Research in Higher Education, Vol. 37, No. 6, 1996

CRITICAL MODELING PRINCIPLES WHEN TESTING FOR GENDER EQUITY IN FACULTY SALARY

William Balzer, Nancy Boudreau, Peter Hutchinson, Ann Marie Ryan, Todd Thorsteinson, James Sullivan, Robert Yonker, and Deanne Snavely

Multiple regression procedures are commonly used to investigate gender equity in faculty salary. However, a review of 19 case studies indicates that many of them fail to (a) adequately develop a regression model that examines the linear and nonlinear (i.e., interactive and curvilinear) relations between predictors and the criterion and (b) appropriately apply regression diagnostic statistics throughout salary model development. A seven-step process is presented as a comprehensive framework for testing allegations of gender discrimination in faculty salary. Steps include (a) identifying predictors of faculty salary, (b) identifying and establishing criteria for interpreting statistical tests and diagnostic procedures, (c) determining the criterion variable used in the salary model, (d) developing a salary model, (e) testing for gender discrimination in pay, (f) conducting diagnostic procedures to confirm the appropriateness of the final salary model, and (g) testing the assumptions of the regression model. An application of this model is presented using a case study (N = 725 faculty).

As the composition of faculties has changed over time, gender equity in faculty salaries has become an increasingly important policy issue on college and university campuses. In order to address this issue in a rational and appropriate manner, it is necessary to construct multivariate statistical models to prevent drawing erroneous conclusions from anecdotal perceptions or spurious bivariate relations between gender and salary. A variety of such statistical approaches has been undertaken. This paper presents the results of a case study that illustrates and emphasizes several critical modeling principles when analyzing gender equity in faculty salaries.[1]

The most common approach for testing for the presence of gender equity in

Send correspondence to: William Balzer, Department of Psychology, Bowling Green State University, Bowling Green, OH 43403. e-mail: wbalzer@bgnet.bgsu.edu.

faculty salary is multiple regression (e.g., Johnson, Riggs, and Downey, 1987; Lassiter, 1983; Moore, 1993; Ramsay, 1979), a statistical technique used to develop an explanatory model or capture an existing policy using a number of predictor variables. In a study of gender equity in salaries, salary may be used as the criterion variable and the variables that are believed to influence salary (e.g., years at the university, rank, a market salary variable) would be included as predictor variables. A regression equation is then determined, based on the least squares criterion, that estimates the contribution each predictor or set of predictors makes to salary, and these estimates can be tested for statistical significance. In this manner, one can reasonably conclude that these predictors contribute to explaining salary and are not due to chance.

Various approaches using multiple regression include the Salary Kit method (Gray and Scott, 1980; Scott, 1977), the Oaxaca method (Oaxaca, 1973), reverse regression (Birnbaum, 1985; Conway and Roberts, 1986), compa-ratio analysis (Bereman and Scott, 1991), and differential regression (Arvey, 1986; Taylor, 1988). These various regression approaches differ along a number of dimensions, including (a) specification of predictor and criterion variables, (b) modeling specifications, and (c) the evidence or criteria used to determine the presence or absence of pay discrimination (Moore, 1993).

Despite the widespread popularity of multiple regression in examining gender equity in pay, this approach is not without its critics. Several concerns regarding multiple regression have been raised. First, variables used in regression models may be imperfect proxies for the real variables of interest, and systematic bias in proxy variables (e.g., age as an index of experience) can lead to inappropriate conclusions (Barrett et al., 1986; Barrett and Doverspike, 1989). Second, the omission of important predictor variables from the model will affect the estimates of predictor variables in the model, resulting in inappropriate conclusions about the presence or absence of gender bias in pay (Barrett and Doverspike, 1989; Hengstler and McLaughlin, 1985). Third, salary models may violate the regression assumptions of linearity and homogeneity of variance (Everett, 1990; McFatter, 1987). Finally, the regression equation should consider nonlinear effects of gender on salary (i. e., the interaction of gender with other predictor variables). If not, the regression results could underestimate the total effect of gender on pay, leading to an inappropriate conclusion about the presence or absence of gender discrimination (Burkhalter et al., 1986; Smart, 1991).

Although there are many articles on how to deal with these criticisms, two points remain relatively unaddressed in the literature on gender equity and pay: (a) the development of a regression model that examines the linear, interactive, and curvilinear relations between predictors and the criterion; and (b) the appropriate application of diagnostic statistics throughout model development. Therefore, the purpose of this study is to present a case study that focuses on these two primary modeling issues and describes the procedures for designing and evaluating such a regression model to determine, in a conceptually appropriate and statistically defensible manner, whether gender discrimination in faculty salary is present. Although these two issues have received little attention in previous case studies of gender discrimination in pay, they are critically important in developing an analysis of gender discrimination that will be acceptable to, and accepted by, administrators, faculty, and other constituencies.

DEVELOPING REGRESSION MODELS TO TEST FOR GENDER INEQUITIES IN PAY: CONCEPTUAL CONSIDERATIONS

There are a number of issues that must be considered when using regression analyses to test for gender inequity in pay. First, it is critically important that the salary model identifies and considers for inclusion, where possible, all relevant predictor variables. In addition to considering commonly used predictor variables, researchers should include predictor variables that capture either special or campus-specific situations or groups of faculty where impact on salary is not captured by the commonly used variables. The omission of important predictor variables could lead to specification error, which can occur when an important predictor variable is left out of a model. When this happens, the effect of the omitted variable on the criterion variable is inappropriately attributed to any related predictor variable(s) included in the model, reducing the general ability of the model to predict salary and biasing estimates of the contributions to salary of those predictor variables in the regression equation.

Simultaneously, researchers must be careful that problems of multicollinearity are not introduced by the inclusion of predictors (Hengstler and McLaughlin, 1985). Multicollinearity will occur when a predictor variable is so highly correlated with another predictor variable, or with a linear combination of at least two other predictor variables, that the standard errors of regression coefficients are inflated. The presence of multicollinearity may cause the parameter estimates to be statistically nonsignificant, or stated alternately, may cause the confidence intervals derived from the estimates to be very large. Reducing multicollinearity will result in more precise estimates with lower standard errors, smaller confidence intervals, and greater statistical power. Of course, some degree of multicollinearity is present in all observational data (McLaughlin, Zirkes, and Mahan, 1983), but it is not commonly viewed as a problem unless certain rules of thumb are violated: (a) average pairwise correlations of .80 or higher, (b) large standard errors of regression coefficients, and (c) regression coefficients with counterintuitive signs (Taylor, 1988). Variance inflation factors (VIFs) may also be calculated to help identify this statistical problem.

Multicollinearity is suspected when the VIF (the reciprocal of $1 - R^2$, where R^2 is obtained by regressing the predictor variable in question on all other predictor variables in the model) for a predictor variable exceeds 10.0 (Neter, Wasserman, and Kutner, 1989). Moore (1993) noted, however, that generally the impact of multicollinearity in salary inequity studies is minimal. However, only 4 of the 19 case studies we located in our literature review discussed the issue of multicollinearity (Bereman and Scott, 1991; Chasin, Goldfarb, and LaNoue, 1989; Swartzman, Seligman, and McClelland, 1992; York, Henley, and Gamble, 1987) and only two (Ervin, Thomas, and Zey-Ferrell, 1984; York, Henley, and Gamble, 1987) described methods for assessing whether it was present in the faculty data examined (case studies are noted as such in reference list). Little guidance is provided on alternative solutions to resolve problems of multicollinearity short of not using those predictor variables contributing to the problem, such as the use of principal components analysis or other techniques that combine highly correlated predictors into a conceptually meaningful concept.[2]

It is also necessary to model the appropriate relationship between the predictor variables and salary in the regression equation used to predict salary. Many researchers include only linear main effects, as was the case in 9 of the 19 case studies (Baker et al., 1988; Bereman and Scott, 1991; Chasin, Goldfarb, and LaNoue, 1989; Danielson and Smith, 1981; Ervin, Thomas, and Zey-Ferrell, 1984; Fisher, Motowidlo, and Werner, 1993; Lassiter, 1983; Swartzman, Seligman, and McClelland, 1992; York, Henley, and Gamble, 1987). A linear model, however, assumes that (a) the impact of a predictor variable on the dependent variable is the same for all levels of the other predictor variables and that there is no interaction between two or more predictor variables, and (b) there is no curvilinear relation between the predictor variables and the criterion variable. The inclusion of interaction and curvilinear terms can model interaction and curvilinear effects respectively, potentially improving the specification and predictability of the regression model.

For example, the inclusion of an interaction term for gender with years of university service makes it possible to examine statistically whether the influence of years of university service is the same for males and females rather than assuming that its impact is equal for males and females. This could indicate that while males and females hired recently do not differ on salary, females hired many years ago are underpaid relative to their male cohorts. Two case studies examined gender interactions (Pfeffer and Ross, 1990; Smart, 1991); two other studies included interactions but did not include those interactions with gender (Borjas, 1983; Gomez-Meija and Balkin, 1992).

The inclusion of curvilinear effects makes it possible to test for the presence of such things as a decelerating relation between years of university service and salary. This would indicate that as years of university service increase, salary would increase up to some point, after which additional years of university service would have less impact on salary. The reason we might anticipate this diminishing return with years of service is potentially due to salary compression as years of service to the university increase. There may be some trade-offs to consider, however, because coding for these quadratic effects can introduce multicollinearity problems. This is due to the fact that the quadratic term is the value on the original predictor variable squared, and the correlation between X and X^2 is typically very high, especially in narrow ranges of X. Centering procedures in which the mean value is subtracted from each observed value of the original predictor variable can be applied to quantitative predictor variables to minimize these multicollinearity problems without changing any statistical inference decisions (Aiken and West, 1991; Neter, Wasserman, and Kutner, 1989). Five of the case studies reviewed examined or recommended examining curvilinear relations (Borjas, 1983; Ferber and Green, 1982; Haberfeld, 1992; Raymond, Sesnowitz, and Williams, 1988; Weiler, 1990).

Finally, it is important to determine whether several statistical assumptions underlying multiple regression are satisfied if valid statistical inferences are to be made (Bohannon, 1988; Hengstler and McLaughlin, 1985). The first assumption is that the error term is normally distributed; failure of this assumption would make it inappropriate to employ statistical tests of significance. That is, if the error term is not normally distributed, the computed p values used in standard statistical inference may be wrong since the test statistic may not be distributed as t or F. One case study reviewed examined the distribution of the error term (Schau and Heyward, 1987). Several diagnostic procedures are available to determine whether the assumption that the error term is normally distributed has been met, such as a visual inspection of the normal probability plot and the correlation between residuals and their expected values under normality. In addition, the Shapiro-Wilk statistic provides a statistical test of a reliable deviation of the error term from normality. If this assumption is not met, it may be possible to reexpress or rescale the criterion variable to minimize or eliminate this violation.

The second important assumption in multiple regression is that the variation of actual values of the criterion variable around the regression equation is constant at various values of the predictor variables (i.e., homogeneity or constant variance of the error term). Violations of this assumption may adversely affect the validity of any statistical tests of significance that are conducted as well as inflate the total error variance for the regression model. Schau and Heyward (1987) was the only case study that examined the homogeneity of variance. Residual analysis is a common diagnostic procedure used to visually analyze the homogeneous variance assumption across all levels of the regression equation, and Bartlett's test of equality of variances can be used to test for heteroscedasticity (i.e., nonconstant variance). Reexpression or transformation of the salary variable is one possible strategy for reducing or eliminating this violation.

A third assumption is that the model is appropriate for the observed data and the regression model has not failed to capture any systematic nonlinear relation in the data (i.e., curvilinear effects or interactions). Violations of this assumption may result in inappropriate estimates of each predictor's influence on salary and a less-than-optimal regression model. Testing for interactions and curvilinear relations between predictors of salary and actual faculty salary is highly recommended (Risher and Cameron, 1982), but only one of the case studies examined the inclusion of curvilinear and interactive effects (Borjas, 1983). Diagnostic procedures, such as examining residuals for systematic patterns and using stepwise regression procedures to test whether the addition of nonlinear effects significantly improve the regression model, may be employed.

Overall, there are several important assumptions that should be met to support the adequacy of the salary model developed to assess the presence of pay discrimination. Several researchers have noted that these assumptions are often violated in pay equity studies (e.g., Buford, Wilmoth, and Burkhalter, 1983; Schau and Heyward, 1987; Taylor, 1988). One potential explanation is that the linear regression model is generally robust to some violations of these underlying assumptions (e.g., Cohen and Cohen, 1983, p. 52). However, violations of these assumptions can bias the probabilities associated with the significance tests used to conclude whether a predictor variable of salary has a statistically significant and reliable relation with salary, potentially affecting whether faculty gender (or any other predictor) is identified as being related to salary. Thus, it is strongly recommended that diagnostic procedures and statistical tests be used to determine whether the assumptions have been violated (Belsley, Kuh, and Welsch, 1980; Neter, Wasserman, and Kutner, 1989) and, if so, the influence of this violation on the interpretation of the salary model must be carefully considered.

APPLICATION: A CASE STUDY

Background

A study was commissioned at Bowling Green State University (BGSU) in response to allegations of gender inequity in faculty salary. What follows is a description of the salary model developed to confirm, disconfirm, or modify the allegations of discrimination following the recommendations noted above. A seven-step procedure was followed: (1) identifying predictors of salary and developing a faculty salary database; (2) identifying and establishing criteria for interpreting statistical tests and diagnostic procedures; (3) determining the criterion variable used in the salary model; (4) developing the salary model; (5) testing for gender discrimination in pay; (6) conducting diagnostic procedures to confirm the appropriateness of the final salary model; and (7) testing the assumptions of the regression model. These steps are described in detail below.

Step 1: Identifying Predictors of Salary and Developing a Faculty Salary Database

The choice of predictor variables was based on a number of factors: (a) our hypothesized conceptual model of what variables should influence faculty salary; (b) faculty, university, or external market characteristics that have been used in previous research on gender discrimination in pay; and (c) legal and statistical considerations (e.g., the inadvisability of using age as a proxy measure of experience; the avoidance of multicollinearity among predictors). The inclusion of conceptually relevant predictor variables increases the likelihood that the salary model would not be misspecified (i.e., exclude important predictors of faculty salary) and thus would more adequately predict faculty salary. Academic-year salary was used as the criterion measure of interest. Predictor variables described below represented our conceptual framework that a faculty member's appointment (i.e., rank, possession of the discipline's terminal degree, membership on graduate faculty, and appointment to an endowed chair by the state's board of regents), his or her seniority (i.e., years in rank, years at the university, and experience prior to joining the university), and the external salary market (i.e., external market influence on salary, administrative appointment at hiring, appointment at the main or regional campus) should influence faculty salary. Although faculty productivity (e.g., merit pay) should be conceptually related to faculty salary, such measures were not included for reasons described below. Faculty gender was included to statistically test for discrimination in pay.

Academic-Year Salary (AYSALARY). The 1993–1994 nine-month academic-year salary was calculated for each faculty member. This salary figure excluded stipends received by faculty, and adjusted the salary of those individuals on fiscal-year contracts to a 9-month basis (.818[Fiscal-Year Salary − Stipend]).

Gender (GENDER). Faculty gender was dummy coded 1 for female and 0 for male faculty, and was verified against faculty personnel files.

Rank (RANK). The inclusion of faculty rank is one of the most controversial points in gender equity studies of salary (Gray, 1985; Barrett and Sansonetti, 1988). Rank has consistently been shown to be *the single best* predictor of salary (Osborne, 1990; Raymond, Sesnowitz, and Williams, 1988; Stacy, 1983). Stacy and Holland (1984) argue that given the typical differences in the proportions of males and females at each rank, ignoring rank more than doubles any apparent salary disparity. Scott (1977) and the AAUP Salary Kit Method (Gray and Scott, 1980), however, argue that because rank is a university-awarded honor or recognition, it may be awarded in a discriminatory fashion. One reasonable solution to this dilemma is only to include rank in the salary model if there is no evidence of discrimination in promotional practices (Baker et al., 1988; Bereman and Scott, 1991; Danielson and Smith, 1981; Ervin, Thomas, and Zey-Ferrell, 1984, Fisher, Motowidlo, and Werner, 1993; Moore, 1993; Raymond, Sesnowitz, and Williams, 1988, 1993; Riggs et al., 1986; Schau and Heyward, 1987; Swartzman, Seligman, and McClelland, 1992). A comprehensive analysis of time in rank promotion rates, and turnover rates for female and male faculty at BGSU over a 21-year period was conducted, and the available evidence was inconsistent with the hypothesis that the awarding of rank was gender biased (Balzer et al., 1995). Therefore, in the absence of evidence to the contrary, rank was included in the salary model. Faculty rank was represented with three dummy coded vectors (R1 = Professor, R2 = Associate Professor, R4 = Instructor or Lecturer; the rank of Assistant Professor served as the baseline for comparison).

Years in Rank (YRRANK). The salary model hypothesized that, other things equal, faculty in rank for a longer period of time would be compensated at a higher level than those in rank for a shorter period of time. Years in rank indicated the total number of years of full-time contractual faculty appointment at BGSU in the current rank held by the faculty member.

Possession of Terminal Degree (TERMDEG). The salary model hypothesized that faculty who possessed the terminal degree in their field would be compensated at a higher level than those not possessing the terminal degree. In some fields, a degree other than a doctorate represents the highest degree attainable in that field. Thus, possession of the terminal degree, rather than possession of the doctorate degree, is the appropriate method of determining whether a faculty member has achieved the highest academic credential available in the field. Terminal degree was coded 1 for individuals possessing the terminal degree and 0 for individuals not possessing the terminal degree.

Years at BGSU (YRBG). The salary model hypothesized that faculty who were at BGSU longer would be compensated at a higher level than those with fewer years at the university. Years at BGSU indicated the total number of years of full-time contractual faculty appointment at BGSU.

Prior Experience (PRIOREXP). The salary model hypothesized that faculty who possessed more prior experience before joining the university would be compensated at a higher level than those with less prior experience. In order to represent a faculty member's professional experience since receipt of the highest degree but prior to joining BGSU, PRIOREXP was computed by subtracting YRBG from years since degree was obtained. One potential limitation of this variable is the assumption that individuals gained work-related experiences during this interim period (as opposed to being unemployed, putting career on hold for family issues, etc.).

External Market Salary Factor (SALFAC). The salary model hypothesized that faculty in a department/specialty area with a higher salary factor would be compensated at a higher level than those in departments/specialty areas with lower salary factors. Given that external market factors influence faculty salaries at the university, several procedures were considered to include salary differences among academic disciplines in our salary model. Creating a set of dummy variables to represent each academic department is one potential approach commonly used in gender equity studies. This approach consumes a large number of degrees of freedom, however, particularly if one chooses to test for interaction effects in the salary model and tends to limit the statistical power of the model. In addition, there is the very likely possibility that there will be a large number of empty cells when testing for interactions, particularly in the smaller academic departments. This approach also assumes that departmental differences at BGSU accurately reflect real external market differences. Thus, an external market salary factor entitled SALFAC was created using national salary data provided by the National Association of State Universities and Land Grant Colleges (NASULGC).[3]

Because BGSU is a NASULGC institution, respondents in this survey represent a reasonable set of peer institutions. SALFAC represents, for a given discipline/major field, the ratio of the national average academic-year salary of full-time faculty in that discipline/major field to the national average academic-year salary for full-time faculty of all disciplines/major fields. For example, a salary factor of .96 for mathematics implies that the national average salary for faculty in mathematics is 96% of the national average salary of all faculty in all disciplines/major fields. This single continuous salary factor variable has been recommended and used in previous faculty salary studies (e.g., Raymond, Sesnowitz, and Williams, 1988; Simpson and Rosenthal, 1982). Department chairs provided the specific codes to classify all faculty within their department.

Graduate Faculty Membership (GRADFAC). The salary model hypothesized that individuals appointed to the graduate faculty would be compensated at a higher level than those not appointed to the graduate faculty given their additional responsibilities in directing master's- and doctoral-level students. Membership on the graduate faculty was coded 1 for members of the graduate faculty and 0 for faculty who were not members of the graduate faculty.

Administrative Experience (ADMIN). The salary model hypothesized that faculty with certain types of administrative experiences would be compensated at a higher level than those without these administrative experiences. Previous research (Raymond, Sesnowitz, and Williams, 1988; Regan, 1990) suggested that individuals who either previously held administrative appointments other than department chair or school director (e.g., deans) or who initially joined the university with an administrative appointment as department chair would receive higher levels of compensation due either to their administrative contributions to the university or external market forces extant at their hiring. Thus, ADMIN was coded 1 for those faculty who held a prior administrative appointment at BGSU other than department chair and for those individuals initially hired as department chairs or school directors. Those without these characteristics were coded 0.

Other. Because three individuals hired as Ohio Board of Regents Eminent Scholars represented an atypical faculty appointment with a higher starting salary, a dummy coded variable (EMINENT) was created. Faculty currently holding appointments as Eminent Scholars were coded 1, and those not holding such an appointment were coded 0. The salary model hypothesized that Eminent Scholars would be compensated at a higher level than other faculty.

Finally, BGSU includes both a main campus and a regional campus, and faculty for these two campuses may be compensated differently for a number of reasons (e.g., recruited from national versus regional labor markets, respectively). Thus, a dummy coded variable (FIRELAND) was created, with faculty holding appointments on the main campus coded 0 and faculty holding appointments on the Firelands campus coded 1. The salary model hypothesized that main campus faculty would be compensated at a higher level than those at Firelands because they are more generally recruited in national markets.

Because allegations of gender discrimination were based on 1993–1994 faculty salary data, this same time period was used in the present study. Faculty members' values for each variable in the database were based on information contained in the university's information system and individual faculty personnel files. The database was carefully checked for errors. Specifically, database information from a random sample of 5% of faculty on academic-year contracts and 50% of faculty members on fiscal-year contracts was compared with information contained in actual personnel files. No errors were found. Table 1 presents descriptive information on the criterion and predictor variables used in the salary model as well as the correlations between the variables.

Notably absent from our salary model was a measure of faculty productivity or merit. Moore (1993) noted that measures of productivity and merit are absent from almost all pay equity studies because it is difficult to distinguish between true productivity/merit (i.e., a faculty member's actual research productivity, excellence in teaching, and service to profession and university) and assessed productivity/merit (i.e., merit ratings and/or merit salary increments recommended by the individual department or program based on estimates or subjective judgments of performance). Using assessed productivity/merit as a proxy for true productivity/merit leads to several problems, some generic and some specific to BGSU (e.g., potential gender bias in the awarding of merit raises, difficulty determining the time period over which merit should be estimated,

TABLE 1. Descriptive Statistics and Correlation Matrix for Variables in Salary Model: 1993–1994 Faculty Salary Database (N = 725)

	Mean[a]	S.D.	1	2	3	4	5	6	7	8	9	10	11	12
1. ADMIN	.03[c]	.17	—											
2. EMINENT	.00[d]	.06	−.01	—										
3. FIRELAND	.05[e]	.21	.04	−.02	—									
4. GENDER	.30[f]	.46	−.06	−.04	−.03	—								
5. GRADFAC	.72[g]	.45	.02	.04	−.35**	−.16**	—							
6. PRIOREXP	2.90	4.80	.25**	.27**	.01	−.02	.05	—						
7. RANK[b]	—	—	−.15**	−.07*	.16**	.40**	−.48**	−.05	—					
8. SALFAC	.94	.15	.04	−.01	.02	−.19**	.08*	−.04	−.04	—				
9. TERMDEG	.86[h]	.34	.04	.03	−.15**	−.27**	.55**	−.10**	−.58**	.06	—			
10. YRBG	12.53	9.90	.01	−.06	−.05	−.28**	.03	−.24**	−.66**	−.04	.20**	—		
11. YRRANK	7.40	7.04	.07	−.03	−.06	−.22**	−.10*	−.10**	−.43**	−.02	.12**	.85**	—	
12. AYSALARY	47801	13992	.20**	.23**	−.12*	−.40**	.36**	.13**	−.78**	.33**	.42**	.62**	.55**	
13. LOGSALARY	10.73	.30	.19**	.16**	−.12**	−.42**	.40**	.10**	−.81**	.32**	.48**	.64**	.55**	

*p <.05.
**p <.01.
[a]Although quantitative variables were centered in the regression model to minimize multicollinearity problems with its squared term, means on the original variables are reported here.
[b]Because rank is a set of variables (R1, R2, R4), correlations with rank in the table reflect the multiple correlation with the other variable.
[c]Reflects the proportion of faculty with prior administrative experience (N = 21).
[d]Reflects the proportion of eminent scholars on the faculty (N = 3).
[e]Reflects the proportion of faculty with primary affiliation at Firelands (N = 35).
[f]Reflects the proportion of female faculty members (N = 214).
[g]Reflects the proportion of faculty holding graduate faculty membership (N = 523).
[h]Reflects the proportion of faculty holding the terminal degree in their discipline (N = 626).

differing merit reference groups and merit philosophies over time and across departments, absence of merit raises for several years due to state budget cuts). Therefore, assessed productivity/merit was excluded as a predictor of faculty salary. Excluding productivity or merit could result in a misspecified salary model, given that merit salary increments at BGSU are based on assessments of meritorious performance in teaching, research, and service. However, its exclusion from the model does not compromise the major focus of this research, which is to test whether gender discrimination in faculty pay exists. Our test of gender discrimination in pay would only be affected by the exclusion of productivity and merit measures if gender were in fact related to true productivity or merit, that is, if male faculty were on average more productive than female faculty. The assumption of a correlation between gender and productivity is inconsistent with the research literature (Bloom and Killingsworth, 1982; Green and Ferber, 1984; Pfeffer and Ross, 1990). Thus, while the omission of true productivity/merit variables may reduce the explanatory power of our salary model, it will have no effect on estimates of the influence of gender on salary (Gray, 1985, 1987; Pfeffer and Ross, 1990).

Step 2: Identifying and Establishing Criteria for Interpreting Statistical Tests and Diagnostic Procedures

Decision rules are needed to test hypotheses both about the legitimacy of allegations of gender discrimination in pay and the appropriateness of the salary model. While there is often consensus on the specific decision rules used to test hypotheses, their application should depend on their appropriateness for the database and the research hypotheses under investigation. Thus, it is important to articulate a priori the decision rules that will be used both for statistical tests and diagnostic procedures.

Appropriateness of Statistical Tests of Hypotheses. The appropriateness of tests of statistical significance for developing the salary model in general and testing for gender discrimination in particular was considered. It could be argued that the 1993–1994 faculty salary database of 725 full-time faculty is the entire population of observations. Following this logic, the obtained regression coefficients would be viewed as population parameters rather than estimates of population parameters, and any nonzero regression coefficients would reflect nonzero population parameters (Moore, 1993). Under this assumption, tests of statistical significance (drawing inferences about the population from a sample) would be inappropriate. Alternatively, it could also be argued that faculty salary databases are indeed drawn from a larger population, and any one database reflects a sample drawn from the population at one point in time (Bloom and Killingsworth, 1982; Taylor, 1988). Also, on a more theoretical level, the 1993–1994 salary outcomes can be viewed as one realization of an infinite number of realizations from the salary outcomes that might have resulted from the institutional salary decision process. Under these assumptions, tests of statistical significance would be appropriate. The latter arguments are statistically more compelling. Thus, the 1993–1994 database is viewed as one possible outcome from a larger population of outcomes across points in time, making the use of statistical tests of inference appropriate. Use of statistical significance also represents standard procedure in legal applications regarding gender discrimination (Campbell, 1984; Moore, 1993) because of this underlying statistical rationale.

Alpha Levels and One- Versus Two-Tailed Significance Tests. The alpha level for concluding that any relations observed in the database were reliable and not due to chance or sampling error must also be determined a priori. Because there was no compelling reason to use a more stringent or liberal alpha level, the traditional $\alpha = .05$ was used. Two-tailed (as opposed to one-tailed) significance tests were conducted, given that both positive and negative relations in the database would be of interest in terms of significant predictors of faculty salary (e.g., whether years of service were positively or negatively related to salary) and testing for gender discrimination in pay (e.g., whether female faculty were paid less or more than male faculty) (Ferber and Green, 1982).

Criteria for Diagnostic Procedures. Multiple diagnostic procedures are often used to evaluate the adequacy of a regression model. For example, both influence statistics (Belsley, Kuh, and Welsch, 1980; Bohannon, 1988) and visual inspection of residuals are used to diagnose the influence of outliers. Multiple indices are also examined to diagnose the presence of multicollinearity among predictors (Bohannon, 1988), including the level of zero-order correlations between predictors, variance inflation factors, and the observed versus expected sign (positive or negative) of the relations between predictor and criterion variables. What is unclear, and for which no rules of thumb are available, is whether the results of all diagnostic procedures must be consistent to conclude that outliers, multicollinearity, and so on are present or absent. Thus, professional judgment must be exercised. In all cases, a full report of all diagnostic statistics is included.

Step 3: Determining the Criterion Variable for the Salary Model

AYSALARY was the original choice for the criterion variable in our model. Preliminary analyses during model development, however, indicated that using AYSALARY as the criterion variable violated the homogeneity of variance assumption of multiple regression (see Step 7). Specifically, Bartlett's test for equality of variances was applied to five groups of faculty arranged in increasing order of predicted salary (e.g., lowest 20%, 21–40%, etc.). The statistical test was significant, $\chi^2(4) = 86.948$, $p < .000$, and results indicated that error variance increased with the size of predicted salary. AYSALARY was therefore reexpressed using the natural log transformation (LOGSALARY), as recommended and used in previous salary discrimination studies (Baker et al., 1988; Chasin, Goldfarb, and LaNoue, 1989; Raymond, Sesnowitz, and Williams, 1988; Taylor, 1988; Wetton, 1990).

For our data, Bartlett's test continued to indicate that the problem of nonconstant variance was not totally eliminated through reexpression, $\chi^2(4) = 12.959$, $p = .012$, although the logarithmic transformation did remove the systematic pattern of increasing error variance according to the value of salary. The salary model using LOGSALARY, however, yielded basically the same inferential results as the model using AYSALARY. For example, most predictor variables or interaction terms that were significant using AYSALARY were also significant using LOGSALARY, and most nonsignificant predictor variables and interaction terms using AYSALARY were nonsignificant using LOGSALARY. In both AYSALARY and LOGSALARY models, the coefficient for gender and interactions between gender and other predictor variables were nonsignificant. Although most economists would probably prefer the model using LOGSALARY, using the natural log transformation of the criterion variable changes the predictor variable coefficient to reflect an estimate of the effect of that variable on salary proportional to the value of salary (as opposed to the absolute impact of that variable on salary regardless of the value of salary when using untransformed salary).

Becker and Goodman (1991) found that when salary was reexpressed via log transformation, serious problems in interpreting the meaning of the gender coefficient occurred for those not well versed in multiple regression analysis, providing the potential for exploitation by a university when settling claims of gender discrimination. In addition, because our salary database was large (725 cases), the tests for gender bias are robust regardless of whether the original or transformed salary data are used. Therefore, because (a) statistical results using AYSALARY and LOGSALARY were inferentially similar, (b) use of LOGSALARY did not totally remove heteroscedasticity (i.e., nonconstant variance), (c) the tests for gender bias were robust to the violation of equality of variance regardless of the model chosen, and (d) transformed criterion data might be interpreted and used inappropriately, AYSALARY was retained as the criterion variable. Complete results using LOGSALARY as the criterion variable are available from the authors.

Step 4: Developing the Salary Model

Faculty Salary Model: Linear Relations Between Predictor Variables and AYSALARY. AYSALARY was regressed on the linear forms of rank (R1, R2, and R4), YRRANK, TERMDEG, YRBG, PRIOREXP, SALFAC, GRADFAC, ADMIN, EMINENT, and FIRELAND to assess their influence on the prediction of salary for both male and female faculty. Results, shown in Table 2, indicated a statistically significant relation between the linear forms of the predictor variables and AYSALARY, $F(12, 712) = 275.87, p < .001$. The adjusted R^2 index, a corrected estimation of the relation between predictors and criterion in the population, was .82, indicating that together the predictor variables used in this model predicted AYSALARY very well (i.e., 82% of the total variation in AYSALARY is explained by this set of predictors).

Nine of the 12 predictor variables made statistically significant contributions to the prediction of salary; only FIRELAND, TERMDEG, and YRBG were not found to contribute to the prediction of faculty salary. The nine statistically significant predictor variables contributed to the prediction of salary in the direction hypothesized (i.e., their parameter estimates had the hypothesized signs). For example, the salary model indicated that, holding all other predictor variables constant, faculty received $185.38 more for each year of experience prior to joining the BGSU faculty, $721.45 more for each additional year in rank, $282.17 more for each additional .01 value in SALFAC above 1.00, $2,986.87 more if a member of the graduate faculty, $4,613.40 if either previously held an administrative appointment other than department chair or ini-

TABLE 2. Salary Model Excluding Gender for 1993–1994 Faculty Salary Database: Linear Relations

Variable	Parameter Estimate	Standard Error	*p*-value	Variance Influence Factor (VIF)
INTERCEPT	38859.00	1024.40	.000	0.00
PRIOREXP[a]	185.38	59.59	.002	1.68
YRBG[a]	−58.24	73.68	.430	10.93
R1	16487.00	1085.57	.000	5.24
R2	5910.73	752.89	.000	2.48
R4	−7585.46	1143.93	.000	2.73
YRRANK[a]	721.45	74.95	.000	5.73
TERMDEG	337.37	1038.40	.745	2.62
SALFAC[a]	28217.00	1519.45	.000	1.05
GRADFAC	2986.87	714.49	.000	2.11
ADMIN	4613.40	1409.84	.001	1.15
EMINENT	37496.00	3600.29	.000	1.10
FIRELAND	1033.76	1119.73	.356	1.18

$N = 725$.
$F(12, 712) = 275.87, p <.000$.
Adjusted $R^2 = 0.82$.
[a]These continuous variables were centered as recommended by Aiken and West (1991).

tially hired as a department chair, and \$37,496 more if appointed as an Eminent Scholar. In addition, holding all other predictor variables constant, full professors earn \$16,487 more on average than assistant professors, associate professors earn \$5,910.73 more on average than assistant professors, and lecturers and instructors earn \$7,585.46 less than assistant professors (i.e., R1, R2, and R4 parameter estimates, respectively).

Faculty Salary Model: Linear and Nonlinear Relations Between Predictor Variables and AYSALARY. In addition to linear relations between the predictor variables and AYSALARY, curvilinear or interactive relations between the predictor variables and AYSALARY were hypothesized. Although testing for these interactions and curvilinear relations between predictors and the criterion is highly recommended (Risher and Cameron, 1982), it has frequently been ignored in faculty salary studies. In order to implement this approach, all quantitative variables were centered to minimize problems of multicollinearity when testing for interactions and curvilinear relations (Aiken and West, 1991). Predictor variables to test for curvilinear contributions to the salary model were created by squaring each of the three centered experience-related predictor variables (PRIOREXP, YRBG, and YRRANK). Predictor variables to test for interactions were created by using the cross products of the predictor variables (e.g., YRBG*SALFAC), again using the centered quantitative variables. Because it could not be specified a priori which nonlinear patterns between the predictor variables and AYSALARY might significantly improve the prediction of salary, a combination of hierarchical (i.e., the construction of multistage regression equations where hypothesized predictor variables are added to the salary equation based on conceptual or statistical grounds) and stepwise (i.e., the inclusion or removal of predictors in a series of steps based solely on their unique statistical contribution to the prediction of salary) regression procedures was used (Cohen and Cohen, 1983). Forcing the linear components to remain in the model, tests were conducted to determine whether, as a set, either any of the curvilinear or any of the interaction predictor variables significantly improved the salary model. Statistical levels for inclusion to, and removal from, the model were set at $\alpha = .10$. This approach allowed variables that may not have been statistically significant upon entry at the $p \leq .05$ level but were statistically significant when other variables were removed to be considered during model development. However, only those interactions and/or curvilinear predictor variables with $p \leq .05$ were included in the final model.

The three curvilinear predictor variables of experience were added to the salary model containing just the linear terms, and there was no significant improvement to the salary model, $F(3, 709) = 1.75$, $p = .152$, $\Delta R^2 = .001$. There was, however, a significant improvement to the model when interaction terms were added, $\Delta R^2 = .051$, $F(52, 660) = 5.11$, $p < .001$, indicating that at least one of the interaction terms was making a statistically significant contribution to the prediction of faculty salary. Stepwise procedures were then used at this point to determine which of the two-way interaction terms contributed significantly to the prediction of faculty salary. As shown in Table 3, 15 two-way interactions were found to make unique and statistically significant contribu-

TABLE 3. Salary Model Excluding Gender for 1993–1994 Faculty Salary Database: Linear and Nonlinear Relations

Variable	Parameter Estimate	Standard Error	*p*-value	Variance Influence Factor (VIF)
INTERCEPT	40665.00	1181.75	0.000	0.00
PRIOREXP[a]	−236.77	82.48	0.004	4.15
YRBG[a]	−51.61	75.22	0.493	14.73
R1	15893.00	1066.71	0.000	6.54
R2	6366.09	698.70	0.000	2.77
R4	−11241.00	1341.28	0.000	4.85
YRRANK[a]	427.07	99.85	0.000	13.14
TERMDEG	−1205.73	1187.10	0.310	4.42
SALFAC[a]	16392.00	3688.08	0.000	8.01
GRADFAC	1715.81	724.12	0.018	2.80
ADMIN	−738.15	1606.90	0.646	1.93
EMINENT	29737.00	3531.45	0.000	1.37
FIRELAND	4479.51	1617.01	0.006	3.20
PRIOREXP*YRBG	−71.97	13.04	0.000	9.06
PRIOREXP*R1	580.24	110.70	0.000	3.61
PRIOREXP*YRRANK	61.13	15.91	0.000	6.47
PRIOREXP*SALFAC	−1363.21	325.56	0.000	1.17
YRBG*YRRANK	−14.98	4.79	0.002	2.73
YRBG*SALFAC	−1068.92	163.16	0.000	1.37
R1*YRRANK	706.05	102.74	0.000	4.56
R2*YRRANK	229.91	83.82	0.006	2.89
R2*ADMIN	9860.49	3081.98	0.001	1.39
R4*TERMDEG	4153.74	1918.69	0.031	1.86
R4*SALFAC	−23503.00	5625.73	0.000	2.49
R4*GRADFAC	6621.25	2523.33	0.009	1.16
YRRANK*ADMIN	431.87	175.71	0.014	1.46
TERMDEG*FIRELAND	−4846.48	2031.88	0.017	3.23
SALFAC*GRADFAC	16851.00	4041.76	0.000	6.90

$N = 725$.
$F(27, 697) = 166.867, p = .000$.
Adjusted $R^2 = 0.861$.
[a]These continuous variables were centered as recommended by Aiken and West (1991).

tions ($p \leq .05$) to the prediction of faculty salary. The adjusted R^2 for this final model was .86 (i.e., this set of predictor variables accounted for 86% of the variance in academic-year salary), and the estimated standard deviation for this model was $5,219.68.

Initial review of the parameter estimates in Table 3 might seem to indicate that the signs of several of these estimates are counterintuitive. This is not the case for several reasons. First, the parameter estimates of several predictor variables are not significantly different from zero; thus, their signs reflect differences due to chance. In addition, the total influence of any predictor variable on salary must take into account the influences of the interaction terms. The parameter estimate for the linear predictor variable cannot be viewed as the total influence of that predictor variable on salary once an interaction term involving that variable is included in the model. An example of this latter case might be helpful. As indicated in Table 3, the parameter estimate for FIRELAND of +4,479.51 should not lead one to conclude erroneously that a Firelands faculty member earns $4,479.51 more than his or her main campus counterpart, holding all other predictor variables constant. One must take into account the significant TERMDEG*FIRELAND interaction, with a parameter estimate of −4,846.48. According to this interaction, holding all other predictor variables constant, Firelands faculty possessing the terminal degree earn $366.97 less than their main campus counterparts (i.e., 4,479.51 − 4,846.48), which is consistent with the hypothesized salary model. The interaction also indicates that, holding all other predictor variables constant, Firelands faculty not possessing the terminal degree earn $4,479.51 more than their main campus counterparts. This larger salary is likely due to the relatively different compositions of nonterminal-degree holders on the two campuses. On the main campus, this group consists of faculty in traditional academic disciplines who have not achieved the terminal degree. At Firelands, however, there is a much higher concentration of nonterminal-degree holders in applied fields related to two-year technical programs in which practitioners are often hired, and that have higher market-driven salaries in order to attract potential faculty away from other employment opportunities outside education.

Step 5: Testing for Gender Discrimination in Pay

Addition of Gender to Faculty Salary Model. Once the best salary model had been developed in Steps 3 and 4, the next goal was to examine whether gender discrimination in pay exists. Gender discrimination in pay would be demonstrated if the inclusion of faculty gender information in the salary model developed above contributed significantly to the prediction of faculty salary. If gender were found to explain a significant proportion of variance in faculty salary, this would suggest that gender is a factor related to salary at the university and the correct parameter estimates for the influence of the other predictor variables should be those used in the model including gender. As a first step, simultaneous multiple regression (i.e., all variables entered into the regression model at the same time) was used to regress AYSALARY both on the set of predictor variables included in the final salary model above and GENDER. The addition of GENDER did not lead to a statistically significant improvement in the salary model, $\Delta R^2 = .000$, $F(1, 696) = .25$, $p = .618$, or equivalently the parameter estimate for GENDER was not statistically different from 0, $t(696) = -.50$, $p = .618$. Thus, faculty gender was not linearly related to salary.

Addition of GENDER and Interactions Between GENDER and Other Linear Predictor Variables to the Faculty Salary Model. Tests for interactions between GENDER and the other predictor variables were then conducted in the final salary model, as has been recommended by others (Risher and Cameron, 1982). Predictor variables to test for interactions between GENDER and the significant predictor variables in the salary model (linear and interaction terms) were created (e.g., YRBG*GENDER, YRBG*SALFAC*GENDER). Forcing GENDER as well as all significant predictor variables and interactions to remain in the model, all two-way and three-way interactions with GENDER were added to the salary model. The addition of these interactions including GENDER did not result in statistically significant improvements to the salary model, $\Delta R^2 = .000$, $F(26, 671) = .845$, $p = .689$. None of the parameter estimates for any of the GENDER interactions was statistically significantly different from zero, all p's $> .05$. Thus, none of the predictors of salary appeared to operate differently for females than for males. Iterative stepwise regression procedures were also used to test whether any individual gender interactions were significant. This was done because the test of significance for the whole set of interactions may have low statistical power due to the large number of degrees of freedom consumed when this large set of terms is being tested. None of the interactions was found to be statistically significant, $p < .05$.

Step 6: Conducting the Diagnostic Procedures to Confirm the Appropriateness of Final Salary Model

Influence statistics were calculated for the final salary model. Specifically, studentized residuals were calculated for each of the 725 faculty members to determine whether any individuals were found to be outliers from the salary model regression equation. The rule of thumb used to determine whether an individual was an outlier was a studentized $t > 2.0$. Using this rule, 38 faculty were determined to be outlier observations; 28 faculty (27 males and 1 female) were positive outliers and 10 faculty (8 males and 2 females) were negative outliers. Residual analysis was also conducted, and results indicated that 53.7% of the females had negative residuals with a mean residual of $-\$130.10$ (S.D.

= \$4,184.69). Similarly, 52.6% of the males had negative residuals, with a mean residual of \$54.48 (S.D. = \$5,469.11). Individuals identified as outliers, as suggested by Becker and Goodman (1991), were examined to determine whether there was any commonality across these individuals that might indicate the salary model was misspecified (i.e., was missing an important predictor variable). This examination indicated no common identifiable characteristics that would explain their deviation from the estimated regression.

Step 7: Testing for the Assumptions of the Regression Model

Diagnostic procedures (VIFs, residual analysis) were conducted to confirm the appropriateness of the final salary model. As shown in Table 3, the VIFs for most predictor variables were less than 10, indicating minimal multicollinearity problems for those predictors. However, the VIFs of two (YRBG and YRRANK) were greater than 10, suggesting potential multicollinearity problems related to these predictors. These two variance inflation factors were 14.73 for YRBG and 13.14 for YRRANK. A high correlation between YRRANK and YRBG (r = .85) was also consistent with a multicollinearity problem. Additional analyses including and excluding each of the linear predictor variables suggested that the salary model was correctly specified when these two predictor variables were included. Specifically, the model in Table 2 was analyzed without YRRANK and then with YRRANK but without YRBG. While the standard errors for most of the variables decreased when each of these variables was left out, the exclusion of YRRANK led to a significant drop in R^2 ($F(1, 712) = 92.52$, $p < .001$, $\Delta R^2 = .02$). Because of the incremental variance accounted for by YRRANK, it was retained in the model. And while YRBG did not contribute incremental variance over YRRANK, it was retained in the model because (a) it was believed a priori to be related to salary and (b) the moderate increase in the size of the standard errors when the variables were left in was not enough to change the significance of any of the variables in the model. Therefore, YRBG and YRRANK were retained in the final salary model despite their moderate collinearity to avoid a more serious specification bias.

An examination of the residuals around the regression equation predicting faculty salary suggested that residuals were larger at higher salary levels than at lower levels, a potential violation of the homoscedasticity assumption of linear regression. Bartlett's test of the equality of variances confirmed statistically significant differences among variances along the regression line, $\chi^2(4) = 89.33$, $p = .000$. As noted earlier, efforts were attempted to reduce this potential problem. AYSALARY was reexpressed using its natural logarithm (i.e., LOGSALARY), a transformation that has been suggested to remove problems of heteroscedasticity (Neter, Wasserman, and Kutner, 1989, pp. 141, 145–149). The process of building a faculty salary model (i.e., Step 4) was then repeated using this new criterion variable. A model was fit using the same linear predictor variables as before, and stepwise procedures indicated that 14 of the 15 two-way interactions from the final model using AYSALARY were again significant (only R2*YRRANK was not statistically significant) as were three additional interactions (i.e., PRIOREXP*GRADFAC, YRBG*R4, and R4*YRRANK). The adjusted R^2 for this regression model containing linear effects and the 17 significant interactions on LOGSALARY was .883.

GENDER was then added to the model, and it failed to make a statistically significant contribution to the prediction of LOGSALARY. Iterative stepwise procedures were then used to include gender interactions; none were statistically significant. These results also supported the earlier tentative conclusion that there was no evidence of gender inequity in faculty salary. However, Bartlett's test continued to indicate significant differences in variances across salary levels, although there was no systematic pattern to these differences across salary levels. Diagnostic procedures indicated that five of the VIFs for the model using LOGSALARY were greater than 10, ranging from 11.01 to 31.92, whereas in the model using AYSALARY only two of the predictors had VIFs greater than 10. Because of the more serious concerns about multicollinearity, the inability to totally eliminate problems with heteroscedasticity, and the potential difficulty in explaining the meaning of parameter estimates to those not well versed in multiple regression analysis when LOGSALARY is used as the criterion variable, the salary model using AYSALARY was retained.

Finally, residuals from the salary model using AYSALARY were also examined to assess whether the regression model's assumption of a normally distributed error term was met. Several procedures were used, following the recommendations of Neter, Wasserman, and Kutner (1989). A visual examination of the normal probability plot indicated a linear relation (i.e., between quantile scores for the actual residual values plotted against quantile scores for the standard normal distribution), which is consistent with the assumption of normality. Second, the correlation between residuals and their expected values under normality was .99, surpassing Neter et al.'s rule of thumb that a correlation of .90 or greater is consistent with the assumption of a normally distributed error term. Finally, the Shapiro-Wilk statistic indicated that the error term was not normally distributed, $W = .98$, $p = .001$. Neter et al., however, suggest that this violation of normality is likely to occur when the homoscedasticity assumption is violated (as noted above). In addition, the obtained Shapiro-Wilk statistic of .98 is very close to the value of 1.0, which would indicate perfect normality in the data; the statistically significant difference may therefore be due to the large sample size in our study, causing small (practically speaking) differences to be statistically significant. Overall, the pattern of results is inconclusive as to whether the assumption of a normally distributed error term was violated. In light of this, and given the general robustness of the regression model to violations of the normality assumption, it was concluded that there was no reason for concern regarding normality. In the model using LOGSALARY both the correlation between residuals and their expected values and the Shapiro-Wilk statistic supported normality.

As a final evaluation of the robustness of the salary model before adding gender, backward stepwise regression procedures (i.e., all the predictor variables are entered simultaneously and the one making the smallest contribution is eliminated) were used to determine if a different set of two-way interactions would be selected. All main effects were forced to remain in the AYSALARY model. Using backward stepwise regression, 14 of the 15 two-way interactions selected using the iterative stepwise procedure were statistically significant ($p \leq .05$) as were five additional two-way interactions. The adjusted R^2 value for this new model was .86, and the estimated standard deviation for the model was \$5172.19, not much different from the previous model. Seven of the 31 terms (main effects and the significant two-way interactions identified using backward stepwise regression) in the salary model, however, had VIFs greater than 10 (ranging from 11.28 to 62.72), including three of the five new two-way interaction terms. Because this backward stepwise regression model appeared to have more severe problems with multicollinearity compared with our working model, and there was little improvement in the adjusted R^2 and the estimated standard deviation using this salary model, the originally developed model was retained.

Summary

Multiple regression procedures were applied to a faculty database to develop a model that best predicted faculty salaries. The final salary model included interactions between predictor variables as well as linear relations between the predictor variables and academic-year salary. Faculty gender, or the interactions of faculty gender and other predictor variables, did not improve the prediction of salary. Diagnostic procedures also indicated that the model adequately and appropriately described the university's current salary policy. These analyses failed to support the allegation of systematic gender discrimination in faculty salary at BGSU.

CONCLUSIONS

Gender equity continues to be an important issue on university campuses. The presence of discrimination in faculty salary due to gender (or race, ethnicity, age, religion, etc.) is both illegal and abhorrent, and reflects negatively on the university and faculty. It is important to investigate all allegations of discrimination in pay, and if these allegations are well founded, to redress this discrimination. If support is not found for these allegations, it is important that the process and results used to reach this conclusion be well publicized and available for review. The methods and statistical analyses used to test for the presence of discrimination must be consistent with professional guidelines and practice and allow for an objective and credible evaluation of the issue.

In order to accomplish these goals, it is essential that appropriate modeling techniques be implemented to prevent drawing inappropriate conclusions regarding this critical issue. The case study presented here demonstrates a modeling approach based on multiple regression analysis. Particular attention has been drawn to the necessity of analyzing both the linear and nonlinear effects of predictor variables when testing for gender bias in salary. A second focus has been the appropriate use of statistical tests and regression diagnostics to assure that the underlying regression approach is appropriate for the faculty data set under analysis. Attention to these two principles is essential to assure proper inferences are drawn from the gender equity analysis and to assure its credibility on campus.

NOTES

1. Because our research focused on gender equity in faculty pay and not the development of a model of faculty pay, this work does not specifically address the broader economic issue of differences in male and female pay or external market factors on faculty pay. There is considerable literature on pay, both within and outside academic settings, that addresses the influence of external markets, internal decisions, and other factors outside any one institution's control (e.g., Barbezat, 1988; Bowen and Sosa, 1989; Clotfelter et al., 1991).
2. The authors thank an anonymous reviewer for this suggestion.
3. This classification scheme was done with one minor exception. CIP codes contained in the NASULGC survey were too broad to represent all specialty areas within the College of Business Administration accurately, and in some cases were not able to differentiate fields in the College of Business Administration from similar fields in colleges of arts and sciences and colleges of law at other NASULGC institutions. Therefore, salary data gathered in the same manner by the American Assembly of Collegiate Schools of Business (AACSB) from university business schools were used. These data were obtained from 71 of the 77 universities that had responded to the NASULGC survey. Therefore, the data were comparable and more accurately represented the appropriate CIP codes or market factors of the business faculty at BGSU.

REFERENCES

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Arvey, R. D. (1986). Sex bias in job evaluation procedures. *Personnel Psychology* 39(2): 315–336.

*Baker, D. E., Lattin, D. L., Siddons, L. M., and Thomas, J. (1988). Report of the faculty services subcommittee: Study of male/female faculty salary differences. *American Journal of Pharmaceutical Education* 52: 421–424.

*Case studies.

Balzer, W., Boudreau, N., Ryan, A. M., Snavely, D., Sullivan, J., and Yonker, R. (1995). Gender equity in faculty salary at Bowling Green State University: A report based on 1993–1994 faculty salary data. Unpublished document, Office of the President, Bowling Green State University, Bowling Green, OH.

Barbezat, D. (1988). Gender differences in the academic reward system. In D. W. Bereman and T. I. K. Youn (eds.), *Academic Labor Markets and Careers*. New York: Falmer Press, pp. 138–164.

Barrett, G. V., and Doverspike, D. (1989). Fundamental errors and erroneous beliefs in using regression analysis as a test for sex discrimination. *Law and Psychology Review* 13: 1–24.

Barrett, G. V., and Sansonetti, D. M. (1988). Issues concerning the use of regression analysis in salary discrimination cases. *Personnel Psychology* 41: 503–515.

Barrett, G. V., Alexander, R. A., Anesgart, M. N., and Doverspike, D. (1986). Frequently encountered problems in the application of regression analysis to the investigation of sex discrimination in salaries. *Public Personnel Management* 15: 143–157.

Becker, W., and Goodman, R. (1991). The semilogarithmic earnings equation and its use in assessing salary discrimination in academe. *Economics of Education Review* 10(4): 323–332.

Belsley, D. A., Kuh, E., and Welsch, R. R. (1980). *Regression Diagnostics*. New York: John Wiley & Sons.

*Bereman, N. A., and Scott, J. A. (1991). Using the compa-ratio to detect gender bias in faculty salaries. *Journal of Higher Education* 62: 556–569.

Birnbaum, M. H. (1985). Relationships among models of salary bias. *American Psychologist* 40: 862–866.

Bloom, D. E., and Killingsworth, M. R. (1982). Pay discrimination research and litigation: The use of regression. *Industrial Relations* 21: 318–339.

Bohannon, T. R. (1988). Applying regression analysis to problems of institutional research. In B. D. Yancy (ed.), *Applying Statistics in Institutional Research*. San Francisco: Jossey-Bass, pp. 43–60.

*Borjas, G. J. (1983). The measurement of race and gender wage differentials: Evidence from the federal sector. *Industrial and Labor Relations Review* 37: 79–91.

Bowen, W. G., and Sosa, J. A. (1989). *Prospects for Faculty in the Arts and Sciences*. Princeton, NJ: Princeton University Press.

Buford, J. A., Wilmoth, J. N., and Burkhalter, B. B. (1983). Auditing the compensation function for sex-based salary differences: Some needed refinements. *Compensation Review* 15: 33–41.

Burkhalter, B. B., Wilmoth, J. N., Essex, N. L., and Buford, J. A. (1986). Auditing the compensation function for race- and sex-based salary differences: Further needed refinements. *Compensation and Benefits Review* 18: 35–42.

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.

*Chasin, S. H., Goldfarb, M. G., and LaNoue, G. R. (1989). An exploration of gender and other influences on the earnings of pharmacy school faculty. *American Journal of Pharmaceutical Education* 53: 1–6.

Clotfelter, C. T., Ehrenberg, R. G., Getz, M., and Siegfried, J. J. (1991). *Economic Challenges in Higher Education*. Chicago: The University of Chicago Press.

Cohen, J., and Cohen, P. (1983). *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Conway, D. A., and Roberts, H. V. (1986). Regression analysis in employment discrimination cases. In M. H. DeGroot, S. E. Fienberg, and J. B. Kadane (eds.), *Statistics and the Law*, pp.107–168. New York: John Wiley & Sons.

*Danielson, J. L., and Smith, R. (1981). The application of regression analysis to equality and merit in personnel decisions. *Public Personnel Management Journal* 10: 126–131.

*Ervin, D., Thomas, B. J., and Zey-Ferrell, M. (1984). Sex discrimination and rewards in a public comprehensive university. *Human Relations* 37: 1005–1025.

Everett, J. E. (1990). Discrimination measures using contingency tables. *Multivariate Behavioral Research* 25: 371–386.

*Ferber, M. A., and Green, C. A. (1982). Traditional or reverse sex discrimination? A case study of a large public university. *Industrial and Labor Relations Review* 35: 550–564.

*Fisher, B D, Motowidlo, S., and Werner, S. (1993). Effects of gender and other factors on rank of law professors in colleges of business: Evidence of a glass ceiling. *Journal of Business Ethics* 12: 771–778.

*Gomez-Meija, L. R., and Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal* 35: 921–955.

Gray, M. W. (1985). Legal perspectives on sex equity in faculty employment. *Journal of Social Issues* 41: 121–134.

Gray, M. W. (1987). The halls of ivy and the halls of justice: Lightening the burden. *Academe* 73: 61–63.

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

Green, C. A., and Ferber, M. A. (1984). Employment discrimination: An empirical test of forward versus reverse regression. *Journal of Human Resources* 22: 557–569.

*Haberfeld, Y. (1992). Employment discrimination: An organizational model. *Academy of Management Journal* 35: 161–180.

Hengstler, D. D., and McLaughlin, G. W. (1985). Statistical issues and concerns in court cases. In W. Rosenthal and B. Yancey (eds.), *The Use of Data in Discrimination Issue Cases*, pp. 65–82. San Francisco: Jossey-Bass.

*Johnson, C. B., Riggs, M. L., and Downey, R. G. (1987). Fun with numbers: Alternative models for predicting salary levels. *Research in Higher Education* 27: 349–362.

*Lassiter, R. L., Jr. (1983). The development and use of a faculty salary model for a higher education system. *Research in Higher Education* 18: 333–358.

McFatter, R. M. (1987). Use of latent variable models for detecting discrimination in salaries. *Psychological Bulletin* 101: 120–125.

McLaughlin, G. W., Zirkes, M. B., and Mahan, B. T. (1983). Multicollinearity and testing questions of sex equity. *Research in Higher Education* 19: 277–284.

Moore, N. (1993). Faculty salary equity: Issues in regression model selection. *Research in Higher Education* 34: 107–126.

Neter, J., Wasserman, W., and Kutner, M. H. (1989). *Applied Linear Regression Models*, 2nd ed. Homewood, IL: Irwin.

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

Osborne, J. S. (1990). A strategy towards accountability: Identifying faculty salary policies at Kentucky public regional institutions. Unpublished doctoral dissertation, University of Kentucky.

*Pfeffer, J., and Ross, J. (1990). Gender-based wage differences: The effects of organizational context. *Work and Occupations* 17: 55–78.

Ramsay, G. A. (1979). A generalized multiple regression model for predicting college faculty salaries and estimating sex bias. In T. R. Pezzullo and B. E. Brittingham (eds.), *Salary Equity: Detecting Sex Bias in Salaries Among College and University Professors*, pp. 37–53. Lexington, MA: Lexington Books.

*Raymond, R. D., Sesnowitz, M. L., and Williams, D. R. (1988). Does sex still matter? New evidence from the 1980s. *Economic Inquiry* 26: 43–58.

Raymond, R. D., Sesnowitz, M. L., and Williams, D. R. (1993). Further evidence on gender and academic rank. *Quarterly Review of Economics and Finance* 33: 197–215.

Regan, C. E. (1990). Faculty salary differences at selected institutions of the State University of New York. Unpublished doctoral dissertation, SUNY Albany.

Riggs, M. L., Downey, R. G., McIntyre, P. E., and Hoyt, D. P. (1986). Using discriminant analysis to predict faculty rank. *Research in Higher Education* 25: 365–376.

Risher, H., and Cameron, M. (1982). Pay decisions: Testing for discrimination. *Employee Relations Law Journal* 7: 432–453.

*Schau, C. G., and Heyward, V. H. (1987). Salary equity: Similarities and differences in outcomes from two common prediction models. *American Educational Research Journal* 24: 271–286.

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

Simpson, W. A., and Rosenthal, W. H. (1982). The role of the institutional researcher in a sex discrimination suit. *Research in Higher Education* 16: 3–26.

*Smart, J. C. (1991). Gender equity in academic rank and salary. *Review of Higher Education* 14: 511–526.

Stacy, D. R. (1983). Handling faculty claims of sex discrimination in pay. *Employee Relations Law Journal* 9: 348–357.

Stacy, D. R., and Holland, C. L. (1984). Legal and statistical problems in litigating sex discrimination claims in higher education. *Journal of College and University Law* 11: 107–177.

*Swartzman, L. C., Seligman, C., and McClelland, W. J. (1992). Detecting gender discrimination in university salaries: A case study. *Canadian Journal of Higher Education* 22: 1–13.

Taylor, S. H. (1988). Conceptual and methodological issues in the assessment of sex-related pay discrimination. Unpublished doctoral dissertation. University of Minnesota.

*Weiler, W. C. (1990). Integrating rank differences into a model of male-female faculty salary discrimination. *Quarterly Review of Economics and Business* 30: 3–15.

Wetton, C. E. (1990). Salary differentials of men and women in the British Columbia service: A human capital approach. Unpublished master's thesis, University of Victoria.

*York, R. O., Henley, H. C., and Gamble, D. N. (1987). Sexual discrimination in social work: Is it salary or advancement? *Social Work* 32: 336–340.

Received July 20, 1995.