2-1998

Gender-Based Pay Gaps: Methodological and Policy Issues in University Salary Studies

Myra Marx Ferree
University of Connecticut, mferree@ssc.wisc.edu

Julia McQuillan
University of Nebraska-Lincoln, jmcquillan2@Unl.edu

Published in *Gender & Society* 12:1 (February 1998), pp. 7–39.

Gender-Based Pay Gaps: Methodological and Policy Issues in University Salary Studies

Myra Marx Ferree and Julia McQuillan
University of Connecticut

Abstract

Methodology is often a point of contention in gender-based salary studies. Although this debate seems at first to be merely about technical issues, it also has an important conceptual dimension. We argue that there are two competing implicit conceptions of discrimination, one institutional and the other individual, that underlie many such debates. We first contrast the preferred methodologies advanced by each side, the policy capturing approach and the flagging approach, and explore the theoretical meaning of their statistical models. We then describe a practical application of both methodological approaches in one specific salary inequity study. In conclusion, we reflect on the implications of such practical statistical choices, discuss how such models can be combined, and make suggestions for sociologists who act as statistical experts or work with them in gender-based salary inequity studies on their own campuses.

Over 100 studies have evaluated possible gender-based discrimination in salaries at academic institutions in the United States (Gray 1990; Haignere et al. 1996; Scott 1977). There are generally two opposing sides interested in these discrimination studies, a faculty committee or union representing employees that believes gender-based discrimination exists, and an administration team that often believes none exists. A central issue for both sides involves the choice of method for determining the existence of a possible gender-based pay dif-

The authors appreciate the considerable help provided by participants as well as reviewers of the article in various forms. In particular; we thank Nancy Day Adams, Mark Abrahamson, Nancy Breen, Christine Bose, Leslie Cutler, Paula England, Gloria A. Gronowicz, Mary Casey Jacob, Timothy Killeen, Larry Pittman, Natalie Sokoloff, Wendy Soneson, Carole Turbin, Wayne Villemez, Dena Wallerson, and Sue Williams, as well as the entire Working Group on the Status of Women Faculty and the Joint Dental and Medical Councils Committee on Women's Pay Equity. We appreciate the willingness of all the participants in this study to be rigorously reflective and to share this process with others in the interest of improving gender equity at other institutions as well. However; the interpretations and views expressed in this article are those of the authors alone and do not represent those of the administration or the committee.

Corresponding author — Myra Marx Ferree, Department of Sociology U-68, University of Connecticut, Storrs, CT 06269-2068; email: ferree@uconnvm.uconn.edu

ferential (summarized in Gray 1990). In this article, we argue that these methodological debates have an important conceptual dimension. To illustrate our case, we draw on a salary study that was conducted at a state university medical and dental school.

This salary study, like many nationwide, came into being as a result of discussions between faculty and administration. The school's vice president established a formal committee of women faculty and charged them with ascertaining if gender-based inequality in pay existed at this institution. Both faculty and administration agreed from the start that the issue was inequity, or discrimination in the broad sense, whether or not it was actually illegal. The goals endorsed by both faculty and administration were to discover if there was a problem in setting pay, whether conscious or unconscious, and if so, to remedy it. Later, disagreement arose between this committee and the administration over several specific aspects of the study, particularly the most appropriate methodological approach for identifying the existence and nature of any such problem. We suggest that one underlying tension was between competing, largely implicit understandings of how pay inequities arise and consequently how research should be appropriately designed to evaluate them. We refer to the two approaches as "institutional" and "individual."

The committee assumed a model of salary inequity that operates at the institutional level. The institutional approach conceptualizes wage discrimination as a gender stratification process that is not only historical but also pervasive and ongoing (Acker 1989; Clayton and Crosby 1992; Connell 1987). As a theory, the institutional approach assumes that salary setting processes for men incumbents are the institutional norm, women having historically been excluded from the system and thus being "added in" after the fact. Paying women less in the same jobs was completely legal until only 30 years ago; when women were hired in conventionally men's occupations, their wages as a class were systematically lower. The acceptability of paying women less remains an implicit social norm. Moreover, the institutional model sees all specific salary-paying organizations as being part of a labor market that as a whole still discriminates against women, leaving women in a less strong market position than men when negotiating salary with their organization, which can then benefit by continuing to pay women less. An institutional model assumes that no individual prejudice or hostility by a supervisor to an individual woman employee is necessary to pay her less, systematically, than a similarly skilled and experienced man in a comparable job (Deaux and Major 1987).

Institutional inequity is often hard to see in everyday interactions. When something is institutionalized, it becomes part of the shared expectations about who deserves what and who should be responsible for specific positions. These shared expectations are eventually built into established patterns of behavior; in other words, they become institutionalized as simply the way things are done (Benokraitis and Feagin 1995). No conspiracy or policy to pay women less is necessary. Identifying inequities at the institutional level requires understanding the salary setting process as it is applied to men in the organization as a whole and establishing if, when, and how women's salaries as a group are

systematically set differently. Addressing institutionalized inequity involves taking proactive steps to change institutional culture and functioning.

The administration appeared to us to assume a model of inequity that operates at the individual level. This approach conceptualizes discrimination as the result of isolated prejudiced individuals in positions to decide salaries who deliberately choose to pay individual women less than they abstractly merit. In the vice president's own words, he wanted to know, "Did I have a bad actor here? And if so, I wanted to get that person out of a leadership role." In this conceptual schema, some percentage of the income the individual woman has earned is being arbitrarily withheld "just because she is a woman" by an individually prejudiced administrator or supervisor, the "bad actor" in the above quote. To model what an individual woman "really earned" while "removing" her individual gender, this approach seeks to identify either an individual man or a group of men who are as close as possible to doing exactly what she does in as similar a setting as can be determined. Once the comparable man or men can be found, a "true" value of her earnings can be estimated and the extent of her underpayment, if any, measured. In general, the individual model assumes that such discrimination would be rare rather than systematic, being a reflection of individual prejudice (the "bad actor") rather than of institutional functioning.

The individual approach also reflects the managerial demands on administrators. As managers, higher-level administrators can more readily identify and sanction "bad judgment" by individual lower-level decision makers than address the impact of widespread practices within the institution. Smaller, more internally homogeneous groups are also more comprehensible, and literally more "manageable" to administrators who have oversight over large, complex educational institutions. For those administrators who think of prejudice-and-discrimination as a single phenomenon and are thus committed to an individual model at the conceptual level, to apply the institutional approach seems to suggest that they believe all of their subordinates are sexist. To them, the across-the-board remedy that the institutional model offers may seem like a "bad performance review" applied indiscriminately rather than to specific "bad actors." As managers, they want to control decision making over issues like salary and to apportion rewards (raises) and blame (for biased decisions) selectively. They also have incentives to keep the cost of any potential salary settlement as low as possible.

There have long been debates between these two models at the conceptual level (see, e.g., Gray 1993). Institutionalists argue for the importance of understanding sexism (and racism) in defining job queues for occupations (Reskin and Roos 1990), structuring patterns of power and influence in corporations (Cockburn 1991 ; Kanter 1977), and systematically influencing the wages and salaries thought to be appropriate for specific jobs and classes of workers (Acker 1989; Bielby and Baron 1986; Jacobs and Steinberg 1990). Defenders of letting market forces alone determine salaries see discrimination merely as an "imperfection" reflecting individuals who are prejudiced, and argue instead that most or all systematic differences reflect real differences in market value

between individuals' human capital, productivity, or occupational supply and demand (e.g., Becker 1985; Filer 1983; Polachek 1985). Neoclassical economists often believe strongly that markets do eliminate discrimination in all but rare instances and are theoretically committed to the individual model.

In this article, we take the conceptual debate and examine it at the level of social statistics and show how the choice of methodology in applied work on salary discrimination in academia reflects such theoretical positions, however poorly articulated they may be by the contending parties (cf. Blalock 1991). In doing so, we hope both to illuminate a struggle in which many of us, as academics, have a personal stake and to indicate the dangers and opportunities we have when we act as statistical experts in such controversies.

In the first part of the article, we contrast the preferred methodologies advanced by each of the conceptual sides and explore the theoretical meaning of the two statistical models in the abstract. In the second part, we describe a practical application of both methodological approaches, by providing a case study of a salary equity study that illustrates the types of models and arguments used. We conclude with a section that raises some broader reflections on our experience and make practical suggestions for sociologists who are enlisted as "statistical experts" or who listen to such experts in gender-based salary studies on their own campuses.

We offer the specific case study not as a "finding of discrimination," but as an illustration of a collaborative process of addressing pay inequity.[1] "Discrimination" is a legal concept that is constructed by a changing constellation of state and federal laws and judicial interpretations. Had we been faced with the problem of advancing a statistical case through the courts, we would have had to tailor our presentation of evidence to the legal standards of "proof." There are three legal standards for a finding of illegal discrimination. "Evil motive" demands direct evidence of discriminatory intent (which has become rare), and "disparate treatment" involves showing the application of different standards of evaluation or reward to "similarly situated" individuals. The third, "disparate impact," requires evidence that, although the same standard was applied to "members of a protected class" (those groups named in and covered by Title VII of the Civil Rights Act or another relevant piece of specific legislation) as to others, the impact of the standard used was disproportionately negative for members of the protected class *and* its use was not a business necessity. These legal standards of *evidence* should not be confused with deeper theoretical understandings of the operation of the *processes* that give rise to such evidence. When we speak of the institutional and individual model, we are referring to these underlying theoretical conceptualizations, not the evidence that courts might consider relevant to each legal standard. What the courts might see as disparate treatment by the organization as a whole (and respond to women as a group in a class-action suit) would be conceptualized here as an institutional model, while the adherents of the individual model would see disparate treatment only in the form of discriminatory outcomes concentrated among certain individuals within the organization (not a class as a whole).

To date, U.S. appellate courts have not held that failure to pay equally for comparable work (rather than for identical work) constitutes illegal pay discrimination (cf. England 1992, chap. 5). However, in the case we study, the state legislature had previously passed a law providing comparable worth settlements to state employees when a joint union-management study identified pay inequities and developed a joint plan for redressing them. Because the faculty we study were not unionized, they did not fall directly under the provisions of the law, but both faculty and administration to some degree used the earlier comparable worth studies and settlements with unionized employees as guidelines for what the process in this case should be. The controversies that arose when re-applying the model developed for a unionized faculty to this new setting clarified for us the differences in thinking about inequity that we are calling the institutional and individual model. These two models are, of course, ideal types rather than rigid dichotomies. However, such underlying and implicit theory helps to explain some of the ways in which methodological disagreements are not just about statistical approaches.

The Methodological Debate

Some issues are, indeed, narrowly methodological. Multiple regression analysis is widely used for studying social inequality (see, e.g., Blalock 1991). Up to a certain point, both sides in salary discrimination disputes agree about this methodology: Each finds appropriate multiple regression models built on all available data that might "legitimately" predict salary. In typical academic salary studies, the administration and the complaining faculty group jointly designate a list of factors potentially relevant to salary for which data exist, among which are rank, department, time at the institution, and years since degree. In this list, both sides usually agree to include rank and department but not productivity measures. These decisions put limits on the results, but they are probably necessary if the goal is to create usable models (Haignere et al. 1996; Steinberg and Haignere 1987).

Rank can be problematic as a variable. On the one hand, it is necessary to include rank in a statistical model in order to ensure that differences in pay attributable to women being more junior (in the sense of being better represented in more recent cohorts of PhDs and thus appropriately in lower rank) are not attributed to gender itself. On the other hand, rank and pay discrimination may be variably related to each other. In many institutions, rank is set by different decision makers than pay and including rank may be capturing various effects of promotion discrimination on women: Women's pay may be unfairly low but not seen by the model as such if their experience and achievements go unrecognized in both promotion and pay, or women's pay may seem to be high for their rank if they get routine raises but are especially less likely to be promoted to a higher rank when their age and achievements would predict promotion (Gray 1990).[2] In either case, these various effects of rank on pay by gender require a separate study of promotions, which we—like most institutions—did not have adequate data to conduct. Because we are focused only on salary ineq-

uity, it is appropriate to include rank. However, including rank as a variable in the salary model produces conservative estimates of discrimination whatever specific regression model is employed (cf. Bellas 1994).

"Productivity" is always a controversial element in such studies. The additional time, expense, and effort necessary to gather information about scholarly productivity, such as publications and research grants, is daunting in itself (and can be frequently used as an excuse for why no credible study can be done at all). Even if full vitae were available and coded, the challenge inherent in deciding the worth of different journals, books versus articles, and types and amounts of grants is contentious within departments and probably not feasible across specialties. Scholarship is also not the only type of productivity that is rewarded, and measures of teaching and service contributions are also poorly developed. Pilot studies by Gray and Scott (1980) on salary data suggest that the variables already typically included in the model capture much of the relationship between productivity and salary and that specific measures of productivity did not significantly improve their models. In any case, few administrations have such data quantified, making the inclusion of a variable measuring productivity in either type of model moot.

Finally, all regression models include some sort of control for departments or groups of departments. The size and composition of these groupings is often controversial, as we discuss below, because the grouping process inevitably demands a judgment as to what academic work is "the same" or different. Departments themselves are not homogeneous in the type of work performed (e.g., qualitative sociology of the family or quantitative analysis of class stratification), and similar specialties may be found in more than one department (e.g., social psychologists in both psychology and sociology departments). Grouping into categories larger than departments is usually necessary on pragmatic grounds, to avoid having no women at all in some categories or ranks (making it impossible to calculate one or more coefficients). At one extreme, a comparable worth approach might argue that all faculty at an institution are doing comparable work and all groupings should be avoided lest the number of women in a group bring down the average pay for that group. At the other extreme, all groups might be resisted on the argument that every individual's job is specifically different in training and market position, making all statistical analysis *a priori* impossible. Other research has shown that the percentage of women in a department may depress salaries, but its effects as a variable also must be separated from the clustering effects that reflect market and human capital differences between departments (Bellas 1997). Because differences between departments or other groups (such as medical specialty) are known to be due not only to the numbers of women in them, excluding all such grouping would undermine the validity of the study. The specific nature of the groups constructed, as we will see below, reflects the theoretical approach taken.

More than merely methodological disagreements arise in relation to the specific nature of the regression models preferred by each side. In these instances, the statistical disagreements are reflections of more fundamental assumptions

about the processes generating inequity. In the following sections we elaborate the methodological and theoretical positions involved in advocating each model.

The Institutional Model

The institutional approach uses multiple regression in a way designed to detect a pattern of salary inequity at the institution as a whole (Ferree and Killeen, reported in Geetter 1988). The basic model is recommended by the American Association of University Professors (Gray 1990) and is also briefly described in Norusis (1988, 175). As we will explain below, the residuals of this regression approach can also be used to test the plausibility of the individual model. We call this in general the "policy-capturing approach." This policy-capturing approach uses the shared list of salary-relevant factors to construct the best-fitting regression possible to predict *men's* salaries, taking the variation of residuals around the mean to reflect the influence of unmeasured variables (research productivity, impact as an undergraduate teacher, service and visibility on campus, administrative favoritism, and other unmeasured factors) on salary.[3] Even with a well-fitting regression model (that is, one that explains 70-80 percent of the variance in men's salaries), a considerable residual in dollar terms remains evident in the distribution of individual men's residuals around the men's mean (which is, by the mathematics underlying regression, always zero).

When a policy-capturing regression is found that is the best-fitting and most appropriate model for men's salaries, the regression coefficients thus obtained are applied to the women's data and used to predict women's salaries. After fitting the women to the men's model, the means and shapes of the residuals obtained are then examined. The difference between the men's mean (zero) and the women's mean (positive or negative) indicates the gap between what women as a group are paid and what they would have if they received the same value as the men for their salary-relevant characteristics. In the institutional model, men constitute the "norm" in male-sex-typed occupations such as university professor. This does not mean that men's salaries are "fair" in some absolute sense; historically gendered processes of salary setting no doubt inflated men's salaries (by restricting competition, etc.) as well as suppressing those of women. As men still constitute the bulk of the faculty (in this case, 71 percent), it makes theoretical sense from the perspective of this model to treat them as the institutional norm.[4] Reversing the regression models—fitting men's data to a women's model—and comparing the resultant salary differential usefully indicates the magnitude of the interaction between membership in the group and rates of return to members of that group (Jones and Kelley 1984). When there is no such interaction, the two models are equivalent and would yield the same results. When the rates of return differ and implicitly one prefers raising women's rates of return to those typical of men, as in this case, Jones and Kelley argue (p. 340) that fitting women to the men's model is more theoretically appropriate.

The institutional understanding of gender-based discrimination identifies the salary gap found by fitting women to the men's model as the amount of gender-based salary discrimination between women and men and, assuming the value is negative and statistically significant, considers this the gender-based pay gap. However, if the analysis stops at this point and only looks at the means, it may be seriously misleading. Such average differences do not tell the whole story. As Gray (1990) argues, it is very important to carefully examine the spread of the residuals. If the assumptions of multiple regression are met, the men's residuals will be approximately normally distributed, spreading evenly above and below the average in a bell-shaped curve. Women's residuals provide further, crucial information.

One way to examine the merits of an institutional explanation of discrimination is by examining the shape of the women's residuals. If the average difference between men's and women's residuals is explained by some individual women being grossly and unfairly underpaid—as the individual model suggests would be the case—there will be a bimodal curve, with a second "hump" near the tail end of the negative residuals.[5] Such a "hump" would represent a cluster of women with abnormally low salaries who pull down the average for the whole group, most of whom earn for their salary-relevant characteristics about what men do (see Figure 1). Which particular women are part of the "cluster" that is experiencing discrimination and which women are part of the normally expected distribution of negative residuals cannot, of course, be deduced solely from the curve. Rather than identifying just those individual women who have exceptionally large negative residuals as the problem cases, however, the logic of the institutional model suggests looking for underlying patterns that may produce this clustering: Are women in the "hump" at the bottom, for example, disproportionately likely to be members of a certain department or school? If so, the possibility raised by the individual model that there is a "bad actor" or prejudiced administrator setting salaries in that department should not be discounted. But the women whose salaries should then come under closer scrutiny should also include those in that same school or department who may fall at the mean or above in the overall distribution of residuals, since they might well deserve high positive residuals on the basis of their scholarly productivity or other unmeasured factors that were not being appropriately rewarded by the possibly prejudiced administrator. In some cases, a legitimate factor may be recognized and added to the model (for example, the interaction of having a certain type of specialty in a certain type of department) that would eliminate the clustering effect at the bottom.

Another possible shape anomaly, one that indicates institutional discriminatory practices, would be a "glass ceiling effect" (e.g., no high-end residual tail, or a sharply truncated one, for women). In this case, most women may be earning about what the men's model would predict, given their characteristics, but the women "stars" would be receiving salaries like those of merely average performers rather than seeing their unmeasured high level of productivity (or other unmeasured but positively valued characteristics) translate into high residuals in salary as these do for men (see Figure 2). Again, note that the residu-

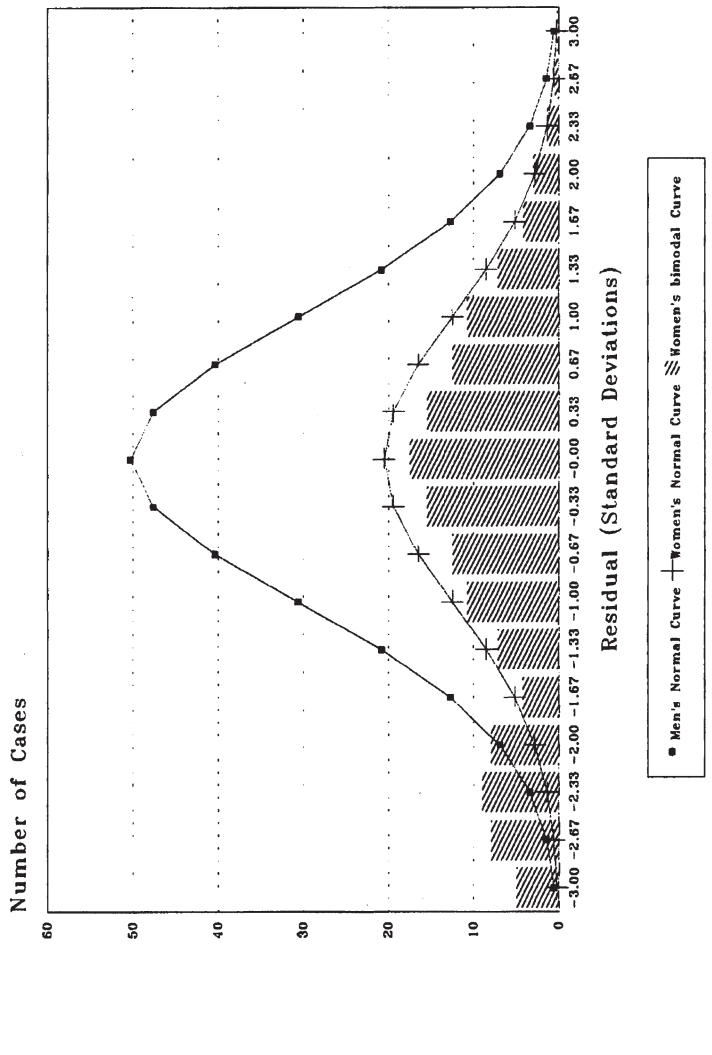

Figure 1. Simulation of Salary Residuals with a Small Extremely Underpaid Group.

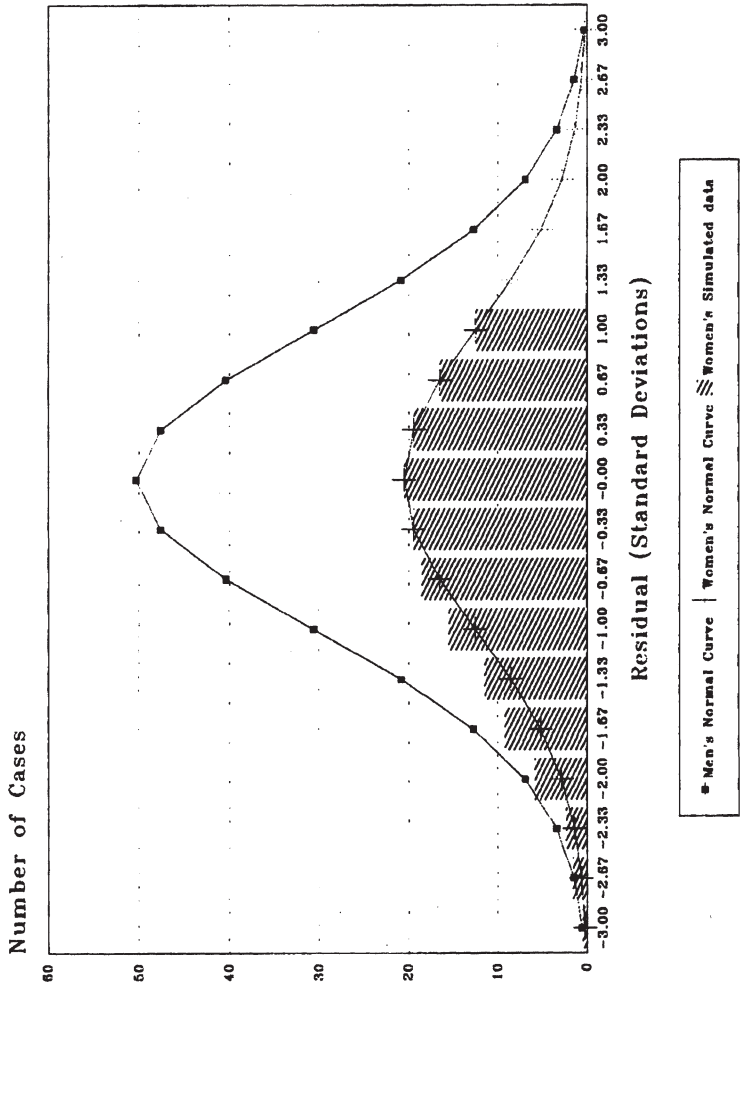

Figure 2. Simulation of Salary Residuals with a Glass Ceiling Effect.

als will not serve to identify these underpaid individual "stars," but rather suggest the nature of the cases that merit special scrutiny or extra redress. Note that in each case, there is a rebuttable presumption that women and men do not systematically differ in these unmeasured characteristics; in other words, the other side may argue back that women do constitute a disproportionately large share of the "dead wood" in the institution (and hence belong in a clump at the bottom) or that women simply are not found among the "stars" at all.

The residuals, however, may show neither a "negative hump" nor a "truncated top" effect, and instead indicate a consistent negative shift in the salaries of all the women taken as a group relative to the group of men (see Figure 3). This model most closely reflects the expectations of the institutional discrimination model. Such a systematic shift downward between the men's and women's distributions could only be explained by productivity factors if one assumes that across the entire range of individuals, women are less productive than men (e.g., women "stars" are less productive than men "stars," "average" women less productive than "average" men, and women "dead wood" somehow even more "dead") (Haignere et al. 1996).

While in the case of the "negative tailhump" or "truncated top" it can be argued that the women affected by discrimination are a subset of all women and need to be individually identified so that their distinctively depressed salaries can be raised, in the case of a "consistent downward shift" it appears that all women are affected by institutional discrimination by an amount approximately equal to the mean residual for the women. If gender does not interact with any of the independent variables in the model, this amount should be the same as a regression coefficient for gender in a model that pools the men's and women's data. Hence, it is important to run a pooled model (combining the men's and women's data) with a variable for gender and compare the regression coefficient for gender with the amount of the women's mean residual when the men's model is applied to the women's data. This comparison, between the mean residual and the regression coefficient, is a global indication of interactions and another important clue to whether or not it is plausible that gender-based pay gaps are different among specific groups of women. If there is a difference, looking at possible interactions, say, between gender and rank, should be further explored.

Assuming there is no indication of interactions in the comparison of the pooled model with that fitting women to the men's model and that close examination of the *pattern* in the residuals also indicates that gender-based inequity is not concentrated on women of a particular type, there is strong statistical evidence that thc wage gap applies to the class "women" as a whole. The appropriate remedy would thus be an across-the-board salary adjustment that would bring the mean residual of the women's distribution up to that of the men (and move the rest of the bell curve of residuals accordingly).

This type of redress seems inherently objectionable to the believers in the model of individual discrimination and managers who focus mostly on individuals, because it applies to a class (all women) something that they believe should be appropriately given only to "deserving individuals," that is, the individual

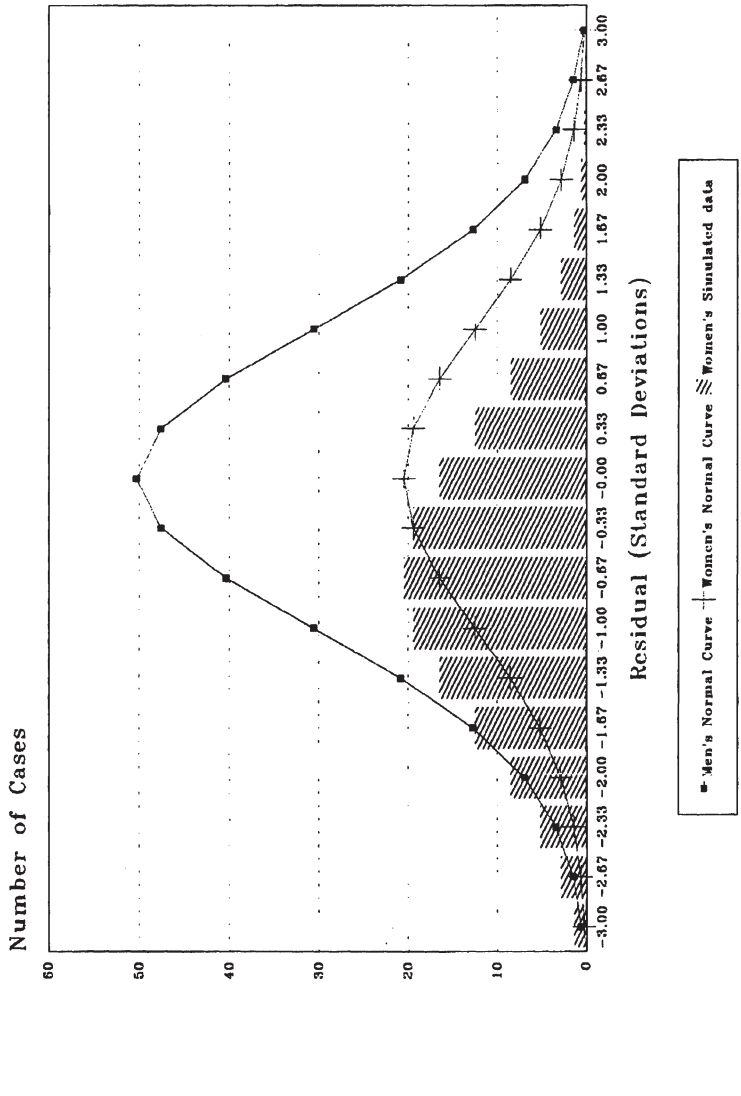

Figure 3. Simulation of Salary Residuals Women Uniformly Shifted Down.

women who have been shown to be the victims of prejudiced administrators who have not given them the salary they "really earned." Their objections are articulated in the case they make for a "more individualized approach.

The Individual Model

The regression approach that reflects the individual model first constructs small groups of "similarly situated" women and men, usually in the same or similar departments, and then runs separate regressions for each such small group. This approach either estimates separate gender coefficients for each group or models men's salaries first and then puts women in the regression. Because the groups are much smaller, the standard errors are typically larger, effects less likely to be statistically significant and the shape of the residuals not necessarily as normal (see, e.g., Holmes-Rovner et al. 1994). We call this the "flagging approach" because it aims to identify ("flag") individuals rather than patterns.[6] Indeed, looking for patterns in such small sets of residuals makes little or no statistical sense.

Instead, the individual approach tests whether gender effects are significant and, if they are, then tries to estimate what an individual woman would be earning "if she were a man" in a particular group (assuming the rest of her characteristics, such as rank and tenure, remain unchanged), by treating her negative residual as evidence of the extent of "underpayment." Then only the women with negative residuals greater than some arbitrary amount have their individual cases examined to see if such a negative residual is warranted or not, and only those women whose negative residuals are found to be "unjustified" by their work are slated to receive compensatory raises. One such study of compensation equity (Holmes-Rovner et al. 1994) created cohorts of men faculty for each woman in the population, averaged salaries within each cohort, and compared the women's salaries with this average. They then looked for patterns of inequity within each rank, by noting the number above 4 percent, within 4 percent, and below 4 percent of the average for the men. They consistently found negative residuals (considered underpayment) of women, with the exception of women full professors.

The use of multiple regression on small groups to flag individuals with negative residuals fits with an understanding of discrimination as the result of prejudiced individuals.[7] It assumes that such cases are relatively rare and should stand out statistically, even in regressions done with such small n's that only the most dramatic effects reach statistical significance. It "finds" those individual women who are not paid at the level of their peers, clearly controlling for factors other than gender, such as type of work or specialty, that might explain differences in salaries, by looking for cases with more or less extreme negative residuals. This approach is intuitively appealing because it is a nice fit with how Americans generally think about salary discrimination, that is, as an isolated occurrence happening to a few unfortunate women. Seeking out those wronged, evaluating their situation, and remedying individual cases is a familiar approach.

This approach may sometimes also rest on a different understanding of what makes "good science." The analysis envisioned by the individual approach is similar to the case-matching approach in medical research, It attempts to create conditions similar to true experiments by selecting cases such that everything but the effect of interest is "the same." This approach defines gender as the only difference between the two "groups" or individuals, if and when they are "matched" on all other important factors (but see Campbell and Stanley 1963, 49, 70-71, for critiques of quasi-experimental matching designs). Since productivity differences cannot be estimated as such, administrators may decide that they can "adjust" for such factors based on their knowledge of the individuals involved. Thus, conceptually, they imagine they are "looking at individuals" matched on particular characteristics, with "only gender" differing between them, rather than at the institutional patterns that gender, productivity, and other unmatched factors produce. This appeals to a sense of how "good science" is conducted, using an experimental model to define "science."

In sum, using the individual model of discrimination as a conceptual framework, as administrators may be prone to do, the appropriate way to proceed is to create several small groups considered comparable for salary purposes, interpret those women with negative residuals as underpaid, and rely on particular individuals designated as evaluators (usually administrators) personally to judge whether the underpayment is warranted or not. Although this approach has intuitive appeal, there are several problematic assumptions in this statistical approach.

First, no individual woman who is earning a higher-than-typical salary for her measured characteristics is considered to be potentially underpaid, even though she may be a "star" performer (as some of the full professors in the Holmes-Rovner et al. study may have been). In other words, some women with positive residuals (e.g., one standard deviation above "average") may merit even higher salaries (e.g., two or more standard deviations above the mean). Unlike men, who may receive exceptionally high salaries because of their unmeasured performance characteristics, women are assumed to deserve no more than average pay (but possibly less) than men of their cohort. As Gray and Scott (1980, 180) argue:

> If an examination may show that on a "quality" basis an individual woman "deserves" a salary which falls below the regression line, we must suppose—unless we assume that women are qualitatively inferior as a class—that on a "quality" basis some women deserve to receive salaries even further above the regression line than are their present salaries.

By reviewing only women with negative residuals, it is impossible to identify women who have positive residuals but are underpaid given their individual, unmeasured characteristics (including productivity). These women may be outstanding in ways that deserve even higher salaries than the men in their group, but because they will not be flagged for review, this inequity will not be

detected. Analyzing only women with negative residuals assumes that women with positive residuals are "overpaid" and implies that an across-the-board settlement is unfair because they also receive an award. In other words, the approach assumes that women cannot deserve more than the average man within the small group in which comparisons are being made. Moreover, by focusing attention only on the lower half of the distribution, there is a risk of confirming possible prejudices that women are "less meritorious" generally (as the highly meritorious women with merely average salaries are exempted from administrative scrutiny).

Second, the flagging approach confuses negative residuals in general with "underpayment" (which then is treated as either justified or not). We know from the assumptions of multiple regression that a normal curve of residuals, both positive and negative, is expected and appropriate. Looking at only women with negative residuals mistakes a characteristic of the institution (which should have a spread of salaries that reflects a spread of productivity-based and arbitrary or chance variation in salaries) for a characteristic of an individual (defining a negative residual as a sign of a problematic underpayment). A negative residual, like a positive residual, simply indicates that the model does not perfectly predict some salaries. Most salaries are close to accurate (most cases fall within one standard deviation from the mean) and some are quite inaccurate (three standard deviations out in either direction). These inaccuracies—in either direction—may well be accounted for by unmeasured (productivity) characteristics, but unless one assumes that all women, at all performance levels, are systematically inferior to men in these unmeasured characteristics, the systematic negative shift found in the institutional model cannot be thus explained away.

The confusion between negative residuals and underpayment is inherent in the small-group flagging approach because its small groups do not allow for meaningful patterns of residuals. The confusion may also arise if the "policy-capturing" regression for the institution as a whole is treated simply as generating negative residuals that are then used to "flag" individual women as underpaid. In the case of no discrimination, half of all women (like half of all men) should have negative residuals. Finding an extreme negative tail or hump pattern in the policy-capturing model is an indication of discrimination that more closely fits the individual model and should help to identify cases and conditions associated with "excess underpayment" concentrated in a specific subgroup, but the possibility that there are other members of that subgroup who are not in the tail or hump but no less underpaid should not be disregarded.

When one focuses on the negative residuals alone (only one form of inaccuracy) and treats them as indicators of "underpayment" for women, the leap to considering them also as indicators of underpayment for men is not large. Because the individual model rests on regressions with small numbers of cases and uses them to find "potentially underpaid" women for individual scrutiny, two undesirable outcomes are encouraged: The women whose unmeasured characteristics tend to make them less valuable to the institution on average

will receive the bulk of the scrutiny (and most cases of "discrimination" will be found to be "justified") and men who are also receiving below-average salaries relative to their measured characteristics will be encouraged to claim they, too, deserve salary review and compensatory payments. Having thus provoked them, the administration may feel it has little choice but to give "compensation" to these men as well, thus eliminating the possibility that raises for women might actually mitigate the gender gap.

Summary of the Two Models

The model used for analysis (policy-capturing or flagging) is thus inextricably bound with the theory of discrimination used (institutional or individual) and the politics of redress (across-the-board or individual awards) preferred. As Gray and Scott (1980, 180) suggest,

> If a case-by-case analysis is to be used to discover factors not taken into account by the regression analysis, then obviously one has discounted the notion of treating the discrimination on a class basis. … The appropriate remedy for statistically revealed discrimination is a statistical one; the regression line for women needs to be brought up to that for men. This means raising the salary of each woman, whether her current salary is above or below the regression line for men's salaries.

This approach, however statistically sound, remains unpalatable to those who refuse to consider institutional discrimination real and are committed to a search for individual acts of discrimination. When we consider how widespread an individual understanding of discrimination is, it is not difficult to understand why an institutional approach might be ignored or rejected. In addition, academic administrators are managers of individuals and thus want to take the people involved into account in forming "qualitative" judgments and to exercise discretion and control over those whom they manage. However, given that considerable evidence suggests both types of discrimination operate in American society (see, for example, Benokraitis and Feagin 1995, for a review), it is perilous to focus only on individuals.

Given the politics of salary-setting, choosing between approaches on statistical grounds alone will not work. The methods preferred by each side reflect their assumptions about how discrimination operates. Perhaps because most studies of this sort are applied evaluations, less attention is focused on assumptions and theory and more on the merits of statistical techniques (Gray 1993). However, arguing over what method should be used without addressing the theoretical assumptions of each results in both sides talking past each other. As we illustrate in the following case study, this conflict of perspectives is real and consequential in applied research on salary inequity.

Health Center Case Study

In 1992 the authors were enlisted as statistical consultants for a committee of women faculty appointed by the administration at a university medical/dental school (hereafter, Health Center) to determine if there were gender-based inequities in salaries.[8] There had been two prior studies of women's pay at the Health Center over the previous decade, each of which had found substantial disparities, but both were statistically weak (primarily relying on ANOVA models with few control variables), and neither had had a practical outcome (Cipes 1983; Dadalt 1986). The women faculty on the main campus had won a salary adjustment of $1,800 per woman faculty member in 1988, and the Health Center faculty women considered that a study done with the same methodology had a good chance of being accepted as a reasonable approach for their campus as well.[9]

The policy-capturing approach had been used on the main campus, and the committee suggested that we apply the same approach to the base salaries of the Health Center faculty. The initial data were 1992 human resources payroll data for 535 faculty, 29 percent of whom were women and 90 percent of whom were white. Women faculty made up the majority of the instructors (60 percent of 55), close to half of the assistant professors (41 percent of 233), but few of the associate (15 percent of 102) or full (8 percent of 125) professors. Of 19 department heads, only 1 was a woman. Women were least likely to be in the dental department group (4 percent of 24) and surgery (19 percent of 35), and most likely to be in pediatrics (45 percent of 53) and at off-site locations (39 percent of 33). Of the 161 faculty in basic science departments, 28 percent were women, compared to 30 percent of the 373 faculty in clinical departments. Women were, on average, younger (41 vs. 46 years for men), likely to have less time at the institution (6 vs. 9 years), likely to have earned their degrees more recently (13 vs. 18 years ago), and likely to have salaried employment at the Health Center represent a lower percentage of their time (82 percent vs. 88 percent).

At the time the study began, the mean difference between men's and women's average base salaries was $28,500. By mutual agreement between the faculty committee and the administration, all clinical bonuses (analogous to profit-sharing), research assistants and associates, medical and dental residents, visiting faculty, and "superstars" (men with salaries that were more than six standard deviations from the mean) were excluded from the analysis. As with the decision to include rank in the model, these constraints also tend to make any estimates of gender-based inequities conservative (low). All base salaries as of June 26, 1992, for the 1992-1993 academic year were expressed as 12-month salaries. Because variation in salaries in such a narrow occupational-institutional context is much smaller and less skewed than in the general population, we concluded that converting salaries to logarithms offered more cost in loss of ready interpretability than it was worth. All effects are thus expressed in dollars.[10]

Table 1: Policy-Capturing Model: Multiple Regression Predicting Male Faculty Salaries, 1992

Departments	B (dollars)	SE *B*	Beta
Administration	22,518*	9,724	.0709
Offsite	–4,740	5,664	–.0248
Community	10,880	6,260	.0595
Psychiatry	2,924	5,367	.0163
Dental clinical	12,651	7,086	.0707
Pediatric	–5,001	4,920	–.0311
Diagnostics	12,955*	4,293	.1026
Basic science	5,437	5,494	.0458
Surgery	40,070*	3,619	.3718
Degrees			
Mixed	–31,212*	13,099	–.0647
Masters	–28,755*	8,712	–.0967
PhD	–26,069*	4,408	–.2820
Dental	–41,700*	4,928	–.3216
Type			
Basic	–27,672*	4,822	–.2986
Tenure			
Not yet	2,412	4,209	.0175
Tenured	7,994	4,139	.0893
Additional variables			
Years at the institution	–497*	231	–.0794
Time since degree	778*	392	.1820
Percentage employed	342*	47	.2170
Age	123	405	.0270
Asian	–8,351	7,232	–.0313
Former administrator	33,213*	7,739	.1183
Rank			
Associate	11,290*	3,445	.1111
Professor	27,930*	4,653	.3006
Department head	58,010*	6,940	.2887
Dean	44,412	25,331	.0533
Constant	48,482*	5,775	
R^2	.764		
Adjusted R^2	·75		

The comparison group for department is medicine; for degree it is M.D.; for appointment type it is clinical; for tenure track it is in-residence; for racial/ethnic group it is white, black, Hispanic, and other; for rank it is instructors/assistants. The dependent variable is 1992 salary.

* $p < .05$

The policy-capturing regression model (Table 1) gives the estimated effect in dollars of a particular characteristic (e.g., academic degree) while controlling for all other characteristics (e.g., rank) for men only. Gender is not included in the regression as a variable because this approach examines gender by bringing women into the men's model. Variation around the mean in the men's model is expected; it reflects the unmeasured productivity factors that also produce le-

Table 2. Departmental Groupings for the Original Committee Model

Variable	Group Name	Departments
Pedidept	Pediatrics	Pediatrics, Pediatric Module, Pediatric Dentistry, and Orthodontics
Dentdept	Dental clinical	Periodontics, Prosthodontics, Restorative Dentistry/ Endodontics
Psychdpt	Psychiatry	Psychiatry
Surgdept	All surgical	Anesthesiology, Obstetrics/Gynecology, Oral Surgery, Orthopaedic Surgery, surgery
Medicine	Medicine	Medicine, Neurology, Student Health, Travelers Center on Aging
Diagdept	Diagnostics	Laboratory Medicine, Nuclear Medicine, Oral Diagnosis, Pathology, Radiology
Sciedept	Science departments	Anatomy, Biochemistry, Biostructure and Function, Microbiology, Physiology, Pharmacology
Commdept	Community departments	Behavioral Sciences and Community Medicine, Community Medicine
Offsite	Offsite clinical	Family Medicine, Burgdorf, CRMHC
Admindept	Administration	Dental Dean, Student Affairs-Dental, Dean's Office-Medical, Academic Affairs-Medical, Minority Student Affairs

Source: Gronowicz et al. (1993).

gitimate salary differences.[11] The variables include department group (see below), degree, rank, tenure, type of appointment (basic science or clinical), age, race, and present and past administrative responsibility. The coefficients from the regression equation can be interpreted as the "worth" of various characteristics, such as department or degree.

Because there were over 40 departments on the payroll system, it was necessary to group departments for the analysis. The committee combined departments considered to have similar types of activities (see Table 2). Since the model already directly included as variables the type of degree, clinical versus basic science appointment, and rank, department groupings provided an additional level of information about salary setting. Theoretically this grouping accepted broad market forces as legitimate salary setting factors, even though markets may be biased against fields with a higher proportion of women (cf. Bellas 1997).[12]

The 1992 data were flawed and each year after 1992 the data were improved (errors were detected and corrected) and additional refinements were added. Unfortunately, we had no way to clean the 1992 data in 1996 because over 100 people were no longer at the institution. Therefore, the model presented here is *not* the best possible model, but is the historical model presented to the faculty. It should be used as an example of how to proceed with this type of analysis, not as providing a specific "finding."

Once the men's model was established, it was applied to the women's data to ascertain the fit using residual analysis (comparing predicted salaries to actual salaries). A model was also constructed using the women's salary data,

Table 3a. Example Mean Salary Information

	Actual Salary	Salary Predicted by Characteristics	Unexplained Difference
Men	98,224	98,224	0
Women	69,708	74,438	–4,731
Difference	28,516	23,786	–4,731

Mean salary, mean salary predicted by characteristics, remaining unexplained difference between men and women's salaries, 1992 faculty.

Table 3b. Example Residual Statistics

	Minimum	Maximum	Mean	Standard Deviation
Men				
Predicted salary	26,738.0	209,653.0	98,224	37,356
Residual	-70,537.0	83,869.0	0	20,761
Standardized residual	-3.38	3.9	-.0002	.9993
Total cases = 380				
Women				
Predicted salary	9,871	145,113	74,438	30,882
Residual	-68,881	60,594	–4,731	20,735
Standardized residual	-3.2	2.8 -.22		0.96
Total cases = 155				
Multiple R for women = .76 (R^2 = .58)				
Multiple R for men = .87 (R^2 = .76)				

Residual statistics for men and women from the regression of men's faculty salaries on the salary-setting characteristics.

and then the coefficients were applied to the men's data in order to rule out the possibility of a negative mean residual for men (Gray 1990). The men's residual was positive, and over $9,000. As Jones and Kelley (1984) discuss, this difference in estimate of the size of the wage gap reflects how the interaction between the effect of gender *per se* and the effects of gender differences in the slopes (coefficients) for the individual variables (overall differences in rate of return) is treated. In the case of such differences in slope, they argue, the most appropriate model uses men as the norm.[13]

As Table 3 shows, when women are put in the men's model, the average residual for the women faculty was –$4,731. Eighty-three percent of the gross wage difference, or $23,786, was explained by the characteristics included in the policy-capturing model and thus treated as "legitimate."[14] The most important predictor variables included being a former administrator, rank, percentage employed, years at the institution and years since degree, type of appointment (basic or clinical), type of degree, and appointments to three of the departments (administration, diagnostics, and surgery). The model explained 76 percent of the variation in men's salaries (R^2 = .764).[15]

Despite the systematic downward shift of –$4,731, the residuals for the model applied to the women's data are approximately normally distributed, as are the men's (although the men's distribution is not quite normal, being particularly narrow and tall, that is, leptokurtic). Based on the values for the measured characteristics, only 40 percent of women earn more than predicted and 60 percent earn less (see Figure 4). No second hump or long tail at the bottom is to be seen, providing no evidence of an especially discriminated against group of women bringing down the average. There may be some indication of a glass-ceiling effect, suggesting a special need for concern about high-performing women. However, there is no evidence that gender significantly interacts with the variables in the model; when the same regression is run on all of the faculty data simultaneously and a dummy variable for gender included, the coefficient for women faculty is essentially the same –$4,683. This method also establishes a significance level for the effect of gender (p = .04), as Ikeda (1995) and Hauser and Mason (1993) suggest.[16]

The Working Group on the Status of Women Faculty reported the results of the study to the administration along with its recommendation that all women's salaries be raised by $4,731. The administration challenged the study on the basis of errors in the data and the type of analysis and pressed for an analysis of small market-based groups of faculty, with most of the same variables controlled. The stated goal of the new analysis was to look for women who were negative outliers (i.e., had large negative residuals) and to rectify errors in the data. Concerns about particular women who fit the model poorly, objections to being given a solution, and unfamiliarity with the analytic strategy added to their desire for a new approach.

A series of negotiations ensued about the appropriate variables to include in the model, groups that should be excluded, additional data cleaning, and different clusterings of departments. The primary sticking point was the administration's desire to identify specific women who might be "underpaid" (i.e., "flagging" particular women with negative residuals). There was no objection to re-running the basic model with some adjustments suggested by the administration. The addition of an interaction term for individuals in the surgery department without an MD degree did improve the model. In addition, removing various groups, such as faculty at off-site locations, made the sample more homogeneous and increased the R^2. None of the changes suggested by the administration, however, reduced the average gap between men's and women's salaries below –$4,000. Some of the changes increased the gap to as much as –$6,000.

The administration did not accept the committee model. Their concerns focused on the accuracy of the human resources data, the exhaustiveness and accuracy of the variables in the model, and the correct people to include in the analysis (groups to include or exclude). They requested a type of comparative analysis of individuals in groups and refused to accept the across-the-board settlement. We interpret the administration's commitment to a "flagging" approach to find possible individuals who might be unfairly underpaid as a result of an individual understanding of discrimination. No progress on an across-the-board settlement was made in academic year 1992-93.

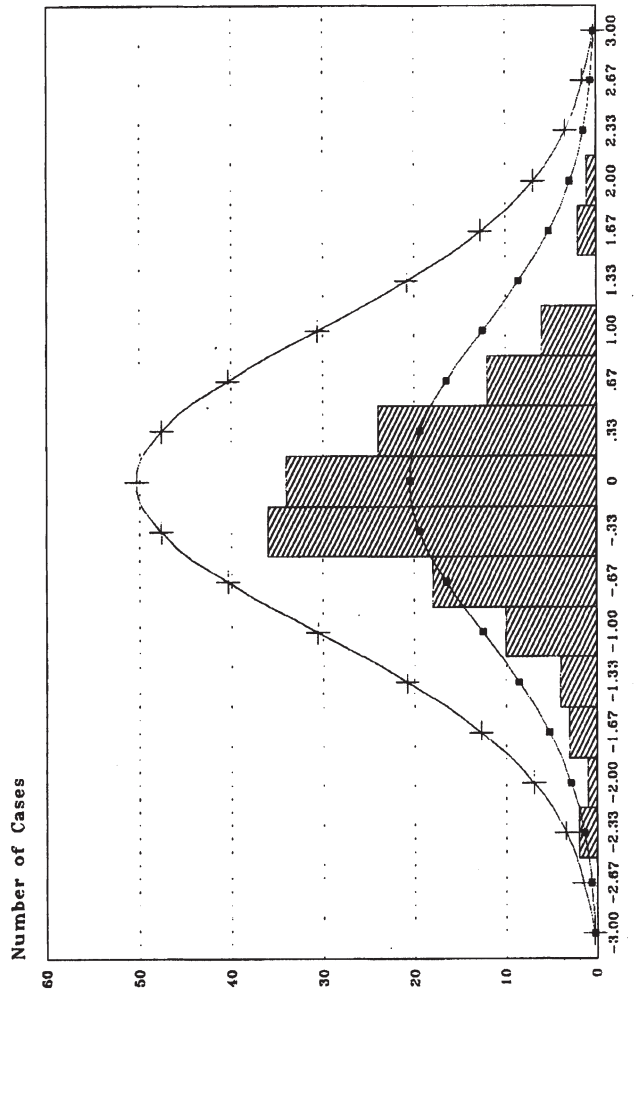

Figure 4. 1992 Actual Case Study Salary Data Distribution of Salary Residuals

In academic year 1993-94 the administration requested a new set of analyses to identify individual women who might have unfairly low salaries. Part of the argument advanced for this approach was that women who are "adequately compensated" should not receive raises, and some women who are not adequately compensated deserve more than the proposed $4,731. The administration took an active role in the continued analysis. The departmental groupings were revised based on assumed market value criteria rather than on the task similarity criteria used in the original model. Separate regression models were run for each group with the explicit goal of identifying individual women who might be underpaid relative to comparable men, according to a Holmes-Rovner et al. (1994) type model. This "searching for individuals" approach was consistent with the individual prejudice conception of how discrimination operates, but it also reflected a laudable desire to "look closely at the data." By looking at faculty in groups assumed to have similar characteristics that determine pay, it was possible to directly see who was being compared to whom.[17]

Despite our objection to defining cases with negative residuals as underpaid individuals, with the advice and consent of the faculty committee, we accepted the administration's request for further analysis. Although throughout the study we were technically hired by the administration, we had primarily worked with the committee (also appointed by the administration). At this point the administration took on a more active role. We ran the requested regressions on the small groups of faculty and flagged women whose residuals were greater than one standard deviation below the mean (an arbitrary cut point). Models were run for separate groups created by the administration. These were (1) basic scientists in basic departments, (2) basic scientists in clinical departments, (3) administrators, and (4) five groups of clinical scientists in clinical departments (clustered by specialty). Table 4 provides a summary of these results. In effect, these eight groups take the place of the ten department groups in the first institution-wide analysis, but the regressions are run on each subgroup separately, making the number of cases too small for a meaningful analysis of the residuals. As a cross-check, after analyzing these separate groups for women outliers, we re-ran the institutional model substituting dummy variables for these new market-based groups for the previous task-based department group dummy variables. This approach did not improve the fit of the model (R^2); when applied to 1993-94 data, it also did not significantly reduce the size of the pay gap between men and women. This interchangeability of the statistics as such should emphasize the extent to which the preference for a type of model reflected an underlying difference in focus of concern: the small group approach turns attention away from looking for patterns in the residuals and toward equating negative residuals with possible inequity. Throughout the process, the models were frequently re-estimated as individuals came and left the institution, as individual women received salary adjustments, or as new groups were considered ineligible and were excluded, but the size of the overall gap in the institutional model remained the same.

Table 4. Results of the Small Group Analysis (individual) Compared **to** the Policy-Capturing Model (institutional) with 1994 Data

Model	Men *N*	Women *N*	Adjusted R^2	Mean Residual	*p*
Basic in basic	72	17	.82	–2,836	.23
Basic in clinical	37	22	.71	–2,983	.43
Clinical in clinical #1	82	35	.66	–3,403	.25
Clinical in clinical #2	33	16	.54	–9,852	.15
Clinical in clinical #3	38	6	.35	–14,783	.5
Clinical in clinical #4	7	1	.96	7,377	—
Clinical in clinical #5	38	4	.94	343	.95
Deans/administrators	14	3	.32	–1,718	.95
Institutional model	342	120	.75	–1,052	.64

Clinical #1: General Pediatrics, General Medicine, Neurology, Family Medicine, Psychiatry. Clinical #2: Laboratory Medicine, Pathology, GI Medicine, Neonatology, Dermatology, Cardiology, Emergency Medicine. Clinical #3: OBGYN, general surgery, Urology, ENT, Nuclear Medicine, Radiology, Anesthesiology. Clinical #4: Neurosurgery, Orthopedics. Clinical #5: all Clinical Dentistry.

This modeling process had an interesting effect. It appeared to us that the administration became even more familiar with the details of how salaries were set by investigating the circumstances of women who were negative outliers in the small groups. Low salaries in some departments for some individuals became a concern. As a result, there was a recommendation to the board of trustees that the board develop some salary-setting guidelines, "without limiting their freedom to reward excellence and recruit and keep top scholars." Moreover, the administration has now set up a biennial salary review to ensure that gender-based differences do not creep back into the system. While applauding these steps to improve gender salary equity at the institution, we continued to emphasize the lack of concern this individual approach showed to women who were "top scholars" and underrewarded their productivity (despite their positive residuals). As of 1997, there were some indications that even these issues might be addressed in a follow-up study.

In June of 1994, the administration increased the women faculty's pay across the board by $1,000 (or a prorated equivalent for part-time workers). This increase was announced as a sign of good faith that an institution-wide problem "may exist." The institutional and individual models were then reapplied to salaries after the June increases. Despite the limitations of the administration's flagging model, primarily that underpayment was equated with negative residuals and only women with negative residuals were reviewed, it did generate evidence that some of the women reviewed still had salaries significantly lower than their peers. There were women who appeared grossly underpaid when compared to men with similar characteristics and work situations. This was now considered convincing evidence of unfairness, both by the administration and by the predominantly male faculty councils of the medical and dental schools, which in fall 1994 set up a joint committee to con-

tinue to improve the analysis and to monitor "progress toward resolution" of the inequities. Additional salary awards were made to individuals in the 1994-95 academic year.

The joint council made an additional contribution to the analysis, moving it further away from the comparable worth precedent it had originally followed. Instead of grouping departments and specialties according to assumed market values, they used published salary statistics for specialties in order to make empirically sound market-value-based groups (described in Jacob et al. 1996). Such purely market-based groupings automatically erase any evidence of society-wide discrimination that takes the form of underrewarding women in specialties with high proportions of women (cf. Bellas 1997). The remaining analyses, both institutional and individual, were conducted using these new groupings.[18]

The joint council salary review committee—consisting of two men and two women representing the medical and dental schools, clinical and basic science faculty—reported in February of 1996 that as a result of review by the administration and the committee, "28 percent of the full-time women faculty salaries were identified as outliers, reviewed, and 22 *percent were adjusted.* Additionally, an undetermined number of women received equity increases July 1, 1995 in addition to the cost-of-living increase received by most faculty" (Jacob et al. 1996, 4).[19] Indeed, both the institutional model and the small group analysis were conducted on newer data (1995 salary data), after both the $1,000 raise had been applied and the additional salary increases given to individual women who were seen by the administration as underpaid. In the 1995 institutional models, after these salary adjustments were made, the gender gap is reduced to about $1,000 and is no longer significant statistically. The small group analyses also now showed no systematic gender bias remaining for any group, with all *t* tests showing no significant differences (Table 4). However, this type of analysis, relying on negative residuals only, could not detect if any of the women with positive residuals were also underpaid, because none were flagged for review.

Discussion

Rather than reaching a stalemate over incompatible methodologies, in this particular case, a compromise process was found that allowed both sides to claim victory. On the one hand, the women faculty all received at least a token award of $1,000 and some women faculty received substantial salary adjustments, which was a breakthrough after years of discontent and unsuccessful claims for change. After several years of negotiations and individual wage adjustments, there was no longer a statistically significant difference between men's and women's average salaries. On the other hand, the administration succeeded in having its individualized approach to identifying pay inequity adopted, and so kept the responsibility for making any salary changes within the discretion of the administrators themselves. Since the actual award

amounts remained confidential, the ability of the faculty to see how arbitrary or well-founded these decisions were was quite limited.

In some respects, both sides moved toward each other over the course of the negotiations. While the committee and the administration continued to disagree with respect to theoretical approach (and consequently statistical models), there was considerable evidence of the administration's increased commitment to identifying and eliminating gender-based salary inequity. The administration took several steps to redress identified inequities, including across-the-board compensation, individual settlements, the joint council monitoring committee, and the recommendation to the trustees regarding salary setting guidelines. The committee allowed more individuals to be excluded from the model (e.g., those in off-site positions) and shifted gradually from task-based to market-value-based definitions of similarity, accepting society-wide structural arrangements that may disadvantage women.

A shared victory can be noted as well: Monitoring pay in and of itself can promote more equity. Several of the faculty committee members provided anecdotal evidence of increased awareness of and decreased tolerance for gender differences in salaries over the four years of studies. It is also clear that studies alone do not bring change. We suspect, but have no systematic evidence for this suspicion, that the involvement of several prominent women faculty and their reports to both the dental and medical school faculty councils contributed more to a changing climate than the studies themselves. Major's (1987) work on gender and entitlement suggests that making male pay norms highly visible can produce discontent and pressure for action from women. The steady persistence of the committee members kept the issue on the agenda and, we suspect, contributed to the actions taken in response to the study.

Application of the theoretical models of individual or institutional discrimination to the arguments and policy preferences of the administration and the women's faculty committee helped us as outside statistical experts to reach a resolution in this case. Despite our misgivings about the theory and methods preferred by the administration, we chose to run the administration's models for them as well (with the full knowledge and consent of the faculty committee). At first, we were surprised that our argument for an across-the-board settlement—that it would be fairer to the best women on the faculty—was rejected. Ultimately, we imputed this resistance to both the administration's tacit view of discrimination as an individual process and their desire to keep salary information and decisions in their own hands. We concluded that if adequately rewarding all women, including the ones who are "twice as good as men," demands rethinking basic institutional practices, many administrators may be unwilling to consider such a change. As Gray (1990) suggests, people who use statistics cannot always "hear what statistics have to tell."

Our experience in this particular case study of salary inequity, as well as our general analysis of the competing models of discrimination likely to be held by each side, leads us to some general conclusions and caveats about the pro-

cess of analyzing discrimination with the goal of making and supporting sound claims for gender-based salary adjustments.

First, we think that women faculty and administrators should be aware of the implicit model of inequity processes that they hold and thus also consider the type of statistical evidence that they will accept as appropriate. We are especially concerned that many women faculty seem to accept the definition of a negative residual as "underpayment." Committees that take (or accept) this approach disregard the potentially underpaid women "stars" even as they encourage men with negative residuals to define themselves also as "underpaid." Highlighting cases with negative residuals puts the spotlight on the women and men who are more likely than most to have lower values on the unmeasured productivity variables. This inadvertently sets up the study to encounter a maximum of political backlash, since a focus on the low end of the distribution risks both confirming stereotypes about women as low performers and stirring up resentment among men who feel it unjust for any "comparable" woman to be earning more than they do. Flagging (and trying to contest "justifications for") some women's low salaries also does not even begin to address the "glass ceiling" for women achievers. When gaining and maintaining the strong support of the most respected women faculty on campus may be key for reaching a resolution politically, disregarding the pay problems that can affect "above average" women is strategically unwise as well as theoretically unsound.

Theoretically, we think that salary equity committees should consider not just the main effects for gender but also the shape of the residuals as important clues to processes of individual and institutional discrimination for women at all salary levels. We do not doubt that there are instances where concentrated animosity toward women may lead to a cluster of especially underpaid women at the bottom (evidenced in the long negative "tail" or second "hump" in the distribution), but unspoken norms about women not earning more than men of their cohort may also produce a "truncated top" that holds back the salaries of the women who excel. It is a pattern of negative residuals, not the negative residual of any individual, that should guide a committee's interpretation of inequity. If such a pattern is found, then the committee should be attentive to individual cases that may represent inequitable pay but appear anywhere in the distribution. For example, if there is a prejudiced administrator in a specific department, the exceptionally low salaries of some women in that department should trigger a review of all women's salaries in that department, not just of the ones that show exceptionally large negative residuals.

Second, we encourage applied researchers to be aware of and explicit about the theory—and the politics—of the models they employ (cf. Acker 1989; Steinberg 1995). The power that administrators hold over salary setting, especially on nonunion campuses, means that the administration's conception of discrimination is often definitive. It may be necessary to provide evidence that is structured to fit in their implicit model in order to create any acknowledgment that "a problem" may exist. Making compromises to meet their concerns can lead to

a more satisfactory outcome over the long run. In our case study, a "quick victory" in the form of an across-the-board increase might not have had the consciousness and policy-changing effects in both faculty and administration that a longer negotiating process produced. Indeed, on the unionized campus that provided the initial model for our study, the salary awards were made without fanfare and salary setting continued as usual. Despite the across-the-board awards that were won in 1988, it would not surprise us to find salary inequities there again in 1998. Paradoxically, institutional change may demand engagement with just those administrators who do not hold an institutional model of discrimination processes.

Understanding the administration's perspective helped to reach such a long-term resolution. Administrators who think discrimination is rare may be more concerned that an across-the-board settlement will "unfairly reward" a woman who happens not to be underpaid (Type II error) than that their focus on "flagging" negative residuals will miss correcting problems for women who have smaller-than-appropriate positive residuals (Type I error). Administrators' definition of the relative importance of Type I versus Type II error is theoretical rather than empirical, resting at its root on a conviction that salary discrimination is rare rather than typical. Extensively working with the administration on a great number of individual "flagging" models may help to make discrimination seem less rare and contribute to their willingness to consider across-the-board increases (as it did in our case study).

Concern about giving "undeserved" increases to women may also reflect a more deeply felt concern about an "unjust advantage" for women than for the "unjust advantages" of men over the years. Haignere et al. (1996, 88) recount a case in which, as a result of a salary settlement, a woman was to receive $300 more per year than a man who was admittedly her equal: this was felt as a gross injustice. The underpayment she had faced over the previous six years was $7,800. At that rate, however, the advantage would have to persist for 26 years for her to simply earn back the amount she had been underpaid. Greater concern with women earning more than men than the reverse represents exactly the sort of "built-in bias" in the system that the concept of institutional discrimination was created to cover (Major 1989). Trying to eliminate institutional discrimination in salary-setting means confronting such "nonhostile" biases against women, which may demand campus-wide discussions.

Making clear the implicit theories of discrimination operating on both sides of the negotiating process can improve understanding and appreciation for the opposing statistical approaches, and this in turn may help reach a resolution. While the administration may not be willing to change its model, the women faculty may be less easily diverted to "fixing" just a few individuals and the men faculty may be more willing to support them over what is likely to be a lengthy struggle. Statistical experts should not attempt to fool their clients (or themselves) into thinking that theirs is an atheoretical effort to merely find the best-fitting regression model, but they should show how both types of approaches to even the same regression are tests of theory. The applied researcher thus can (and under some circumstances, should) test both competing theories

and try to explain the strengths and weaknesses of each, including the implications of institutional analysis for remedies.

On one hand, using the institutional approach makes it possible to show when discrimination is a systematic pattern and allows for exploration of what kind of pattern it is by examining main effects, interactions, and residuals. The policy-capturing approach itself can increase administrators' awareness of how much certain factors contribute to basic salary-setting decisions and can lead to efforts to make policy more explicit and more fair in principle. On the other hand, using the flagging approach can confront administrators with evidence they find compelling about some individual women, including some instances of gross underpayment well in excess of institutional patterns. In addition, by focusing attention on the small groups in which salaries are actually negotiated, the model may lead administrators to reflect more on how salaries are bargained and be more aware of the need for clearer and more uniform salary-setting policies at the institution.

Overall, we think that looking at institutional as well as individual patterns of discrimination is useful in achieving more than short-term salary adjustments. If the pattern of women's underpayment is not going to simply be reestablished by a return to ordinary, nonhostile practices with discriminatory effect, these practices will need to be identified and discussed institutionally. For example, the common institutional practice of expecting talented high achievers to get competing outside offers to demonstrate their value to their home institution may disadvantage women in that they are more likely than men on the faculty to be married to another faculty member and thus less credible as seriously contemplating a move, making it harder to get outside offers and less likely that such offers will be aggressively matched (Haignere et al. 1996). Such "business-as-usual" approaches simply retain the organizational norms constructed based on professional men's life patterns of the 1950s (Acker 1989). Taking both types of statistical models to their limits may push both parties to begin to discuss such assumptions and identify the problems both types of discrimination may cause. Even this, however, still leaves a distance to go in constructing the political dialogue that will enable academic institutions to acknowledge and overcome their biases.

Notes

1. The empirical model presented here has many limits, and is not offered as evidence of discrimination, whether illegal or not. The analysis only provides an example of what the model and the output might look like using this approach.
2. Pay and promotion reflect different institutional processes. They are often set by different decision makers (in the case of promotion, the support of the full professors of the department may be required, while pay increases may be given by department heads, deans, or higher administrators), using different criteria (the "merit" assessed for pay may more readily include administrative service or teaching, while promotion may focus narrowly on scholarship), and standards of fairness (raises may reflect ideas about inflation, family size, or income demands while promotions do not). Thus it should not be surprising that discrimination in each might vary independently.

3. In regressions, the residual is the difference between the actual and the predicted salary. In sample-based studies, such residuals also represent sampling error. Since salary studies use a full population rather than a sample, the residual is unexplained variance that represents some real but unmeasured process by which salaries are set, even if it is sheer caprice (chance).
4. Following this logic, the institutional norm for female-sex-typed occupations would be the women's salaries, and in these instances the men's data should be fitted to the women's model.
5. This could also appear as a clustering of women just at the most extreme low end of the distribution, producing a long "tail" there, if all the affected women were very extremely underpaid. It is more likely, however, that the center of the separate distribution of "underpaid women" falls somewhere above the lowest point of the distribution of "fairly paid women" even when there are two separate underlying distributions so that the two groups cannot be separated by statistical means alone. The overlap of the two underlying distributions is what generates the observed bimodality.
6. At its worst extreme, this approach eschews statistics entirely, selecting one individual "similarly situated" man to compare to each woman.
7. Unfortunately, some of those who adopt the institutional approach also fall into the trap of confusing the negative residual itself with "underpayment" and simply interpret a regression performed on the entire institution in the same way we discuss for the small group regressions. It is no less a problem in this instance.
8. To be precise, the second author was hired and the administration paid her very modest consulting fees.
9. This study, briefly described in Geetter (1988), was conducted by a joint administration-union team as mandated in the previous collective bargaining contract. There had been nearly a decade of negotiations (between women faculty and the union, between the union and the administration, and between the statewide coalition of state employee unions and the state legislature) that made possible both the study and its relatively rapid translation into salary inequity awards.
10. Using logs lessens the gap between the mean and the high end of the distribution and is often warranted when distributions are very skewed, with a long high-end tail. But in this instance, the "superstars" who formed the high-end tail were removed and the remaining distribution not particularly skewed.
11. Measurement error in the independent variables can also contribute to the spread of the residuals, but most of the variables in this model are relatively straightforward (similar to demographic indicators), such as age, department, rank, type of appointment, and so forth, which tend to have less measurement error than indicators of more abstract constructs. "Error" in the dependent variable is not measurement error (the measured salaries represent what base pay the faculty actually received) but may represent variation that is due to unmeasured but systematic causes (what we are calling productivity) or unmeasured, unsystematic causes (what can be called chance or caprice); we assume the latter is negligible and inseparably confounded with real productivity in ***legitimate*** salary decision-making.
12. There was also disagreement regarding how to group individuals. The original committee chose task similarities and based the groups on different departments. We accepted the limitation of the data (human resources did not list specialties within departments) and attempted to group individuals by department types. We consciously chose not to use market force groupings (groupings based on published data that grouped individuals based on similar pay ranges for type of work performed) because of gender bias in the market that differentially rewards different types of employment (comparable worth issues). Even the committee's task-based departmental grouping confounds similarity of task with proportion female in the departments (e.g., by grouping pediatric departments together). However, administration was very aware of how market forces influence salary setting and felt it was not fair to penalize one institution that is at the mercy of worldwide market forces. Thus, the administration considered the original model inadequate because it did not use market forces to group faculty. Ultimately the later committee and

the administration agreed to use published salary data to group individuals by medical specialty regardless of departmental affiliation. With considerable help from the administration, we were able to get this specialty information for all employees and to add it to the personnel data.

13. When values of measure are arbitrary (as in a 0/1 dummy variable for gender), the difference in slopes between the two models implies a different intercept as well (the distance between two nonparallel lines will vary depending on where the *y*-axis is set).
14. In fact, some differences (those due to race and age particularly) are also illegal and illegitimate, but for the purposes of assessing gender-based inequities, they need to be examined separately (that is, statistically controlled), as well as examined for interaction effects with gender (none was significant).
15. Gray (1990) considers a model explaining more than 70 percent of the variance to be acceptable, in that the size of the gap is not being inflated by a generally poor specification of the variables. Although the specific number is only an arbitrary convention, like the .05 level for statistical significance, it reflects a general principle of seeking a good fit before turning to examination of the residual.
16. In fact, Hauser recommends only running the regression with gender as a variable, but although this method is subsumed in ours, it is less useful because it provides no information to examine the nature of the average difference, as the analysis of residuals does.
17. This small group/individual prejudice approach is also closer to the way variables are controlled in experimental designs (Campbell and Stanley 1963) and also may appeal to physicians and other clinical researchers for this reason. The earlier studies at the Health Center that faculty women carried out also tended to employ comparison grouping and analysis of variance rather than multivariate statistical controls.
18. The data for setting market-based values on departments were drawn from the American Association of Medical Colleges Report on Medical School Faculty Salaries (Smith, 1993-1994) and the American Association of Dental Schools (i993-1994) Faculty Salary Survey.
19. These figures reflect all of the adjustments made up to this point, not just those recommended by the joint council.

References

Acker, Joan. 1989. *Doing comparable worth: Gender; class and pay equity.* Philadelphia: Temple University Press.

American Association of Dental Schools. Faculty Salary Survey. Washington, DC: American Association of Dental Schools. 1993-1994.

Becker, Gary. 1985. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics* 3:S33-58.

Bellas, Marcia L. 1994. Comparable worth in academia: The effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines. *American Sociological Review* 59:807-21.

———. 1997. Disciplinary differences in faculty salaries: Does gender bias play a role? *Journal of Higher Education* 68 (3): 299-321.

Benokraitis, Nijole V., and Joe R. Feagin. 1995. *Modem sexism.* 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Bielby, William, and James Baron. 1986. Men and women at work—sex segregation and statistical discrimination. *American Journal of Sociology* 91:759-99.

Blalock, Hubert M. 1991. *Understanding social inequality: Modeling allocation processes.* Newbury Park, CA: Sage.

Campbell, Donald T, and Julian C. Stanley. 1963. *Experimental and quasi-experimental designs for research.* Dallas, TX: Houghton Mifflin.

Cipes, M. 1983. Analysis of pay equity for men and women faculty at the University of Connecticut Health Center. Prepared for the Ad Hoc Committee on the Status of Women at the Health Center.

Clayton, Susan, and Faye Crosby. 1992. *Gender justice, and affirmative action.* Ann Arbor: University of Michigan Press.

Cockburn, Cynthia. 1991. *In the way of women: Men's resistance to sex equality in organizations.* Ithaca, NY: ILR.

Connell, R. W. 1987. *Gender and power.* Stanford, CA: Stanford University Press.

Dadalt, R. A. 1986. Pay equity analysis. Masters thesis, University of Connecticut Health Center Faculty.

Deaux, Kay, and Brenda Major. 1987. Putting gender into context: An interactive model of gender related behavior. *Psychological Review* 94:369-89.

England, Paula. 1992. *Comparable worth: Theories and evidence.* New York: Aldine De Gruyter.

Filer, R. 1983. Sexual differences in earnings: The role of individual personalities and tastes. *Journal of Human Resources* 18 (1): 82-99.

Geetter, Joan. 1988. Report of the University/AAUP salary inequity committee to the board of trustees. University of Connecticut, Storrs, CT.

Gray, Mary W. 1990. *Achieving pay equity on campus.* Washington, DC: American Association of University Professors.

———. 1993. Can statistics tell us what we do not want to hear? The case of complex salary structures. *Statistical Science* 8 (2): 144-79.

Gray, Mary W., and Elizabeth L. Scott. 1980. A "statistical" remedy for statistically identified discrimination. *Academe* 66: 174-81.

Gronowicz, Gloria, Nancy D. Adams, Andrea J. Apter, Elise Barbarese, Joan Bourbeau, Diana G. Myles, Christine Niekrash, Martha J. Radford, Susan Rersine, Jonelle C. Rowe, Paula Tempro, Myra M. Ferree, and Julia McQuillan. 1993. Report of the Working Group on the Status of Women Faculty at the University of Connecticut Health Center. Farmington, CT.

Haignere, Lois, Y. Lin, B. Eisenberg, and J. McCarthy. 1996. *Pay checks: A guide to achieving salary equity in higher education.* New York: United University Professions.

Hauser, Robert M., and Karl Oppenheim Mason. 1993. Faculty pay equity study: Gender and ethnicity pay equity analysis. University of Hawaii at Manoa, Office of the Senior Vice President for Academic Affairs.

Holmes-Rovner, M., E. Alexander, B. O'Kelly, W. Rome, S. Ming Wu, K. Lovell, and M. Alfano. 1994. Compensation equity between men and women in academic medicine: Methods and implications. *Academic Medicine* 69 (2)*:*13 1-7.

Ikeda, K. 1995. Evaluation research issues in women faculty pay equity adjustments: Evaluation of across-the-board over case-by-case equity adjustments. Paper presented at the 90th Annual Meeting of the American Sociological Association, August, Washington, DC.

Jacob, Mary C., Nancy D. Adams, Robert L. Jungas, and Clarence Trummel. 1996. Joint medical and dental councils committee on women's pay equity: Final report. University of Connecticut Health Center, Farmington.

Jacobs, J. A., and Ronnie J. Steinberg. 1990. Compensating differentials and the male-female wage gap: Evidence from the New York State comparable worth study. *Social Forces* 69:439-68.

Jones, F. L., and Jonathan Kelley. 1984. Decomposing differences between groups. *Sociological Methods & Research* 12 (3): 323-43.

Kanter, Rosabeth M. 1977. *Men and women of the corporation.* New York: Basic Books.

Major, Brenda. 1987. Gender, justice, and the psychology of entitlement. In *Sex and gender,* edited by P. Shaver and C. Hendrich. Newbury Park, CA: Sage.

———. 1989. Gender differences in comparisons and entitlements: Implications for comparable worth. *Journal of Social Issues* 45:99-115.

Norusis, M. J. 1988. *SPSS-X introductory statistics guide.* Chicago: SPSS.

Polachek, Solomon W. 1985. Women in the economy: Perspectives on gender inequality. In *Comparable worth: Issues for the 80's: A consultation of the U.S. Commission on Civil Rights.* Washington, DC: U.S. Commission on Civil Rights.

Reskin, Barbara F., and Patricia A. Roos. 1990. *Job queues, gender queues: Explaining women's inroads into male occupations.* Philadelphia: Temple University Press.

Scott, Elizabeth L. 1977. *Higher education salary evaluation kit.* Washington, DC: American Association of University Professors.

Smith, William. 1993-1994. *Report on medical school faculty salaries.* Washington, DC: American Association of Medical Colleges.

Steinberg, Ronnie J. 1995. Advocacy research for feminist policy objectives: Experiences with comparable worth. In *Feminism and social change: Bridging theory and practice,* edited by H. Gottfried. Chicago: University of Illinois Press.

Steinberg, Ronnie J., and Lois Haignere. 1987. Equitable compensation: Methodological criteria for comparable worth. In *Ingredients for women's employment policy,* edited by C. Bose and G. Spitze. Albany: State University of New York Press.

♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂♀♂

Myra Marx Ferree is professor of sociology and women's studies at the University of Connecticut. Among her recent publications are *Feminist Organizations: Harvest of the New Women's Movement,* coedited with Patricia Yancey Martin (1995),and various articles on gender; family, social movements, and feminism in the United States and Germany.

Julia McQuillan is finishing her PhD at the University of Connecticut and doing outcomes research focusing on mental health in the Department of Family Medicine. Her dissertation examines how gender and marital power explain the division of labor; perceived equity, and distress among dual earner couples.