would like to thank Lou Gordon and Ingram Olkin for helpful conversations and suggestions.

Comment

Joseph L. Gastwirth

Because of the author's dual qualifications as a lawyer and statistician, it was interesting to observe that two themes of the article were the view that statistical studies are used to bolster decisions that the policymaker or judge was leaning to anyway and the reluctance of judges to rely solely on statistics. After first commenting on these and other general issues raised by Professor Gray, I will then discuss the regression analyses used in some of the cases cited. As I previously participated in the discussion of Professor

Joseph L. Gastwirth is Professor, Statistics Department, George Washington University, 2201 G Street, N.W., Washington, D.C. 20052.