Research in Higher Education, Vol. 34, No. 1, 1993

FACULTY SALARY EQUITY: Issues in Regression Model Selection

Nelle Moore

Faculty salary equity is a hot political issue that may have severe legal, monetary, and human consequences. It is also an issue that often requires the use of sophisticated statistical techniques for the determination of inequity. The purpose of this paper is to identify the areas in which human judgment must be made in order to conduct a statistical analysis of salary equity and to provide some informed guidelines for making those judgments. The direction and magnitude of the final results are contingent on the way these statistical decisions are made. Therefore, careful consideration of these issues is essential for conducting a fair and defensible salary equity study. This paper will provide a framework based on four decision elements and four fields of study as the basis for establishing criteria for selecting an appropriate salary equity model. Through this discussion, the author hopes to bring a broader perspective and, if not objectivity, then ethical fairness to the process of designing salary equity models.

PURPOSE

In a hotly political context, objectivity is often sought by means of statistical methods. The hope is that a mathematical model applied to a political issue will produce a result that is free of biased human judgment. This faith in the objectivity of statistical methods is misplaced. Human judgment is an inherent part of any statistical method. Resorting to statistical methods may serve political interests by removing from public view the arena in which judgment is exercised but never by removing the necessity for making those judgments.

Faculty salary equity is a hot political issue that may have severe legal, monetary, and human consequences. It is also an issue that often requires the use of sophisticated statistical techniques for the determination of inequity. The purpose of this paper is to identify the areas in which human judgment must be made in order to conduct a statistical analysis of salary equity and to provide

Nelle Moore, Director of Institutional Research, San Juan College, 4601 College Blvd., Farmington, NM 87402.

Presented at the Annual Forum of the Association for Institutional Research, Atlanta, Georgia, May 1992.

0361-0365/93/0200-0107$07.00/0

some informed guidelines for making those judgments. The direction and magnitude of the final results are contingent on the way these statistical decisions are made. Therefore, careful consideration of these issues is essential for conducting a fair and defensible salary equity study.

DECISION ELEMENTS

The first essential decision that must be addressed is, What is the purpose of the salary equity study? Is the purpose to identify if an institution or department shows a pattern of salary inequity? Which individuals are underpaid? How much money would be required to achieve equitable compensation throughout the institution? There is no formula approach; there is no one-size-fits-all regression model. The methods and models must be selected according to the specific question to be answered. In order to select an appropriate model, one must begin by clearly identifying the goal of the study and then understand the scope and limits of the various methods for achieving that goal.

Having established the focus of the study, there are four primary decision elements that must be addressed in order to conduct a salary equity study.

1. *What group of people should be included in the model?* Is this a study of faculty or another employee group? Are all faculty to be included or only full-time, tenure-eligible positions to be included? Longitudinal or cross-sectional data? All departments? Males and females in one group or separate groups? All males or just white males?
2. *What variables should be included in the model?* Should rank, which may itself be biased, be included in the analysis? Are productivity and performance variables essential? How many variables are necessary? Should all possible variables be included or only a few main ones? How should discipline or department variables be handled?
3. *What statistical model should be used?* Should a regression model be run on white males and then the resulting formula be applied to females? Should males and females be in the same model with the inclusion of a sex variable? Should a standard or a hierarchical model be employed?
4. *Which outcome statistics should be used to interpret the results?* Should regression residuals, R^2, standardized or unstandardized coefficients, or tests of statistical significance be used to assess the outcome of the analysis?

BASIS FOR ESTABLISHING JUDGMENTAL CRITERIA

There are four fields of study that offer insights into statistical criteria for conducting salary equity studies. These are Law, Economics, Statistics, and Institutional Research. The criteria by which the above decision elements can

be addressed will be drawn from an integrated study of these four fields. Each of these fields will contribute in varying degrees toward establishing criteria on the decision elements. Nowhere have these four fields been combined to form a set of guidelines for faculty salary equity studies. More often, these different approaches have confronted each other in the courtroom by representing opposing parties. In this context, the apparent strategy of the litigants is to select the approach that best supports its own point of view.

The foundation for salary equity studies lies in Law since it is this field that defines and requires equity. The law also defines what is and is not acceptable evidence in discrimination cases. However, the law does not specify the details of analytical procedure and is often lacking in a full understanding of statistical methods. A study of Statistics is necessary to ensure that the design and interpretation of the study is methodologically appropriate. Statistical models can vary, however, depending on the applied field of study. For example, in Economics, human capital theorists have developed specific statistical methods for accounting for human characteristics that contribute to differences in salary earnings. And in Institutional Research, several methods have been used specifically to identify salary inequities in higher education.

This paper will not attempt to provide any definitive answers to the decision elements. The answers are a matter of choice depending on the purpose of the salary equity study, the availability of data, the time frame or duration of the study, and the context within which the data will be presented. Rather, this paper will use the decision elements to query each of the four fields in search of criteria for making informed decisions. Through this discussion, the author hopes to bring a broader perspective and, if not objectivity, then ethical fairness to the process of designing salary equity models.

PERSPECTIVES FROM THE FOUR FIELDS

Law

There are two laws under which gender-based salary equity suits are brought to trial: Title VII of the Civil Rights Act of 1964, as amended in 1972, and the Equal Pay Act of 1963. The Equal Pay Act is an amendment to the Fair Labor Standards Act, which specifically exempted executive, administrative, and professional employees. However, in the Education Amendment of 1972, Congress made the Equal Pay Act and Title VII applicable to these professionals in Higher Education.

Equal Pay Act—"Equal Pay for Equal Work"

The Equal Pay Act (EPA) relates specifically to gender-based discrimination in wages.

> No employer shall discriminate . . . between employees on the basis of sex by paying wages . . . at a rate less than the rate at which he pays wages to employees of the opposite sex . . . for equal work on jobs the performance of which requires equal skill, effort, and responsibility and which are paid under similar working conditions except where such payment is made pursuant to (1) a seniority system, (2) a merit system, (3) a system which measures earnings by quantity or quality of production, or (4) a differential based on any other factor other than sex.

The first burden under EPA is for the plaintiff to convince the court that the work of the parties is substantially equal. The equal work standard is defined very specifically. The focus is on the actual job performance and content, and not on job titles or descriptions (*Spaulding v. University of Washington*, 1984). Comparable work is not a substitute for equal work. The proper comparison criterion for skill level is the skills that a job requires and not the skills that the employees possess.

Once the plaintiff has convinced the court that the jobs in question are substantially equal and that the pay is unequal, the burden shifts to the defendant. The defendant must show that the reason for the difference in pay is non-discriminatory under one of the four legitimate exceptions allowed in EPA: seniority, merit, quantity or quality of production, or any other non-gender-based factor. The plaintiff then has the opportunity to show that the reason given by the defendant is pretextual and the pay differential is in fact due to sex discrimination.

A common defense under the fourth exemption is market. The courts generally agree that paying the competitive market price is accepted business practice and does not qualify as employment discrimination (*AFSCME v. State of Washington*, 1985; *American Nurses Assoc. v. State of Illinois*, 1986; *Spaulding v. University of Washington*, 1984). The employer is treating all employees neutrally in setting wages at the market value. The employer did not create the market disparity and therefore cannot be held liable for any inequity in market rates.

Another common defense is merit. The defense of merit, however, has on occasion been rejected by the courts (*Board of Regents of Nebraska v. Dawes*, 1975; *Marshall v. Georgia Southwest College*, 1980; *Mecklenburg v. Montana State Board of Regents*, 1976). The merit exception will not hold if the merit system is "operated in an informal and unsystematic manner," if the employees are unaware of any merit system, or if merit evaluations are conducted in a subjective, ad hoc manner (*Marshall*, 1980). Even if a true merit system does exist, the existence of a system is insufficient (*EEOC v. McCarthy*, 1984). The defendant must show that the difference in pay is a result of that system. This same flaw was found regarding the market defense when the defendant failed to produce specific evidence that it had assessed the current market rates for the employee position under review (*Marshall*, 1980).

An example of a case tried under EPA is *Board of Regents of the University of Nebraska v. Dawes* (1975). This is a case of reverse discrimination brought by male employees of the University of Nebraska. The university decided voluntarily to correct gender-based inequity in salaries. They developed a formula, based on male salaries, which accounted for the factors that determined salaries. The formula was used to compare individual female salaries with the average male salaries. A one-time adjustment was made to those females with salaries below the average male salary. This adjustment left a number of males who had salaries below the formula salary.

The U.S. Court of Appeals ruled in favor of the male employees. "When a university establishes and effectuates a formula for determining a minimum salary schedule for one sex and bases the formula on specific criteria . . . it is a violation of the Equal Pay Act to refuse to pay employees of the opposite sex the minimum required under the formula" (*Board of Regents of Nebraska*, 1975). The EPA is readily applicable in cases such as Dawes where an institution adopts a standard pay scale or formula and applies it differentially on the basis of sex alone.

Title VII—Disparate Treatment, Disparate Impact

Title VII prohibits discrimination in compensation on the basis of race, religion, sex, or national origin without making reference to the equal work requirement. To eliminate any potential conflicts between EPA and Title VII, Congress adopted the Bennett Amendment stating that Title VII does not negate EPA. There has been substantial debate over how to interpret Bennett (*County of Washington v. Gunther*, 1981; Lee 1989; Weeks, 1985). Does Bennett incorporate into Title VII the EPA requirement of proving equal work or does Bennett refer only to the four defenses allowed under EPA? If Bennett is interpreted as allowing the EPA defenses, then Title VII can be used to argue sex discrimination even when the jobs in question are not equal.

The issue was decided by the U.S. Supreme Court in the Gunther (*County of Washington v. Gunther*, 1981) case. In a 5–4 vote, the Court ruled that Bennett did not require the equal work standard for Title VII claims. By this decision, the Court allows gender-based salary discrimination claims to be actionable under Title VII even if they do not meet the equal work standard. The Supreme Court declined to set specific standards for claiming discrimination under Title VII, leaving the lower courts flexible to interpret the facts and evidence on a case-by-case basis. The dissenting judges feared that this interpretation of Bennett would allow claims of comparable worth under Title VII, an outcome that they felt Congress did not intend (Lee, 1989). The issue of comparable worth will be discussed further at the end of this section.

Under Title VII, there are two types of discrimination: disparate treatment and disparate impact. Disparate treatment involves intentional discrimination

with discriminatory motive. Disparate impact occurs when an employer's seemingly neutral policies or practices have a disproportionate effect.

The structure of the cases under disparate treatment and disparate impact are somewhat different. Disparate treatment theory was first developed in *McDonnell-Douglas v. Green* (1973). The plaintiff has the burden of proving a prima facie case of discrimination by a preponderance of the evidence. In a class action suit, the evidence must establish a pattern, practice, or custom of discrimination based on sex. A prima facie case does not require proof of actual or overt discrimination but should be at least suggestive for discrimination (*Barnett v. Grant*, 1975; *Mecklenburg*, 1976). However, the plaintiff must still show that the employer has a discriminatory motive either by direct or circumstantial evidence.

Whether or not a prima facie case has been established depends on the facts of each case. In *Spaulding* (1984), brought by nursing faculty, the Court ruled that it "cannot infer intent to discriminate merely from the existence of wage differentials between jobs that are only similar." While in *Bazemore* (1986), the Court accepted a regression analysis as prima facie evidence "that it is more likely than not that impermissible discrimination exists."

If the plaintiff succeeds in establishing a prima facie case, the burden shifts to the defendant to show that there is some nondiscriminatory reason for the disparate treatment. Legitimate reasons for disparity include the four defenses under EPA, occupational qualifications, and business necessity. A continuation of inequity existing before discrimination became illegal is not an acceptable defense (*Bazemore v. Friday*, 1986; *Sobel v. Yeshiva University*, 1988). If the defendant succeeds in presenting evidence of nondiscriminatory reasons, the plaintiff then has the opportunity to show that the defendant's reasons are not the real reasons but are pretexts for discrimination.

The structure of disparate impact was established in *Griggs v. Duke Power Company* (1971). The burdens of proof under disparate impact differ from disparate treatment only in the first phase where the plaintiff bears the burden. The burden of the defendant and the counter by the plaintiff remain the same as disparate treatment. Under disparate impact, the plaintiff is not required to show discriminatory intent but may simply show that the employer's seemingly neutral policy or practice has a disproportionate impact on females. *AFSCME* (1985) makes it clear that disparate impact claims must "challenge a specific, clearly delineated employment practice," stating that wage scales are "too broad" to be interpretable as policies under disparate impact. Plaintiff must prove that there is a discriminatory impact and also must show the specific practices or policies that produced it. The important issue here is to show the specific employer's acts that have led to the disproportionate impact (*Spaulding v. University of Washington*, 1984).

Comparable Worth

Comparable worth is an assessment of the intrinsic value of jobs to the employer. Jobs within a place of employment are evaluated on a variety of factors important to the employer. A numeric value is then assigned to all jobs based on a common scale. The concept of comparable worth proposes that compensation rates should be set according to the relative value of the jobs as measured on this scale. Jobs with a similar value rating should receive comparable wages regardless of the dissimilarity of the work.

It was the concern of the Rehnquist dissent in *Gunther (County of Washington v. Gunther*, 1981) that the removal of the equal work requirement under Title VII would thereby allow comparable worth claims. Rehnquist felt that the notion of comparable worth was specifically rejected by Congress when they adopted the EPA (Lee, 1989). The *Gunther* Court, however, neither explicitly rejected nor endorsed comparable worth when delineating its interpretation of Bennett.

Even though the Rehnquist opinion is not law, it has probably had a negative influence on the fate of comparable worth. The lower courts have consistently rejected comparable worth as an actionable theory. In *American Nurses Assoc. v. State of Illinois* (1986), the court refused to play a role as evaluator of the value of jobs. "Courts are not authorized to engage in wholesale re-evaluation of any employer's pay structure in order to enforce their own conceptions of economic worth."

Comparable worth is thereby rejected as a theory when it places the court in a position of having to make subjective evaluations of jobs or to assess the validity of an employer's job evaluation system. The courts will not get involved with evaluating pay scales, but they will compel an employer to apply equally a scale that the employer has already adopted (*Board of Regents of Nebraska v. Dawes*, 1975; *County of Washington v. Gunther*, 1981; Weeks and Organ, 1986). This is the distinction made in *Gunther*. By setting aside the equal work requirement, the Supreme Court has allowed cases to be tried where the employer has discriminated in the implementation of its own pay scale.

Comparable worth claims also fail when they rely on market inequities as evidence of discrimination. The courts have accepted competitive market prices as a legitimate defense under EPA's fourth exemption. However, some state legislatures have enacted legislation on comparable worth.

Economics

The contribution of the field of Economics to the study of salary equity comes from human capital theory. Human capital theory views income as a function of market factors. Individuals who invest in human capital can expect

greater returns on their investment in the form of higher earnings. Investments in such things as education and job training increase one's value to an employer presumably because these elements of human capital yield higher productivity (Parcel and Mueller, 1983).

Discrimination, in human capital terms, occurs when the wage of females is lower than the wage that would have been earned if the human capital were valued the same for males and females. Given two individuals of different sex with the same education level, experience, and seniority, differences in salary could be attributed to discrimination. Conversely, if there were no discrimination, then the wage structure for males would also apply to females and the wage structure for females would also apply to males (Oaxaca, 1973).

The human capital analysis of salary differentials between males and females has been operationalized by Oaxaca (1973) in the following formula:

$$W_m - W_f = b_m(X_m - X_f) + (b_m - b_f)X_f$$

where: W = wage, X = human capital characteristics, and b = weight. This formula separates the male-female wage gap into two parts. The first part, $X_m - X_f$, is the amount of wage difference that is attributable to differences in human capital characteristics between males and females times the rate of male return on investment (b_m). The second part, $b_m - b_f$, is the difference between male and female returns on investment for the same human capital characteristics (X_f). The second part of the equation is the measure of wage discrimination (Gunderson, 1989).

The standard procedure for analyzing salary equity is to estimate the earnings equations separately for males and females (Gunderson, 1989). Wages are usually measured in logarithmic terms so that the coefficients reflect the proportionate effect of changes in human capital characteristics on wages. The regression coefficients indicate the rate of return that the market yields for a unit change in individual characteristics.

In studies of academic salaries, the factors usually included in the human capital model are years of experience (years since Ph.D.), seniority (years at the institution), education level (highest degree earned), and field of expertise (academic discipline). The human capital theory assumes that these characteristics are correlated with productivity. Rank is not considered to be an element of human capital since it is conferred by the institution, as well as salary, and is a result of the investment individuals have made in their human capital.

In addition to the typical human capital factors that are at best general correlates of productivity, studies in higher education have attempted to use direct measures of productivity by including the number and type of publications in the model (Abramson, 1975; Bayer and Astin, 1975; Centra, 1974; Lewis and Becker, 1979; Tuckman and Tuckman, 1976). These studies are contradictory,

some showing that males produce more research than females and some showing no difference. In some cases, the analysis has attempted to look at marital status as well. These efforts to incorporate more complete and detailed models have not produced significant reductions in the size of the salary differential. The human capital models fail to explain significant differentials even where data exist on research and teaching performance (Hirsch and Leppel, 1982). In more than 10 years of human capital research, productivity-related variables are unable to account for more than about half of the male-female gap in earnings (Daymont and Andrisani, 1984).

In 1975, Johnson and Stafford published an often quoted study that used human capital theory to examine the salaries of female faculty. They offered an alternative to discrimination as an explanation of the sex differential in salaries. They felt that the salary differential could be the "market's reaction to voluntary choices by females with regard to lifetime labor-force participation and on-the-job training." These choices reflect a woman's expectation to leave the marketplace during childbearing years and therefore lead to an accumulation of less human capital. This conclusion was supported in their minds by data that showed the starting salaries of female faculty as "not much less" than those of males (4% to 11% less in six disciplines) and a "fairly substantial differential" (13% to 23%) 15 years after obtaining the Ph.D. (Johnson and Stafford, 1975).

However, the Johnson and Stafford study did not directly measure leave of absence differentials between males and females. An ACE study conducted by Darland et al. (1973) found that one-fourth of all faculty had interrupted their careers for more than one year and that a greater percentage of men had interrupted their careers than women.

Hirsch and Leppel (1982) also dispute the pattern of salary differential upon which Johnson and Stafford base their argument. Johnson and Stafford (1975) theorize that the sex gap is smallest at the beginning of the career when human capital accumulation is similar for males and females. The salary differential widens with experience as females choose to withdraw from the labor force but may narrow later after the childbearing years. They state that this pattern should be evident at any U.S. university. Hirsch and Leppel (1982) found that women faculty at one university received equal rewards from experience capital during the early years but that the differential widened in later years. Hirsch and Leppel note that, "While we cannot clearly test between the human capital and discrimination explanations for salary differentials, our evidence strongly suggests that universities can and do exercise significant discretion in the awarding of salaries."

The usefulness of a human capital model for explaining salary differences within a specific university is questionable. Human capital theorists tend to employ large national databases to study sex differentials in society as a whole. Their purpose is to gain theoretical understanding of the contributions of differ-

ent human capital characteristics and an explanation of the sex gap in terms of differences in human capital. Human capital models, however, do not ordinarily conduct analyses that are specific to one institution. Even on a national level, the precision of human capital models is not great. They typically account for less than one-third of earnings differences among male workers (Fogel, 1986).

Statistics

Salary equity has not been addressed in statistics literature as a methodological problem. The basic tools used to analyze salary equity belong to the field of statistics, but no one has addressed the issue of how to choose the most appropriate statistical model.

It is generally recognized in salary equity research that some form of multiple regression analysis is the minimum requirement for a believable study. Regression analysis has the ability to simultaneously account for the effects of multiple variables and to extrapolate data points where there may be missing cells. However, within the context of regression analysis, there remains a large number of possible models. The field of statistics does not provide definite rules as to which model should be used. A simple piece of advice is to be sure that a statistical model is selected that is capable of answering the practical question at hand. Therefore, it is essential to keep the purpose of the analysis in mind while trying to decide which model would be most appropriate and which outcome measures should be used to interpret the data (Williams, 1959).

It is unreasonable, however, to expect to find a model without flaws, or to find a model that perfectly reproduces the institution's salary decisions. Fortunately, a perfect model is not necessary in order to gain insight into the salary decision process, but it is important to understand the limits of the chosen statistical method. A limitation common to all models is that no statistical model will prove anything (Baldus and Cole, 1980; Long in Pezzullo and Brittingham, 1979).

Because of the limitations in any given model, it may be useful to analyze data using several different models, each with its strengths and weaknesses (Namboodiri et al., 1975; Miller and Fredericks, 1991). Using multiple models can increase confidence in the outcome when the results are confirmed by more than one method, or multiple models can provide new insights when results differ. When different models produce different results, it may be more reasonable to present salary differentials as lying within a range of numbers than to imply a high level of precision by reporting one number. The difficulty with trying a large number of models lies in indiscriminate selection of alternatives without having a well-thought-out rationale for each model. Lack of forethought can lead to a confusing array of conflicting results without any means of making sense out of them.

The importance of forethought is relevant to another issue, and that is the use of step-wide regression. The author shares the view of many statisticians that the use of step-wise regression for model building is a "thoughtless technique" (Ramsey in Pezzullo and Brittingham, 1979). It is the responsibility of the researcher (not the computer) to identify the variables to include and how they will be entered into the model based on a reasonable assessment of the specific situation.

While choices about which variables to include, whether to analyze males and females in the same or separate models, and whether to use simple regression or a hierarchical approach, are statistical issues, they cannot be made on a purely statistical basis. Statistical methods can be adapted for all of these options. However, there are two purely statistical issues that are worth considering here: (1) multicollinearity and (2) statistical significance.

Multicollinearity

In higher education practice, multicollinearity is one of those words that operates as a discussion stopper. When someone says, "You can't use that variable because of multicollinearity," who can dispute it? Few people understand what it is or how to determine if it is a problem. Multicollinearity occurs when two or more independent variables in the model are correlated to the extent that they form near-linear relationships with each other (Baldus and Cole, 1980). Multicollinearity goes beyond the simple correlations between variables because the simple correlations do not account for the effects of all of the variables in the model on each other. Furthermore, multicollinearity does not occur whenever there is intercorrelation but only when that intercorrelation is near perfect. The problem with a near-linear relationship between independent variables is that there is no unique solution to the equation (Netter et al., 1985; Darlington, 1968). More than one equation can fit the same data; therefore, conclusions drawn from the regression coefficients will be misleading.

Multiple regression does not require independent variables to be uncorrelated, and as long as there is some random error, multicollinearity does not occur. When multicollinearity occurs, it increases the standard errors of the estimated coefficients, which affects the determination of significance. It will produce widely different regression coefficients in different samples that come from the same population.

The impact of multicollinearity for salary equity analysis is minimal. On the first hand, the variables entered into typical salary equity models will seldom approach perfect intercorrelation. Second, where salary equity models are conducted on the population of faculty at one institution, the purpose of the analysis is not to make inferences from sample coefficients to the population. The regression coefficients are not estimates of the parameters but are the parameters themselves; therefore, instability of estimated coefficients is not an issue.

Finally, there is a standard test available to determine if multicollinearity is a problem for a specific model (Baldus and Cole, 1980). If the level of intercorrelation between variables is too high, it can be reduced by transforming the variables.

Statistical Significance

The issues surrounding the appropriate use of significance tests are complex and controversial. The debate has been ongoing since at least the early 1950s and is unlikely to be resolved in this paper. Rather than recap the historical arguments, the discussion here will focus on the use of significance tests in salary equity studies and will certainly be biased toward the side of the debate on which the author stands.

A test of statistical significance is defined as "a formal procedure for making a decision between two hypotheses about some characteristic of a population (parameter) on the basis of knowledge obtained from a sample (sample statistic) of that population" (Morrison and Henkel, 1969).

This formal procedure involves a comparison of the difference between the sample statistic and the parameter that would be expected given the null hypothesis and a normal sampling distribution. The level of statistical significance indicates the probability that a sample statistic of the obtained size would have occurred over repeated samples from the same population if the null hypothesis reflected the true population parameter. If the value of the sample statistic could be expected due to sampling variation (chance), then the null hypothesis that the sample came from the population is accepted and the test is not significant. If, however, the value of the sample statistic is so extreme as to be improbable from sampling variation alone, then the test is significant and the null hypothesis is rejected.

The foundation of tests of statistical significance is probability theory, which stems from random sampling distributions. Statistical inference from a sample to a population has no meaning apart from the assumption of randomness, which is central in all probability sampling designs (Morrison and Henkel, 1969). Tests of statistical significance only have meaning in the context of random samples of finite populations. The population must exist; it must be defined in advance. The sample must be drawn from the defined population according to a deliberate randomization strategy according to a given research design. Only within this context does statistical significance have any meaning or application.

In practice within social science research, statistical significance has accumulated a wide range of additional interpretations based more on wishful thinking and fantasy than on science or probability theory. Carver (1978) lists three of these fantasies that have gained common acceptance and usage. *Fantasy #1:*

odds against chance. This fantasy interprets the *p* value as the probability that the research results were due to chance. This is a subtle fantasy because it uses the right words and therefore sounds true, but it twists the meaning of the test. The hypothesis under test is that chance was completely responsible for the observed difference. The *p* value is used to decide whether to accept or reject that hypothesis—yes or no. The fantasy changes the interpretation of the *p* value to a continuous probability and is evident in misstatements such as "the results were highly significant," or "the results were almost significant." *Fantasy #2: reliability*. Fantasy #2 interprets statistical significance as the probability that a replication of the study would produce the same results. This fantasy misplaces the contribution of repeated sampling to the interpretation of the significance tests. Probability theory is based on the sampling distributions that occur from drawing repeated random samples from a population. The *p* value indicates the probability that the results could have been obtained from the hypothesized population but says nothing about the degree of confidence in the reliability of the results. *Fantasy #3: truth of hypothesis*. Fantasy #3 interprets statistical significance as the probability that the research hypothesis is true. Some studies have used the *p* value as evidence that the research hypothesis was probably true even if the design of the study was questionable. Scientific method is not designed to prove that anything is true. The design of the statistical test is set up only to *disprove* the null hypothesis.

Much of the fantasy surrounding statistical significance has been perpetuated by the unfortunate choice of terminology and the resulting confusion with normal usage of the word *significant*. It is all too easy to confuse statistical significance with substantive importance. In fact, the confusion is deliberate, even if unconscious, since researchers crave an objective measure for evaluating whether a result is important. Statistical significance provides the illusion of objectivity and has become the sine qua non crutch of social science research. The interpretation of statistical significance is bent and stretched to draw whatever specific conclusions are needed in the study while giving the appearance of using scientific method. This is not done through intentional misrepresentation but through misunderstanding of complex issues.

In salary equity studies, the statistical significance of a sex coefficient is often used as an indicator of discrimination. If the sex coefficient is statistically significant after all other variables in the model have been taken into account, then seemingly there is support for a hypothesis of discrimination. It is this author's view that significance tests have no meaning, and therefore should not be used, in salary equity studies. This view is based on the nature of the data used in such studies. At an institutional level, salary equity studies typically include the population of faculty at the institution. (If they don't, they should. There is no reason to draw a sample when the complete data are available.) There is nothing random about the data, about the hiring process, or about the

awarding of salaries. There is no sampling procedure. There simply is no context within which the use of significance tests could be considered appropriate.

A legitimate concern of the courts is that a judgment might be based on a result that may not reflect decisions made over the long run. The courts are wary of "imposing liability on people for events caused by chance. Statistics is an effective tool for measuring that risk" (Baldus and Cole, 1980). Baldus and Cole outline an approach to statistical inference that includes the following reference to significance tests: "Obtain numbers that reveal the likelihood that disparity in the sample would have occurred if those decisions had been selected at random from a large universe of decisions in which there was no disparity whatever."

This interpretation prompts the addition of a fourth fantasy, the *Let's pretend* fantasy. Let's pretend that we have a population and let's pretend that we selected a sample at random from that population. What would our results mean if they were interpreted as if they were sample statistics? This fantasy is perpetuated by the use of computer statistical packages that routinely produce p values regardless of whether the data are population data, convenience samples, or randomly selected samples. Morrison and Henkel (1969) respond here to two examples of the fantasy.

> However, it has been argued that significance tests can be applied to such a set of cases or to any set of cases by treating the set (1) as a probability sample of a hypothetical universe of possibilities (Blalock, 1960; Hagood and Price, 1952), or (2) "as an indication of a probability that an observed association could be generated in the given set of data by a random process model" (Gold, 1969; Blalock, 1960), for instance, by repeated random redistribution of the marginal frequency of a fourfold table into cells.
>
> Both require the assumption that the particular result observed was somehow randomly selected from some larger set of possible results. On what basis this assumption is warranted except on the desire of the researcher to apply the statistical inference model is not clear. Are some or all of the specific benefits of probability sampling available regardless of whether the sample is a probability sample? We doubt it. Statistical inference depends on a statistical theory, but to be applicable, the theory also depends on certain empirical questions in research. To ask whether a given result could be generated by a random process model in the absence of a random process in the generation of the data is simply to raise an irrelevant question; an absolutely crucial feature if the application of the model is missing.

The primary reason that the courts look at significance tests is to eliminate chance as an explanation for the results. There is a more direct way to address this concern. Salary equity studies can be replicated by repeating the study in each of several past years. The cumulative results will be indicative of the "pattern and practice" of the institution and should satisfy the court's concern in this regard. "The popularity of statistical significance would probably decline appreciably if it were more widely recognized that it is not a predictor of the

replicability of research data. It seems best to rely upon direct evidence of replication rather than upon the myth that somehow statistical significance predicts replicability" (Carver, 1978).

Institutional Research

There is very little published in the field of Institutional Research to guide the novice researcher in salary equity model selection. Most of the papers written by institutional researchers are unpublished perhaps due in part to the sensitive nature of the subject. One source of information is the AAUP Higher Education Salary Evaluation Kit (Scott, 1977).

The AAUP salary kit promotes a method for "flagging" individual female faculty who are apparently underpaid relative to white males. According to the AAUP method, a regression model is constructed to predict the salaries of white male faculty from the variables year of birth, highest degree, and year of highest degree within appropriate funding units/departments. The resulting regression weights are then applied to female faculty. All female faculty who receive a lower actual salary than is predicted by the male formula are "flagged" for review by a committee or administrator. Scott notes that although the method may be criticized as "quick and sloppy," it serves its purpose. Indeed it is worthwhile to consider what is the purpose that is served by this method.

According to Scott (1977), higher education administrators "have a strong feeling that salary inequity is an exceptional thing and that class action is not appropriate." However, there is no reason to assume that only a few individuals are discriminated against and that most personnel decisions are free of bias. The issue can and should be settled empirically by examining the institutional data. This issue is important because it influences the type of analysis that is performed and, ultimately, the type of remedy that may be applied.

In the AAUP salary kit, Scott takes the approach of flagging individual females, an approach based on the view that inequity is an exceptional occurrence. Gray and Scott (1980) present an approach based on the view that sex bias impacts the salaries of all female faculty as a class. They argue that regression models are sensitive to systematic bias and that females at all salary levels are paid less than they would have been paid if they were male.

A more recent source from the field of institutional research is the 1985 New Directions for Institutional Research publication of AIR entitled *The Use of Data in Discrimination Issues Cases*. This volume, edited by Rosenthal and Yancey, contains a selection of chapters written to help institutional researchers deal with the data requirements of litigation. It is not the purpose of this book to discuss the specifics of model selection and therefore it has very little to offer in terms of specific guidelines for developing a salary equity model. It

does provide some sound advice and a worthwhile perspective on the role of institutional researchers facing salary equity litigation.

The most useful source of information from institutional research can be found in Pezzullo and Brittingham's (1979) edited volume *Salary Equity*. The contributed chapters in this book raise a number of the issues and problems involved in specifying salary equity models. The approaches represent a range of complexity. The authors also present possible solutions and alternative approaches to specific problems. There is however no attempt to integrate the approaches. Aside from one chapter that critiques a method in a previous chapter, there is no analysis provided by the editors that expresses the "state-of-the-art" in institutional research. After cataloging each of the chapters according to the four decision elements for which population, variables, models, and outcomes were used, there does not seem to be any consensus or common methodology among institutional researchers as of 1979. The influence of AAUP methodology is evident but not predominant.

To get a sense of current practice in institutional research, I solicited information from several related electronic-mail lists and obtained responses from 12 institutions regarding their salary equity studies. These responses ranged from a few paragraphs via e-mail to fully descriptive white papers. A catalog of these responses according to the four decision elements also showed a wide range of approaches. Some variation in approach is to be expected due to the range of possible issues that could be addressed in a salary equity study. Among the studies reported in Pezzullo and Brittingham and the 12 informally solicited reports, there are some common patterns that may begin to describe the "typical" institutional research approach to salary equity studies.

Sadly, stepwise regression is frequently used to select the variables in the model. This practice may be an indication of the power of politics outweighing purposeful model building. In a political rather than a research environment, giving the computer responsibility for selecting variables may be a great relief to the institutional researcher!

There seem to be two predominant models for salary equity: (1) the AAUP method of applying the male formula to females and flagging females with negative residuals and (2) combining males and females in the same model and determining the statistical significance of the sex variable. Most studies use only full-time tenured and tenure-track faculty in the analysis.

The commonly used variables are rank, highest degree, years since Ph.D. (experience), years since hire (service), and discipline (market). Although there is sometimes mention of the issue of possible bias in rank, and despite AAUP's recommendation to exclude it, rank is included in every case.

Conspicuously absent from institutional studies is any measure of productivity or merit. Measures of productivity have been examined using the ACE national database (for example, Tuckman in Pezzullo and Brittingham, 1979)

and on subsamples in some institutional studies (for example, Muffo et al. or Brittingham et al. in Pezzullo and Brittingham, 1979). The absence of productivity or merit measures in institutional studies after almost 20 years of salary equity research is particularly disturbing. Differentials between males and females are often "explained" away on the basis of these factors without ever measuring them. Institutional researchers understandably use the variables that are readily available to them on institutional databases. However, if institutions award salaries and salary increases on the basis of productivity and merit as they claim, it is not unreasonable to expect that these variables would have become a part of the institutional database by now. The fact that these variables remain unavailable and still lack accepted operational measures tends to undermine institutional credibility in this area. Until the institutions can produce valid measures for productivity and merit that can be examined for fairness or bias, models of salary equity can hardly be faulted for not including these measures. Until then, it has not been proven that salary inequities can be explained by sex differences in productivity and merit nor has it been established that institutions truly assign salaries on the basis of these factors and not on some other, less lofty criteria.

CONCLUSION

This paper has presented a framework for selecting salary equity models based on four decision elements and four fields of study. The discussion on the perspectives from the four fields has provided only the initial groundwork without going into detail on each of the decision elements. The next step in this study will be to apply the contributions from the four fields to an analysis of specific salary equity models. This analysis will establish criteria for evaluating the usefulness and fairness of different approaches to determining salary equity.

REFERENCES

Abramson, J. (1975). *The Invisible Woman*. Washington, DC: Jossey-Bass.

AFSCME v. State of Washington, 770 F.2d 1401 (9th Cir. 1985).

American Nurses Assoc. v. State of Illinois, 783 F.2d 716 (7th Cir. 1986).

Baldus, D. C., and Cole, J. W. L. (1980). *Statistical Proof of Discrimination*. New York: McGraw-Hill.

Barbezat, D. A. (1987). Salary differentials by sex in the academic labor market. *The Journal of Human Resources* 22(3):422–428.

Barnett v. Grant, 518 F.2d 543 (4th Cir. 1975).

Bayer, A. E., and Astin, H. S. (1975). Sex differentials in the academic reward system. *Science* 188: 796–802.

Bazemore v. Friday, 106 S.Ct. 3000 (1986).

Berger, M. C., and Black, D. A. (1991). Faculty salary differences at the University of Kentucky: A reconsideration of results. Presented in *Let Facts Be Submitted to a Candid World*. Lexington, KY: University of Kentucky.

Beyond the prima facie case in employment discrimination law: Statistical reproof and rebuttal (1975). *Harvard Law Review* 89: 387–422.

Blalock, H. M. (1960). *Social Statistics*. New York: McGraw-Hill.

Blum, D. E. (1990, June 13). Ten years later, questions abound over Minnesota sex-bias settlement. *Chronicle of Higher Education*, pp. A13–A15.

Board of Regents of Nebraska v. Dawes, 522 F.2d 380 (8th Cir. 1975).

Bock, R. D. (ed.) (1989). *Multilevel Analysis of Educational Data*. San Diego, CA: Academic Press.

Campbell, T. J. (1984). Regression analysis in Title VII Cases: Minimum standards, comparable worth, and other issues where law and statistics meet. *Stanford Law Review* 36: 1299–1324.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review* 48(3): 389–399.

Centra, John A. (1974). *Women, Men and the Doctorate*. Princeton, NJ: Educational Testing Service.

Civil Rights Act of 1964, as amended in 1972, Title VII, 42 U.S. C.

County of Washington v. Gunther, 452 U.S. 161 (1981).

Darland, M. D., Dawkins, S. M., Lavasich, J. L., Sherman, M. E., and Whipple, J. L. (1973). Application of multivariate regression to studies of salary differences between men and women faculty. In *Proceedings of the Social Statistics Section* (pp. 120–132). Washington, DC: American Statistical Association.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin* 69(3): 161–182.

Daymont, T. N., and Andrisani, P. J. (1984). Job preferences, college major, and the gender gap. *The Journal of Human Resources* 19(3): 408–428.

Denny v. Westfield State College, 669 F. Supp. 1146 (1987).

EEOC v. McCarthy, 578 F.Supp. 46 (1984).

Equal Pay Act of 1963, 29 U.S. C.

Fisher (1980). Multiple regression in legal proceedings. *Columbia Law Review* 80: 702.

Fogel, Walter (1986). Class pay discrimination and multiple regression proofs. *Nebraska Law Review* 65(242): 289–329.

Gold, D. (1969, February). Statistical tests and substantive significance. *American Sociologist* 4: 42–46.

Goldin, C. (1990). *Understanding the Gender Gap: An Economic History of American Women*. New York: Oxford University Press, Inc.

Goldin, C., and Polachek, S. (1987). Residual differences by sex: Perspectives on the gender gap in earnings. *American Economic Review* 77(2): 143–151.

Goodman, R., Hoenack, S., and Rasmussen, M. (1989). *Statistical Analysis of Salaries for Tenured and Tenure-Track Faculty at the Twin Cities and Duluth Campuses of the University of Minnesota*. Working paper. University of Minnesota, Office of Management Planning and Information Services.

Gordon, N. M., Morton, T. E., and Braden, I. C. (1974). Faculty salaries: Is there discrimination by sex, race, and discipline? *American Economic Review* 64(3): 419–427.

Gray, M. W., and Scott, E. L. (1980). A "statistical" remedy for statistically identified discrimination. *Academe* 66(4): 174–181.

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Gunderson, M. (1989). Male-female wage differentials and policy responses. *Journal of Economic Literature* 27: 46–72.

Hagood, M., and Price, D. (1952). *Statistics for Sociologists*. New York: Holt.

Hirsch, B. T., and Leppel, K. (1982). Sex discrimination in faculty salaries: Evidence from a historically women's university. *American Economic Review* 72(4): 829–835.

Hoffman, E. R. (1976). Faculty salaries: Is there discrimination by sex, race, and discipline? Additional evidence. *American Economic Review* 66(1): 196–198.

Johnson, G. E., and Stafford, F. P. (1975). The earnings and promotion of women faculty. *American Economic Review* 64(6): 888–903.

Kazal-Thresher, D. M. (1990). *Employment and Earnings Patterns of Stanford MBAs: Gender Comparisons Among Thirteen Graduating Classes*. Unpublished doctoral dissertation, Stanford University, Palo Alto, CA.

Key v. Gillette, 29 Empl. Prac. De. (CCH) 32, 909 (1982).

Keyes v. Lenoir Rhyne College, 552 F.2d 579 (4th Cir. 1977).

LaNoue, G. R., and Lee, B. A. (1987). *Academics in Court: The Consequences of Faculty Discrimination Litigation*. Ann Arbor, MI: The University of Michigan Press.

Lee, Y. S. (1989). Shaping judicial response to gender discrimination in employment compensation. *Public Administration Review* 49: 420–430.

Lewis, D. R., and Becker, W. E. (1979). *Academic Rewards in Higher Education*. Cambridge, MA: Ballinger Publishing Company.

Marshall v. Georgia Southwest College, 489 F.Supp. 1322 (1980).

McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).

Mecklenburg v. Montana State Board of Regents, 13 FEP Cases 462 (1976).

Megdal, S. B., and Ransom, M. R. (1985). Longitudinal changes in salary at a large public university: What response to equal pay legislation? *AEA Papers and Proceedings* 75(2): 271–274.

Miller, S. I., and Fredericks, M. (1991). Postpositivistic assumptions and educational research: Another view. *Educational Researcher* 20(4): 2–8.

Moore, K. (1990). *Gender Differences in Faculty Salaries at the Pennsylvania State University. Report to the Strategic Study Group on the Status of Women*. University Park, PA: Pennsylvania State University.

Morrison, D. E., and Henkel, R. E. (1969). Significance tests reconsidered. *The American Sociologist* 4(2): 131–140.

Namboodiri, N. K., Carter, L. F., and Blalock, H. M., Jr. (1975). *Applied Multivariate Analysis and Experimental Design*. New York: McGraw-Hill.

Netter, J., Wasserman, W., and Kutner, M. H. (1985). *Applied Linear Regression Models*. Homewood, IL: Richard D. Irwin, Inc.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review* 14(3): 693–709.

Ottaviani v. State University of New York at New Paltz, 679 F.Supp. 288 (1988).

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

Penk v. Oregon State Board of Higher Education, 816 F.2d 458 (1987).

Pezzullo, T. R., and Brittingham, B. E. (eds.) (1979). *Salary Equity: Detecting Sex Bias in Salaries Among College and University Professors*. Massachusetts: Lexington Books.

Presseisen v. Swarthmore College, 15 FEP Cases 1466 (1977).

Ransom, M. R., and Megdal, S. B. (1989) *On the Status of Women in the Academic Labor Market in the Affirmative Action Area*. Unpublished paper.

Rosenthal, W., and Yancey, B. (eds.) (1985). The use of data in discrimination issues cases. *New Directions for Institutional Research, No. 48, XII(4)*. San Francisco, CA: Jossey-Bass.

Rothchild. Overview of pay initiatives. *Issues for the 80s*, pp. 119–128.

Scott, E. L. (1977). *Higher Education Salary Evaluation Kit.* Washington, DC: American Association of University Professors.

Senate Committee on Faculty Affairs (1990). *Faculty Salaries: Analysis of Gender, Location, and Minority Status Differences in Salaries.* University Park, PA: Pennsylvania State University.

Siegfried, J. J., and White, K. J. (1973). Financial rewards to research and teaching: A case study of academic economics. *American Economic Review* 63(2): 309–315.

Sobel v. Yeshiva University, 839 F.2d 18 (2d Cir. 1988).

Spaulding v. University of Washington, 740 F.2d 686 (9th Cir. 1984).

Striebel, C. (1989). *Differences in Salary Between Men and Women on the Faculty and Academic Staff at the University of Minnesota* (Petitioner's Statistical Report). Duluth, MN: University of Minnesota.

Strober, M. H., and Quester, A. O. (1977). The earnings and promotion of women faculty: Comment. *The American Economic Review* 67(2): 207–213.

Tabahnick, B. G., and Fidell, L. S. (1989). *Using Multivariate Statistics* (2nd ed.). New York: Harper & Row.

Tuckman, B. H., and Tuckman, H. P. (1976). The structure of salaries at American universities. *The Journal of Higher Education* 47(1): 51–64.

Weeks, K. (1985). Equal pay: The emerging terrain. *Journal of College and University Law* 12: 41–60.

Weeks, K., and Organ, J. (1986). Educational institutions and comparable worth: A doctrine in search of applications. *Journal of Law and Education* 15: 207–228.

Williams, E. J. (1959). *Regression Analysis.* New York: Wiley.