decisions. In some companies, the decisions also involve multiple raters (which might include a self-evaluation), so that there are multiple measures of assessed productivity. Assessment of possible employment discrimination would undoubtedly be easier with a richer data base that included unbiased assessments of productivity and would suggest new methods of analysis.

Finally, I know of at least one company that uses an explicit direct regression approach for assigning salaries by a computer algorithm to insure fairness 1. It would be interesting to examine the data from such companies over time to help isolate market factors that might affect different job positions. It would also be of interest to evaluate the implications of this approach with respect to fairness, personnel costs, quality of the workforce and market competition.

As more sophisticated information becomes available for employment studies, our methods of analysis should also expand to exploit this information in creative ways. In this way, we will continue to improve our understanding of the employment process and foster the development of more realistic models.

ADDITIONAL REFERENCES

BAZELON, D. L. (1979). Risk and responsibility. *Science* **205** 277–280.

BORJAS, G. J. (1978). Discrimination in HEW: Is the doctor sick or are the patients healthy? *J. Law Economics* **21** 97–110.

BRODIN, M. S. (1982). The standard of causation in the mixed-motive Title VII action: A social policy perspective. *Columbia Law Rev.* **82** 292–326.

CONWAY, D. A. and ROBERTS, H. V. (1984). Rejoinder to comments on "Reverse regression, fairness and employment discrimination." *J. Bus. Econ. Statist.* **2** 126–139.

CONWAY, D. A. and ROBERTS, H. V. (1986a). Regression analyses in employment discrimination cases (with discussion). In *Statistics and the Law* (M. H. DeGroot, S. E. Fienberg and J. B. Kadane, eds.) 107–195. Wiley, New York.

CONWAY, D. A. and ROBERTS, H. V. (1987). Analysis of employment discrimination through homogeneous job groups. Working paper, School of Business, Univ. Southern California.

DEMPSTER, A. P. (1979). Statistical concepts of discrimination. Working paper, Dept. Statistics, Harvard Univ.

FURNISH, H. A. (1982). A path through the maze: Disparate impact and disparate treatment under Title VII of the Civil Rights Act of 1964 after Beazer and Burdine. *Boston College Law Rev.* **23** 419–445.

LOPATKA, K. T. (1977). A 1977 primer on the federal regulation of employment discrimination. *Univ. Illinois Law Forum* **1977** 69–168.

MANISHIN, G. B. (1980). Section 1981: Discriminatory purpose or disproportionate impact. *Columbia Law Rev.* **80** 137–170.

TRIEMANN, D. J. and HARTMANN, H. I., eds. (1981). *Women, Work and Wages: Equal Pay for Jobs of Equal Value.* National Academy Press, Washington.

VUYANICH VS. REPUBLIC NATIONAL BANK (1980). F.E.P. Cases 128 (N.D. Texas).

Comment

Joseph L. Gastwirth

Professor Dempster's interesting and thought-provoking article concerns many scientifically fundamental and socially important issues: the relationship between statistical inference and causality, the proper construction of probabilistic models, the effect of omitted predictor variables (OVs) and errors in variables (EVs) on inferences drawn from models fitted to data and the implications of these topics on the statistical analyses relied on in employment discrimination cases and related public policy decisions. His formulas (12) and (15) concerning the effect of omitted variables on direct and reverse regressions, respectively, add to our understanding of these techniques. From my analysis (Gastwirth, 1984, 1988) of EEO cases, I believe Prof. Dempster may have overestimated the potential for "legal mischief" although I agree with him that a proper statistical analysis involves a careful evaluation of the data and model, including consideration of data errors and omitted factors. Indeed, the effects of OVs stemming from Cornfield's analysis of the possible effect of OVs on the smoking and lung cancer association (see Greenhouse, 1982, for details; Rosenbaum, 1987, for recent developments) and the Bayesian view of missing evidence (Lindley and Eggleston, 1983) have implicitly been used by policy makers and judges. I wish Dempster's use of the Bayesian approach was more explicit so we could compare his conclusions with those reached by the judiciary in actual cases. In particular, the process used by the employer in computing the "posterior expected reward . . . employee" is precisely what is at issue in a disparate treatment case.

Dempster emphasizes the importance of careful causal modeling and considers the randomized clinical trial the most convincing statistical design. However, he also notes that the decisions examined

Joseph L. Gastwirth is Professor of Statistics and Economics, Department of Statistics, George Washington University, Washington, D. C. 20052.

by statisticians should be "exchangeable," which implies that the process generating them should be the same. In EEO cases the very act of someone filing a formal charge with the EEOC typically affects the employer's practices. Almost all views of causality (see Swamy and Von Zur Muehlen (1987) for further references and an alternative view to that of Holland, 1986a) place great weight on considerations of time as does Dempster when he questions reverse regression on the grounds that the employer's productivity assessment cannot be a causal determinant of the job-related characteristics of the employee (courts would accept this if the characteristics concerned education and prior experience but might require evidence of fair treatment if the characteristic concerned training provided by the employer). The failure of some courts to realize the importance of time and the need for exchangeability, in my opinion, has led to far more "legal mischief" than some of the technical issues concerning regression analysis that have dominated the statistical literature (Conway and Roberts, 1983; Robbins and Levin, 1983; Ash, 1986; McCabe, 1986; Peterson, 1986). These articles complement Dempster's results by showing that the group coefficient also may contain a component due to regression to the mean effects which arise for a variety of reasons (EVs). A recent article, published after Dempster's was accepted, by Schafer (1987) presents a diagnostic procedure to evaluate the EV problem to help us assess whether the reliability of the measured covariate is so low that it could seriously affect our ultimate inference.

At the outset I should say that I prefer the methods of matching and stratification used in epidemiology (Breslow and Day, 1980; Schlesselman, 1982) and survival analysis (Miller, 1981), possibly incorporating regression to control for a few other variables, to fitting one regression model with a large number of covariates to pay data covering widely divergent job categories. First, the relative importance of different productivity characteristics may depend on the job. My second reason was better stated by Judge Posner in *Riorden v. Kempiners* 44 FEP Cases 1355 (7th Cir. 1987) when the court remanded a sex discrimination case for further hearings as the trial judge did not allow plaintiff to compare her pay to that of several comparable males. With respect to disparate treatment cases he writes:

> "Proof of such discrimination is always difficult. Defendants of even minimal sophistication will neither admit discriminatory animus nor leave a paper trail demonstrating it; and because most employment decisions involve an element of discretion, alternative hypotheses (including that of simple mistake) will always be possible and often plausible. Only the very best workers are completely satisfactory, and they are not likely to be discriminated against—the cost of discrimination is too great. The law tries to protect average and even below-average workers against being treated more harshly than would be the case if they were of a different race, sex, religion, or national origin, but it has difficulty achieving this goal because it is so easy to concoct a plausible reason for not hiring, or firing, or failing to promote, or denying a pay raise to, a worker who is not superlative. A plaintiff's ability to prove discrimination indirectly, circumstantially, must not be crippled by evidentiary rulings that keep out probative evidence because of crabbed notions of relevance or excessive mistrust of juries."

Judge Posner's statement has two implications for statisticians involved in the EEO area. a) The simple regression model assuming a constant shift α as in equation (4) is unrealistic. b) The information concerning discrimination is most likely to be found in the discordant pairs (if we have a matched study). Virtually all highly (low) qualified employees will (not) be promoted. It is in the middle range of qualification levels that discrimination is most likely to appear. Of course, EVs and OVs can have a deleterious impact on epidemiological and survival methods. (These are described in the proceedings of a recent conference at NIH with papers by Carroll, Chen, Fuller, Gleser, Ware, Whittemore and others which will appear in *Statistics in Medicine*, and in Samuels (1986).)

Dempster provides results on the bias of both types of regressions for models of the form

$$\text{(a)} \qquad Y = G\alpha^* + X^*\beta^* + \text{error},$$

which are estimated on data enabling us to fit

$$\text{(b)} \qquad Y = G\alpha + X\beta + \text{error},$$

where $X \subset X^*$. Rather than beginning with the linear model one might consider the males ($G = 1$) alone and note that the best predictor $E_1[Y \mid X^*]$ is being approximated by $E_1[Y \mid X]$. Using $E_0[Y \mid X^*]$ and $E_0[Y \mid X]$ for the females ($G = 0$) and assuming an *additive* group effect, the analysis of Cochran (1968) for the bias due to EVs leads to

$$\text{(c)} \qquad \{E_1[Y \mid X^*] - E_1[Y \mid X]\} - \{E_0[Y \mid X^*] - E_0[Y \mid X]\},$$

which reduces to $\mu_{M_2} - \mu_{F_2}$ in Dempster's equation (12), i.e.,

$$\mu_{M_2} = X^*\beta^* - X\beta \quad \text{for males}$$

and

$$\mu_{F_2} = X^*\beta^* - X\beta \quad \text{for females.}$$

Formula (c) combines the effects of OVs and EVs as some of the X's can be regarded as $X_i = X_i^* + e$. Dempster notes that when males ($G = 1$) have higher productivity values on the omitted variables so that $\mu_{M_2} - \mu_{F_2} > 0$ direct regression yields an overestimate of the coefficient reflecting discrimination. In interpreting this result, however, we must realize that (12) is expressed in terms of the orthogonalized X's, i.e., $\mu_{M_2} - \mu_{F_2}$ is the difference between the groups with respect to the omitted variables *after* their correlation with the variables (X) in the model has been accounted for. Thus, the discussion between formulas (5) and (6) may need amplification as much empirical evidence would concern X and X^* rather than the "remaining effects" after their correlation was accounted for. Similar considerations apply to his formula (15) concerning the bias in reverse regression and confirm his implication that reverse regression underestimates the discrimination effect, especially if one assumes that the measured X variables are the major determinants of productivity.

REMARKS. 1) The review article by Hocking (1976) also presents a computationally convenient formula for the bias in terms of the values of the omitted variables and discusses other consequences of OVs which include higher estimated residual variance, implying lower R^2s, and generally higher variability of the estimated coefficients. These tend to somewhat offset the bias problem in EEO cases as the increase makes rejection of the null hypothesis less likely and regressions with low to moderate R^2s typically receive less weight in the judicial process (Baldus and Cole, 1987). The OV problem has also been addressed in the econometric literature (Theil, 1971, pages 548–556). 2) Formula (c) reminds us that we should assess whether the forms of the prediction equations for the two groups are similar.

I agree with Dempster's ending remarks concerning the need for more realistic models of the wage determination process. They should include interactions, nonlognormal data (Gastwirth and Smith, 1972), feedback mechanisms (Rubinfeld, 1985), etc. Indeed, the assumption that covariance matrices of the X variates are the same in both groups does not agree with empiric evidence (Smith, 1972) from the Harvard reanalysis of the Coleman report. I will leave further detailed technical commentary on the various models to discussants who are econometricians. Nevertheless, it should be noted that the error or disturbance term in econometric models is justified by the principle of parsimony (Stone, 1981) which is a modern version of Occam's razor. The disturbance term includes chance events plus factors that have a *minor* influence as well as human indeterminacy (Stewart and Wallis, 1981). I believe Goldberger's (1984) paper was less concerned with modeling wages than with the identification problem that occurs when both direct and reverse regressions are fitted to the same data. When the reduced form equations are solved, the error term in the two regressions are correlated and the error terms are also correlated with the predictor (independent) variables. Thus, the computation of the conditional expectations and their variances becomes more involved. Hopefully, other discussants and Dempster's rejoinder will clarify this point. In addition, the role of seniority in wage determination and other aspects of labor relations needs to be considered more carefully. Abraham and Medhoff (1984) show that seniority plays a major role in layoff decisions in nonunion as well as unionized firms. They cite other references which indicate that seniority is part of an implied contract between employees and employers, allowing more senior employees to be paid more than they merit while younger ones receive less. The seniority system reduces turnover and training costs and may also contribute to a more amiable workplace as all employees know the system. Thus, the participants in the labor market may have agreed that it is preferable to use seniority to decide close promotion decisions rather than leave them to more subjective evaluations of hard to measure variables. In a sense, they may be using seniority in lieu of some of the OVs. Further, empirical research in this area along the lines of Abraham and Medhoff (1985) and Mills (1985) is needed to quantify this phenomenon.

Before discussing actual legal applications, it is helpful to consider an example illustrating the use of nonrandomized studies in epidemiology and public health as the notions of causality in law (Hart and Honoré, 1985) are similar in spirit to those in epidemiology (Mausner, Bahn and Kramer, 1985). In particular, causality must be related to biological plausibility and other factors should *not* be able to explain the phenomenon. Good causal modeling in social science (Cooley, 1978) and econometrics (Fisher, 1980) also relies on models based on empirical and theoretical evidence from the subject matter area.

In late 1982, after a case control study in Ohio confirmed a strong association (references are given in Gastwirth, 1988) between Reyes' syndrome and a child's ingestion of salicylate compounds during a prior bout with flu or chicken pox, the FDA wanted to require a warning label on aspirin and related drugs. The industry raised a number of questions concerning the validity of the previous studies. As a result, a new study was planned which involved a one-year pilot study and a two-year main study. As critics of the earlier studies had suggested that parents of cases would be under more stress than those of controls,

control groups of children brought to emergency rooms or who were hospitalized for other reasons were used in addition to school and community controls. A question concerning the parents' knowledge of any suspected link between drugs and childhood diseases was added to assess whether people who could not remember what drugs were administered might rely on this knowledge. Other verification procedures, e.g., pictures of brands of child medications, were used to minimize respondent error. The results of the pilot study, conducted in the 1983–1984 flu season were announced in January 1985. They showed a relative risk between salicylate use and Reyes' syndrome of 10 or more in *all* four control groups and the industry launched a publicity campaign in response to the government pressure. The results of the pilot study appeared in the *New England Journal of Medicine* in late 1985 and in April 1986 the *Journal* published letters from critics of the study and a rejoinder from the authors of the study. Everyone seemed to ignore the fact that the incidence of Reyes' syndrome for 1985 declined by over 50%, presumably in response to the publicity given in the media as well as notices posted in drug stores. The 1985 data was issued by CDC in February 1986.

Essentially, the government relied on Cornfield's result to conclude that it was *extremely unlikely* that another factor that would have a 10-fold increased risk of causing Reyes' syndrome and be 10 times more prevalent in the case group than in all four control groups could exist. Furthermore, a logical extension of the missing evidence principle, used in courts and given a Bayesian justification by Lindley and Eggleston (1983), also applies; if a party has data available but does not introduce it, then an *adverse inference* is proper. The industry could have developed a study relating the geographic location of Reyes' cases to sales data. Had the Reyes' cases occurred primarily in areas where nonsalicylate medicines were dominant that might have cast doubt on the results of the case control studies. Although such ecological studies are not considered as reliable or as sound as case control studies, the fact that the critics offered *no* statistical data should be considered in evaluating their criticisms. Because one can speculate endlessly about potential OVs and EVs, I believe that such criticisms should be supported by data before policy makers take them seriously. *Note*: It should be mentioned that I was a statistical consultant to OMB when the pilot and main Reyes' studies were approved.

Statistical data is used to examine equal pay for equal work, the possible disparate impact of a job requirement (if a requirement excludes a disproportionate fraction of minority applicants, then it must be shown to be job-related) and disparate treatment (the unequal treatment of similarly qualified persons due to their race, sex, religion or national origin). Justice Powell's opinion in *Texas Dept. of Community Affairs v. Burdine* 450 U.S. 248 (1981) describes the order of proof in disparate treatment cases as follows: "First, the plaintiff has the burden of proving by a preponderance of the evidence a *prima facie* case of discrimination. Second, if the plaintiff succeeds, . . . , the burden shifts to the defendant to articulate some legitimate, nondiscriminatory reason for the employee's rejection . . . Third, should the defendant carry this burden, the plaintiff must have the opportunity to prove by a preponderance of the evidence that the legitimate reasons offered by the defendant were not its true reasons but were a pretext for discrimination." The opinion notes that the burden of proof (not production of evidence) of intentional discrimination "lies with the plaintiff."

In *Bazemore v. Friday* 106 U.S. 3000 (1986) the North Carolina Extension Service was charged with racial discrimination in pay. The original claim was filed in 1971 and the United States Department of Justice intervened in April 1972. Plaintiffs analyzed salaries for the years 1974, 1975 and 1981 with race and education, tenure and job title as the predictor variables (X). In 1974 and 1975, blacks averaged a statistically significant $350 a year less than comparable whites. By 1981 the yearly difference had declined to a nonsignificant figure. The district court discounted this regression analysis because it did not account for county-wide differences in pay or pay raises and because pre-Act discrimination was still affecting later salaries. The district and appeals courts thought that a regression must include *all* measurable variables thought to affect salary. Another issue in the case was the fairness of the job evaluation system as persons in the lowest quartile did not receive a merit raise. The portion of the Supreme Court's opinion concerning regression analysis was unanimous in stating that a regression analysis need not contain *all* variables but should contain the *major* factors. The Court noted that legal pre-Act discrimination in pay was *not* an excuse for illegal pay discrimination later and that differences between counties could not explain the pay differential as plaintiffs showed that blacks were *not* disproportionately located in the low paying counties although they were not equally distributed among all counties. Plaintiffs also noted that including the quartile ranking in the regressions *increased* the black-white differential. They also compared similarly situated (county, job) black and white employees and showed that blacks generally received the lower salary.

Before examining the decision in light of Dempster's concerns, it should be mentioned that the appellate

opinion, 751 F.2d 662 (4th Cir. 1984) analyzed the quartile rankings by testing the 2×2 tables *separately* for each of the five districts and found that the discrepancy in only one was close to statistical significance. As I show elsewhere (Gastwirth, 1988) had the court used the Mantel-Haenszel (1959) refinement of Cochran's (1954) procedure for combining 2×2 tables, they would have obtained a significant difference. As the Extension Service employed a total of about 90 blacks and 350 whites, by insisting on a significant difference in each district the Fourth circuit made it virtually impossible for plaintiffs to demonstrate discrimination in job evaluation.

In order to assess whether omitted variables and errors in variables problems could explain the significant race coefficient we recall that such a variable must be unequally distributed in the two groups (Carroll, Gallo and Gleser, 1985). We already have noted that location (county) does not meet this condition. Of plaintiffs three variables only education might be subject to serious mismeasurement. As Dempster states in many regression models the role of the variables is often not clearly specified. We will assume that education measures knowledge obtained prior to the job as well as indicating motivation (ability to stick with a program of studies). As Professor Dempster points out, here is where a judge and the reader must make a subjective assessment.

To do this we use the Lindley-Eggleston view of missing evidence. If motivation were the OV or EV, it should be manifest in some measure of work product, e.g., willingness to work overtime, turnover and/or absenteeism data. No such data was offered into evidence. If prior education was mismeasured, we need to extend the logic underlying missing evidence by inquiring as to whether it was obtainable or not. Because grades in school, quality of school, etc. are readily available to an employer (recall most hires are at entry level jobs and transcripts are *free* to an employer) if such data is not obtained it is reasonable to conclude that the employer decided that this information is not relevant or helpful in estimating Y^*. Hence, it is unlikely that an EV problem in education would explain the significant race effect in *Bazemore*.

REMARKS. 1) In *Bazemore* the pay of *all* employees was considered. Most employees had served for a number of years, presumably performing satisfactorily, which should mitigate the EV and OV problems. Further, mathematical models might be developed to incorporate this natural truncation effect, i.e., those employees whose qualifications were highly overestimated may well be terminated or transferred to a more suitable job. 2) The previous argument suggests that EV and OV problems should be important in considering pay at hire, early terminations and time to promotion (especially the first one). In *Lewis v. NLRB* 36 FEP Cases 1388 (5th Cir. 1985), the court allowed the defendant to rebut a statistically significant difference in time to promotion by showing that deleting a few "slow" minority members from the data eliminated the difference and explaining why each "slow" promotion occurred. Hence, in their own way, courts do account for EV and OV problems. 3) A recent article by Koenig (1987) contrasts the casual hiring procedures used by American firms with those used by Japanese firms who give applicants several tests and interviews. As Lord (1960) demonstrated, one way to reduce the EV problem is to use independent estimates of the skill. Although several tests or grades in school may not be totally independent, together they would reduce the EV problem. Furthermore, by obtaining more accurate values of the objective predictors, the employer would obtain better estimates of $E(Y)$. In the context of an employment discrimination case one can question the logic of allowing an employer who does not use available accurate predictors to claim that EVs and OVs overestimate the effect of G, especially when the marketplace has demonstrated that other firms (the Japanese) using the more refined predictors do better. After all, the less predictive power the "objectively measured" variables possess, the more leeway an employer has to place greater weight on subjective factors which are more easily affected by discrimination.

No discussion of statistics in EEO cases can be considered complete without consideration of Judge P. E. Higginbotham's opinion in *Vuyanich v. RNB* 505 F.Supp. 224 (1980) vacated on other grounds 723 F.2d 1195 (5th Cir. 1984). Although we may find some technical issues to quibble with, it is clear that the judge is carefully thinking about scientific evidence and its role in these cases. The opinion notes that there is *no* discrimination when a highly productive white receives more pay than a less productive minority worker. However, courts scrutinize productivity measures as "absent accurate measure they contain great potential for masking differentials actually based on race or sex." The human capital theory of earnings (the theoretical basis for wage equations) and the basic concepts of multiple regression are then presented. In particular, a section of the opinion is devoted to "What can go wrong?" It discusses specification errors, including the importance of using the correct functional form, the importance of the assumption that the errors be independent of the predictors, multicollinearity and omitted variables. The opinion notes that an omitted variable needs to be related to group status in order to have an effect and that when an

employer shows a systematic "bias" in plaintiff's model or vice versa, the model needs to be "sensitive" to it, i.e., the problems must have an effect large enough to impact on the validity of the model's finding of the presence or absence of discrimination. The opinion notes several *serious* flaws with plaintiff's model, in particular the use of age minus 18 instead of experience by plaintiffs was shown to create a serious underestimate of the experience and a resulting overestimate of discrimination as females had two years less experience than males of the same age. As the opinion is over 150 pages long I will only discuss its treatment of omitted variables at pages 310–314, where the potential effect of IQ, analytic ability, creativity, childhood environmental influences, etc. are discussed. The opinion notes, as does Dempster, that in society as a whole, blacks and women may possess a lower amount of some of these characteristics. However, one cannot conclude that the bank's employees reflect these societal differences. Indeed, to the extent that the bank selected individuals with high values of X factors, they are not a random sample of society. Hence, in an actual case, the employer needs to make more than a general assertion; relevant data should be produced. (In *Vuyanich*, the defendant did conduct a small survey of their employees' willingness to move, etc., but did not stratify the results on the basis of job level.) This approach is reasonable from a legal view as virtually no judge would give credence to a plaintiff who said they did not accept an offer of training from an employer because of previous societal discrimination so this particular firm would also discriminate, regardless of qualification. The requirement that data specific to the employment practices at issue in the actual case be examined is also good statistical practice. Because cases are not a representative sample of the nation's employment practices, it is unsound to assume that the relevant characteristics of the employers or employees involved are those of a random sample of the nation.

REMARKS. 1) The results of Parcel and Mueller (1983), who assessed the effect of including social-psychological factors on earnings and found they did not change the discrimination measures enough to reduce them to insignificance may also help courts assess the effect of OVs. On the other hand, most of the data bases used by Parcel and Mueller measured education by years of schooling, as is done in the census surveys. Without special areas of study, school quality and measures of actual achievement EV and OV problems clearly remain. Thus, the results of Dempster and the research on EVs and OVs have more immediate relevance to the many studies in the socioeconomic literature which rely on census public-use tapes, then in EEO cases. 2) More research on the numerical value of (12) for reasonable models and covariance structures would be helpful. Some very detailed data sets such as the one used by Darland, Dawkins, Lovasich, Scott, Sherman and Whipple (1973) could be examined by deleting several variables which may be omitted from ordinary salary studies of professors. 3) The wisdom of Judge Higginbotham's requiring more than vague general findings appears to be borne out in *Watson v. Ft. Worth Bank* 798 F.2d 791 (5th Cir. 1986), currently up for review by the Supreme Court. According to plaintiff's brief, not only did their pay and grade level regressions show discrimination, blacks had higher average values on most of the productivity predictors. In his dissent Judge Goldberg lays out a solid statistical proof of hiring discrimination, although its legal relevance for a promotion case will be determined by the Supreme Court. Because the sample size available to study minority promotions is smaller when they have suffered hiring discrimination, the Supreme Court may need to relax its adherence to the 0.05 significance level in order to assess the promotion and job evaluation data in the case. Otherwise, an employer who discriminates in hiring might be statistically immunized from other discriminatory practices. 4) Recent research on quit rates (Blau and Kahn, 1981) indicating that minority groups do not have higher turnover rates when relevant covariates (occupation, etc.) are considered, also supports the view that data on the actual employees needs to be examined. 5) The *Vuyanich* opinion at page 356 notes that courts resolve doubts concerning job qualifications at the *prima facie* stage in favor of plaintiffs as undercorrection only gives the defendant the burden of explaining group differences by more appropriate qualification measures. Thus, courts are not allowing "legal mischief." They have weighed the costs and benefits of various alternative approaches to the presentation of evidence and have decided on a reasonable one. The remarks toward the end of the article indicate that Dempster has thought about these various trade offs and I hope he develops these ideas in a future article. 6) An alternative to a regression analysis to all employees is to use different regressions for different occupations and then combine them to assess whether an overall pattern of disparate treatment exists. Indeed, this procedure is suggested in *Vuyanich* at page 304. As a result of stratification, the groups should be reasonably well balanced with respect to the X values, hence formula (12) and formula (9.13) of Cochran (1968) suggest that EV and OV effects should be small. However, there may be truncation in range and sample selection bias (Heckman, 1979) problems if there truly is one regression. Judge Higginbotham decided that the court was

unable to resolve these issues, *Vuyanich* at page 314, but the problem is worth further study. This approach was used by the defendant in *Eastland v. TVA* 704 F.2d 613 (11th Cir. 1983) to rebut plaintiff's big regression. Indeed, the defendant analyzed five classes of employees and found a significant race effect in only one. The court apparently did not examine all five together to see whether there was a pattern of blacks receiving less pay. This failure to recombine the results of stratified analyses occurs quite frequently and generally disadvantages plaintiffs rather than defendants although we cannot tell from the opinion whether the error was serious enough to affect the *Eastland* case.

I believe that Dempster is correct in drawing our attention to the way our subjective evaluation of the model invariably enters our assessment of a real data analysis, so his results are useful in many applications. In the context of EEO cases, my experience suggests that several aspects of his model may need further development. First, his claim that there is no restriction on the variables in X^* conflicts with legal doctrine (Finkelstein, 1980; *Vuyanich* at page 277) because some of these can reflect employer discrimination. For instance, if a university included teaching evaluations, research record and service on college committees in pay and promotion decisions (but the Dean rarely appointed members of group $G = 0$ to these committees), the inclusion of committee service would obscure rather than illuminate the issue. Secondly, Dempster assumes a sequential set of approximations. As true productivity is assumed unknown, the employer is limited to equation (3*) and the statistician uses

$$E(Y) = G\alpha + X\beta.$$

It is not clear why defendant's statistician should lack access to X^*. Similarly, plaintiff's expert would have access to all the data they obtained in discovery. When X does not equal X^*, we should apply the missing evidence principle—the data not supplied to the expert is likely to be harmful to their side. Different data sets have been used by the parties in a number of cases and both plaintiffs and defendants have not informed their experts of all available data. When this occurs the assumption that the error terms have mean 0 is severely violated and it is reasonable to assume that the errors are related to G in a way that will bias the estimated coefficient in favor of the side for whom the expert performed the analysis.

Elsewhere (Gastwirth, 1988) the statistical data from a number of EEO cases is described. Plaintiffs do not even establish a *prima facie* case with a statistical presentation unless all objectively measurable factors available to them are included or shown to be unreliable. For example, in *Coble v. Hot Springs School District No. 6*, 682 F.2d 721 (8th Cir. 1982) plaintiffs were told they should have introduced a multiple regression, not simply compared the pay of the sexes one factor at a time. Defendants often rebut *prima facie* cases by showing that the statistical assumptions needed for a model's validity are not satisfied by the data, e.g., *Sobel v. Yeshiva University* 566 F.Supp. 1166 (S.D. N.Y. 1983). They fail to rebut plaintiffs' case when they cannot demonstrate that the flaws are sufficiently severe to affect the inference of discrimination, e.g., *Vuyanich* and *Craik v. Minnesota State Univ. Bd.* 731 F.2d 465 (8th Cir. 1985). The recent results of Schafer (1987) and Dempster's formula (12) should aid courts in assessing the possible impact of EVs and OVs.

Originally I planned to conclude with a discussion of the major errors courts make in assessing data: mixing postcharge data with precharge data (Gastwirth, 1984, 1988) thereby enabling a firm to obscure earlier discrimination, failing to consider power (Fienberg and Straf, 1982; Gastwirth and Wang, 1988), not using proper stratification and recombination methods, and with a discussion of the need for methods which build and test models on the same data (Cohen and Sackrowitz, 1987) as legal and exchangeability requirements often limit the relevant data to a few years prior to the charge. However, the recent news story about Howard University's suit against the National Collegiate Athletic Association for a place in the football tournament for the top 16 schools in its category may be more interesting to many readers. According to Huff and Greenberger (1987), Judge Penn denied Howard's request to delay the tournament but will conduct further proceedings. Howard finished second in its division with a 9–1 record. The NCAA asserted that its schedule included several weak teams (including several primarily black schools—the only ones it could play against in the old days). The article notes:

> "In yesterday morning's hearing, Howard attorneys focused on one particular issue. Howard entered the final weekend of the season tied for No. 20 in the rankings with North Texas State. Howard defeated then No. 14-ranked Delaware State, 12–7, while North Texas State was a 10–5 winner over Louisiana Tech, which finished 3–8."
>
> "Howard was ranked No. 18 in the most recent poll. North Texas State was ranked No. 16. NCAA attorney Donald Bucklin explained North Texas State was moved up higher in the poll due to its overall strength of schedule. Penn asked Bucklin several times, if strength of schedule was

so important, why was North Texas State not ranked higher the previous week?"

"Bucklin continually answered there were a number of factors and it was the organization's contention that the weekly rankings are "only for public interest" and not based on the same criteria as the tournament selections."

This case again shows that issues of unfairness are most likely to occur at the borderline. Hence, the judge is primarily concerned with deciding how the University's team was treated relative to the most comparable ones.

My comments support the author's reservations concerning the applicability of the model described by equation (4) to data (especially data aggregated over many types of job) in EEO cases. The article is important as it should stimulate our profession to continue research on the many problems arising in this area.

ACKNOWLEDGMENT

This work was partially supported by a National Science Foundation grant.

ADDITIONAL REFERENCES

ABRAHAM, K. G. and MEDHOFF, J. L. (1984). Length of service and layoffs in union and nonunion work groups. *Ind. Labor Rel. Rev.* **38** 87–97.

ABRAHAM, K. G. and MEDHOFF, J. L. (1985). Length of service and promotions in union and nonunion work groups. *Ind. Labor Rel. Rev.* **38** 408–420.

ASH, A. (1986). The perverse logic of reverse regression. In *Statistical Methods in Discrimination Litigation* (D. H. Kaye and M. H. Aickin, eds.). Dekker, New York.

BALDUS, D. and COLE, J. W. L. (1987). Supplement to *Statistical Proof of Discrimination.* Shepard's/McGraw-Hill, New York.

BLAU, F. D. and KAHN, L. M. (1981). Race and sex differences in quits by young workers. *Ind. Labor Rel. Rev.* **34** 563–577.

BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research, Volume 1: The Analysis of Case Control Studies.* International Agency for Cancer Research, Lyon.

CARROLL, R. J., GALLO, P. and GLESER, L. J. (1985). Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. *J. Amer. Statist. Assoc.* **80** 929–932.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ square tests. *Biometrics* **10** 417–451.

COCHRAN, W. G. (1968). Errors of measurement in statistics. *Technometrics* **10** 637–666.

COHEN, A. and SACKROWITZ, H. B. (1987). An approach to inference following model selection with applications to transformation-based and adaptive inference. *J. Amer. Statist. Assoc.* **82** 1123–1130.

COOLEY, W. W. (1978). Explanatory research studies. *Ed. Res.* **7** 9–15.

DARLAND, M. G., DAWKINS, S. M., LOVASICH, J. L., SCOTT, E. L., SHERMAN, M. E. and WHIPPLE, J. E. (1973). Application of multivariate regression to studies of salary differences between men and women faculty. *ASA Proc.* 120–132. Amer. Statist. Assoc.

FIENBERG, S. E. and STRAF, M. L. (1982). Statistical assessments as evidence. *J. Roy. Statist. Soc. Ser. A* **145** 410–421.

FINKELSTEIN, M. O. (1980). The judicial reception of multiple regression in race and sex discrimination cases. *Columbia Law Rev.* **80** 737–754.

FISHER, F. M. (1980). Multiple regression in legal proceedings. *Columbia Law Rev.* **80** 702–736.

GASTWIRTH, J. L. (1984). Statistical methods for analyzing claims of employment discrimination. *Ind. Labor Rel. Rev.* **38** 75–86.

GASTWIRTH, J. L. (1988). *Statistical Reasoning in Law and Public Policy.* Academic, Boston. To appear.

GASTWIRTH, J. L. and SMITH, J. T. (1972). A new goodness of fit test. *ASA Proc. Bus. Econ. Statist. Sec.* 320–321. Amer. Statist. Assoc.

GASTWIRTH, J. L. and WANG, J. L. (1987). Nonparametric tests in small unbalanced samples. *Canad. J. Statist.* To appear.

GREENHOUSE, S. W. (1982). A tribute to Jerome Cornfield. *Biometrics* **38** (suppl.) 3–7.

HART, H. L. A. and HONORÉ, T. (1985). *Causation in the Law,* 2nd ed. Oxford Univ. Press, New York.

HECKMAN, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47** 153–161.

HOCKING, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32** 1–49.

HUFF, D. and GREENBERGER, N. H. (1987). *The Washington Post.* Nov. 28 D1.

KOENIG, R. (1987). Exacting employer: Toyota takes pains and time, filling jobs at its Kentucky plant. *Wall Street J.* Dec. 1.

LINDLEY, D. V. and EGGLESTON, R. (1983). The problem of missing evidence. *Law Quart. Rev.* **99** 86–99.

LORD, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. *J. Amer. Statist. Assoc.* **55** 307–321.

MANTEL, N. and HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22** 719–748.

MAUSNER, J. S., BAHN, A. and KRAMER, S. (1985). *Epidemiology: An Introductory Text,* 2nd ed. Saunders, Philadelphia.

MCCABE, G. P. (1986). Regression analysis in discrimination cases. In *Statistical Methods in Discrimination Litigation* (D. H. Kaye and M. H. Aicken, eds.). Dekker, New York.

MILLER, R. G., JR. (1981). *Survival Analysis.* Wiley, New York.

MILLS, D. Q. (1985). Seniority versus ability in promotion decisions. *Ind. Labor Rel. Rev.* **38** 421–425.

PARCEL, T. L. and MUELLER, C. W. (1983). *Ascription and Labor Markets: Race and Sex Differences in Earnings.* Academic, New York.

PETERSON, D. W. (1986). Measurement error and regression analysis in employment cases. In *Statistical Methods in Discrimination Litigation* (D. H. Kaye and M. H. Aicken, eds.). Dekker, New York.

ROBBINS, H. and LEVIN, B. (1983). A note on the "underadjustment" phenomenon. *Statist. Probab. Lett.* **1** 137–139.

ROSENBAUM, P. (1987). The role of a second control group in an observational study (with discussion). *Statist. Sci.* **2** 292–316.

RUBINFELD, D. (1985). Econometrics in the courtroom. *Columbia Law Rev.* **85** 1048–1097.

SAMUELS, M. L. (1986). Use of analysis of covariance in clinical trials: A clarification. *Controlled Clin. Trials* **7** 325–329.

SCHAFER, D. W. (1987). Measurement-error diagnostics and the sex discrimination problem. *J. Bus. Econ. Statist.* **5** 529–538.

SCHLESSELMAN, J. J. (1982). *Case Control Studies.* Oxford Univ. Press, New York.

SMITH, M. S. (1972). Equality of educational opportunity: The basic findings reconsidered. In *On Equality of Educational Opportunity* (F. Mosteller and D. P. Moynihan, eds.). Random House, New York.

STEWART, M. B. and WALLIS, J. F. (1981). *Introductory Econometrics*, 2nd ed. Blackwell, Oxford.
STONE, C. J. (1981). Admissible selection of an accurate and parsimonious normal linear regression model. *Ann. Statist.* **9** 475–485.
SWAMY, P. A. V. B. and VON ZUR MUEHLEN, P. (1987). Further thoughts on causality with econometric models. Special Study Paper. Federal Reserve Board, Washington.
THEIL, H. (1971). *Principles of Econometrics*. Wiley, New York.
WOLINS, L. (1978). Sex differentials in salaries: Faults in analysis of covariance. *Science* **200** 723.

Comment

Gail Blattenberger

I am pleased to see attention directed to the theoretical principles underlying the statistical reasoning process in legal proceedings. Arthur Dempster has raised important questions regarding statistical analysis in employment discrimination cases. Too often statistical advice has been given in legal cases purporting to estimate a discrimination effect without explicit statement or understanding of what is being estimated. Here I offer comment on statistical evidence of discrimination in the legal context. I focus on interpretational questions rather than specific model specification issues such as reverse regression.

Continued active controversies over the meaning of probability render a universally accepted standard definition impossible. Nonetheless, the understanding of probabilistic language has implications for the interpretation of the evidential content of the analysis. Explicit statement of probabilistic modeling assumptions becomes necessary for communication not only between statistical experts and lawyers, but even among statisticians. In this I agree with Dempster.

My own position on probability adopts basically the personal measure of uncertainty meaning that Dempster advocates in this context, but my position is perhaps more extreme in this direction than his. Following de Finetti, I view probabilities as representations of uncertain opinions about the value of unknown but observable quantities. In this context it becomes important to specify whose opinions the probabilistic structure represents and under what circumstances. Probability for me is not a physical property, and estimation of unknown and inherently unmeasurable constructs lacks substance. This has relevance for the specification and interpretation of the probabilistic model.

Within this probabilistic perspective a linear model of the form specified in Dempster's equations (1) and (2) might represent a linear belief structure of an analyst. This model is specified:

$$Y_i = G_i\alpha + X_i\beta + \theta_i. \tag{1}$$

I will play the role of the analyst. From my position equation (1) denotes the process by which I base my opinion about the measurable value of Y_i, employee i's salary, given the measurable values of G_i, the ith employee's gender, and given X_i, a vector of other measured characteristics of the ith employee. Although this equation has the same form as the standard model in Dempster's discussion, the interpretation is different.

Dempster expands his model to include information known by the employer but not the statistician, X^*, and a more comprehensive vector of characteristics, X^{**}, needed to determine the employee's "true worth," Y^{**}. Undoubtedly, the employer does use information available to him, but unknown to me in setting salaries; it is also true that the employer may provide nonmonetary fringe benefits which are unknown to me. I could incorporate recognition of this into my belief structure. I would question, however, the role of unmeasurable or unmeasured characteristics, X^{**}, the existence of the unmeasured and inherently unmeasurable, Y^{**}, and its expected value, Y^*. Dempster admits that the realism of these concepts is questionable, but he assumes that they exist. He proceeds to develop a model based on these concepts and examines its implications for assessing discrimination.

Economists, adhering to the human capital approach, have used the idea of an individual's marginal productivity to indicate the "true worth" of that employee. I have argued elsewhere, Blattenberger and Michelson (1984), that individual marginal productivity is not an intrinsic property possessed by an individual. It is inherently unmeasurable. Dempster does not use the term marginal productivity, but the same arguments are applicable to "true worth." I personally have had recent experience with this issue. In response to state budget cuts, I have participated in a committee

Gail Blattenberger is Associate Professor, Department of Economics, University of Utah, Salt Lake City, Utah 84112.