Journal of Management
1996, Vol. 22, No. 4, 653-673

Latent Variable Models in the Investigation of Salary Discrimination: Theory and Practice

Roger E. Millsap
Ross Taylor
Baruch College

Statistical methods play an important role in salary discrimination litigation. Regression analysis is the most widely used method in this application. In regression analysis, group differences in salary are evaluated among employees who have been matched on measured "merit" variables (e.g., years of job experience). Difficulties often arise because important merit variables may be unmeasured, or because the available measures of merit are fallible. An alternative approach in this case lies in the use of several latent variable path models that have been proposed for the salary problem (Birnbaum, 1979; McFatter, 1987). The theoretical and practical implications of these models are discussed. The use of the models is illustrated in real data.

In the last several decades, statistical analyses have played an important role in legal decision-making (Finkelstein & Levin, 1990; Gastwirth, 1988). One area in which statistical evidence is often crucial is in employment discrimination litigation. Employment discrimination concerns discrimination against one or more protected groups in employee selection, promotion, compensation, or discharge. Here we will focus on the use of statistical evidence in studies of salary discrimination. A wide literature exists on this topic (Birnbaum, 1979; Conway & Roberts, 1983; Dempster, 1988; Goldberger, 1984; Gollob, 1984; McFatter, 1987; Peterson, 1986; Schafer, 1987). Our goal is to describe the use of latent variable models as an alternative to traditional methods in studies of salary fairness. Both the theoretical implications of these models and their practical use will be discussed. We begin with a brief review of the legal background and the definition of salary fairness.

Salary Discrimination and the Law

The allocation of salary and other forms of compensation is regulated by a number of statutes. Title VII of the Civil Rights Act, the Equal Pay Act, and the

Direct all correspondence to: Roger E. Millsap, Baruch College-CUNY, Department of Psychology, 17 Lexington Avenue, New York, NY 10010. E-mail: sapbb@cunyvm.cuny.edu.

Age Discrimination in Employment Act are important examples. Title VII prohibits discrimination against any individual with respect to hiring, discharge, compensation, terms, or conditions of employment because of the individual's race, color, religion, sex, or national origin. The Equal Pay Act prohibits sex-based wage discrimination between jobs that require equal skill, effort, and responsibility, and that are performed under similar working conditions. Note that this Act concerns pay differences between jobs. Pay differences are deemed acceptable only if the differences are based on either a seniority system, a merit-based system, a system that measures earnings by quantity or quality of production, or a factor other than gender. The Age Discrimination in Employment Act prohibits age discrimination in the hiring, discharge, promotion, or treatment of persons over the age of 40. Age may be considered by the employer when creating a seniority system or an employee benefit plan, or when age is a Bona Fide Occupational Qualification (BFOQ). In order for an employer to claim age as a BFOQ, it must be shown that an age limit is necessary to the essence of the business, and that anyone excluded on the basis of age is in fact disqualified.

Under the law, salary discrimination may arise either from "disparate treatment" or "disparate impact." Disparate treatment refers to intentional, willful discrimination against members of a protected class. For example, rules that explicitly provide higher compensation for men than women would create disparate treatment. Disparate impact ordinarily refers to unintentional discrimination against members of a protected group. In this case, the employer's actions appear to be fair in form, or "facially neutral," but these actions result in discrimination against the protected class. For example, various neutral indicators of merit (e.g., tenure, education) may have been considered by the employer in allocating salaries, yet disparities in average salaries across groups may result. Statistical evidence can play a deciding role in such cases.

One form of salary discrimination that will not be discussed in this paper concerns issues of comparable worth. In comparable worth litigation, discrimination against a protected group is alleged to exist due to disparities in pay across different job titles. The different jobs appear to be equal in organizational value and complexity; discrimination occurs because members of the protected group tend to occupy the lower-paid job title. The unit of analysis in such cases is the job, rather than the individual employee. In contrast, our concern in this paper lies with discrimination against groups of employees within a single job title or job class.

Fairness Definitions

Before investigating salary discrimination, a clear definition of salary fairness must be formulated. Millsap and Meredith (1994) distinguished two types of fairness. Both types assume that salary should be allocated on the basis of one or more "merit" variables that in some way determine the employee's worth for salary purposes. These merit variables may, or may not, be readily measurable. Both types of fairness also require that employees be matched on merit prior to comparing groups, but the two differ according to the type of merit considered.

True fairness is defined to hold when among employees whose "true merit" is identical, there are no group differences in salary distributions. True merit includes all of the information that is relevant for the determination of salary, regardless of whether this information is directly measured. True fairness can be formally stated, using conditional probability, as

$$Pr(Y \mid W = w, V = v) = Pr(Y \mid W = w), \tag{1}$$

where Y is salary, W is true merit, V is the group membership variable, and $Pr(X \mid U = u)$ is the conditional probability of X given that the variable U assumes the value u. True merit W and group membership V may be multivariate in the above. The second type of fairness, *observed fairness*, holds when among employees whose observed merit is identical, there are no group differences in salary distributions. Observed merit includes the measures of merit that are available for study. Formally, observed fairness holds if

$$Pr(Y \mid Z = z, V = v) = Pr(Y \mid Z = z), \tag{2}$$

where Z is observed merit (possibly multivariate). These two definitions of fairness are distinct because observed merit often differs from true merit in practice. Measurement error, omitted variables, and measurement bias all contribute to the distinction between true and observed merit.

These two definitions are consistent with much of the literature on salary fairness (Birnbaum, 1979; Finkelstein & Levin, 1990; Gastwirth, 1988; Gollob, 1984; Levin & Robbins, 1983; McFatter, 1987; Peterson, 1986), but other definitions are possible. A different approach would match employees on salary rather than merit, and define fairness by invariance across groups in the merit distributions (Birnbaum, 1979; Conway & Roberts, 1983). This definition is generally inconsistent with the two definitions considered above. Another approach retains the idea of matching employees on merit, but does not require that the entire salary distributions be the same across groups. For example, one might simply require that the average salaries be identical. Gregory (1991) discusses alternative definitions of this type. None of these alternative approaches will be considered here.

Given that true and observed merit differ, are there conditions under which the two forms of fairness are equivalent? This question is important because statistical investigations of salary fairness are generally designed only for the study of observed fairness. The implications of these investigations for true fairness are often unclear. Millsap and Meredith (1992, 1994) described some conditions under which observed and true fairness are equivalent. Under these conditions, conclusions about observed fairness have direct implications for true fairness. We will return to these conditions below when describing the latent variable path models that have been proposed for salary investigations.

In the next section, we review some traditional methods of statistical analysis that have been applied to the salary fairness problem. The difficulties facing these methods in reaching conclusions about true fairness are discussed. We then describe some path models that have been proposed for the salary problem. The implications of these models for traditional methods of analysis, and for the investigation of true fairness, are discussed.

Statistical Method Review

Group differences in average salaries among employees holding the same job could arise for reasons other than discrimination. Commonly, these groups will differ on other variables that are relevant to salary allocations. For example, men and women may differ in average salary but may also differ in the amount of previous job experience. Hence the question becomes: are the group salary differences larger than we would expect, given the differences on the available merit variables? The answer to this question depends on how one "adjusts" for differences in merit. Statistical methods are commonly used in making these adjustments; regression methods in particular are widely used.

Regression Methods

In regression analysis, the salary Y is regressed on an indicator variable V that codes group membership, and on one or more merit measures Z. For a single measure of merit, the regression equation is

$$Y = \beta_0 + \beta_v V + \beta_z Z + e \tag{3}$$

where β_0 is the regression intercept, β_z, β_v are regression coefficients, and e is a regression residual. If multiple merit measures are available, the equation can be expanded to include the additional variables. Observed unfairness in the salary allocation is evaluated by examining β_v. A nonzero value for the estimate of this coefficient is taken to indicate unfairness in the salary allocation, and is tested for statistical significance. A nonzero value for β_v indicates that group differences in salary exist among employees with identical measured merit Z. The adjustment for merit comes about by including measured merit in the regression equation.

One difficulty with the above approach is that the regression of salary on measured merit is assumed to be identical across groups. Under this assumption, if one were to perform separate regressions of salary on measured merit within each group, the regression lines would be parallel. The only difference between the groups in these regressions must lie in the intercepts of the regression lines. The practical implication of this assumption is that unfairness, if present, simply acts as a constant additive increment in salary levels. The size of this increment must not depend on employee merit. This form of unfairness may be common, but more complex forms are possible in which unfairness depends on merit in some way. If present, this type of unfairness may not be detected in the above regression.

An alternative regression procedure that eliminates the additivity assumption is to regress salary Y on measured merit Z separately within each group being compared. Observed fairness is then assessed by testing whether the regression equations are identical across groups. The regressions are identical only if both the regression intercepts and coefficients are identical across groups. Observed fairness also requires that the residual variances about regression be the same across groups. Tests for equality of regression equations are available (Ali & Silver, 1985; Marascuilo & Levin, 1983), but are not part of standard statistical software.

One controversy in the use of regression analysis for salary applications lies in the choice of the criterion variable for the regression. In Equation 3, salary Y is chosen as the criterion. This approach is known as *direct regression* or *forward regression* in the literature (Finkelstein & Levin, 1990; Gastwirth, 1988). An alternative *reverse regression* approach uses measured merit (or a composite of several merit measures) as the criterion, with group membership V and salary Y as predictors (Birnbaum, 1979; Conway & Roberts, 1983). In this reverse regression, unfairness is indicated by a nonzero value for the regression coefficient for group membership, as in direct regression. If both regressions are performed on the same data, the two methods need not lead to the same conclusion regarding fairness (Birnbaum, 1979). For example, direct regression may indicate that among employees of equal merit, the men's average salary exceeds that given to women. On the other hand, the reverse regression may show that among employees receiving identical salaries, men have higher scores on merit than women. The contradiction arises in part because the reverse regression approach is based on a definition of fairness that is different from the definition underlying the direct regression. In general, the reverse regression approach does not provide a useful way of testing observed fairness as defined earlier.

All regression methods require five assumptions, any of which could be violated in practice. These assumptions are:

(Linearity) The salary/merit relationship must be well-approximated by a straight line.

(Homoscedasticity) The variability of salaries around this straight line must not depend on merit, with higher variance at some merit values than at others.

(Conditional Normality) Within any group and among employees with common merit values, the distribution of salaries should resemble a normal distribution.

(Reliability) Merit variables are measured without measurement error.

(Specification) No merit variables that are important to the salary allocation are left out of the regression equation.

Collectively, these assumptions are highly restrictive. The first three assumptions can be checked using the available data and standard methods (Draper & Smith, 1981). The fourth and fifth assumptions are more difficult to evaluate, and are often controversial in actual applications. The choice of which merit measures to include in the regression is highly important. In the next section, we discuss the difficulties encountered when these assumptions may be violated.

Regression is not the only statistical method that has been usefully applied to the salary discrimination problem. Levin and Robbins (1983) presented a method based on a less restrictive statistical model (an "urn model") that allows the investigator to test whether the salary allocation is related to group membership after removing the portion of the group difference that is related to merit. The advantage of this approach is that it requires fewer assumptions. The adjustment for merit can take a variety of forms, and is not limited to a straight-line relationship. The reliability and specification assumptions are still required, however, because the method assumes that the adjustment for merit truly removes any salary differences due to merit.

Measurement Errors and Omitted Variables

Under the above assumptions, the measures of merit used in the regression analysis must be perfectly reliable and must contain all of the important determinants of salary. These requirements are difficult to meet in practice. Some measures of merit that may be relevant to the salary allocation, such as supervisory performance evaluations, are seldom perfectly reliable. Frequently, the available measures of merit do not include all of the merit information that affects salary. For example, a merit measure that is important for university faculty salary determination is the number of publications produced by the employee. The quality of these publications may also be important, but formal measures of quality are usually unavailable. Either measurement error or the omission of important merit variables can lead to inaccuracies in the resulting regression coefficient estimates. These inaccuracies may in turn lead to spurious conclusions of fairness, or unfairness, in the salary allocation.

To illustrate the influence of measurement error, Table 1 presents the results of a small simulation that included two groups of "employees" (100 per group) whose salaries were regressed on a single measure of merit and a dummy-coded indicator for group membership, as in Equation 1. The reliability of the merit measure was systematically varied to study its impact on the estimate of β_v, the coefficient for group membership. The salary and merit data were generated to fit a common factor model in which the latent variable plays the role of "true merit" (Birnbaum, 1979). The measure of merit was thus an imperfect indicator of true merit in this simulation. A path diagram for this model is given in Figure 1. The factor structure was held constant between groups, making the true value of β_v equal to zero. The salary distribution was fair in relation to true merit: among employees with equal true merit, there are no group differences in salary.

The nonzero values for the estimates of β_v in Table 1 represent the distortion in these estimates created by the presence of measurement error in the merit measure. As the reliability of the merit measure becomes smaller, the coefficient

Table 1. Regression Coefficient Estimates in Simulated Data under Varying Merit Measure Reliability

Reliability	β_v	β_z*
.90	1.66	4.35
.80	1.73	4.17
.70	4.15	3.27
.60	4.41	2.99
.50	5.19	2.61

Note: *β_v is the regression coefficient for group membership. Its true value is zero. β_z is the coefficient for merit. Its true value is 5.0.

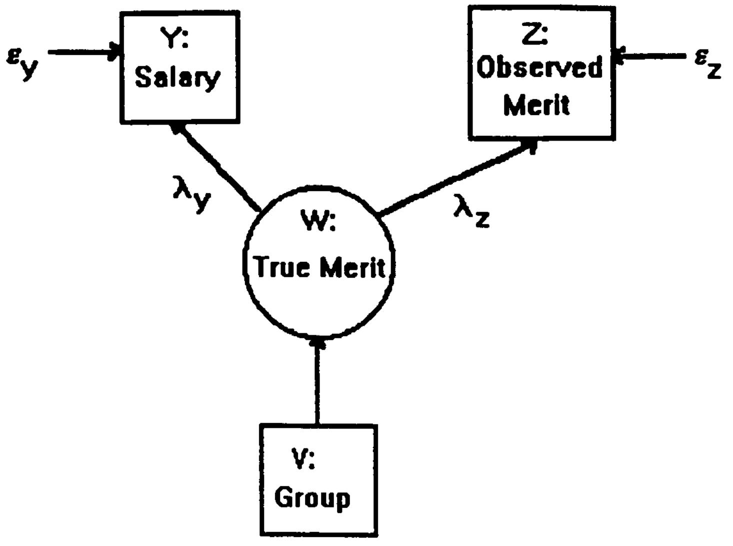

Figure 1. A Common Factor Model

estimate for group membership becomes larger. This distortion is meaningful even with a reliability of .90, as the coefficient estimate of 1.66 is statistically significant. Note that the unreliability in the measure of merit has the expected effect on the regression coefficient estimate β_Z: as the reliability becomes smaller, the coefficient estimate also becomes smaller. A regression analysis of these data would suggest unfairness, although the salary distribution is fair.

The omission of important merit variables from the regression analysis can also distort the estimate of the regression coefficient β_V. This topic has been studied intensively (Holland, 1986; Mauro, 1990; Rosenbaum, 1995). The omission may lead to a nonzero estimate when the true value is zero, as in the simulations. Conversely, the omission may mask the presence of unfairness by reducing the size of the estimates relative to their true values. In this case, the regression analysis would fail to detect the unfairness that exists in the salary allocation. Two conditions are required in order for either of these distortions to result from omitted variables:

1. The omitted variables must be associated with salary, even after controlling statistically for merit variables included in the regression analysis, and
2. The omitted variables must be related to the variables included in the equation (both merit and group membership).

The second condition implies that the groups must differ on the omitted variables (e.g., differences in means). The first condition implies that the measures of merit included in the regression must not exhaust all of the merit information that is relevant to the salary distribution. These conditions are not highly restrictive, and may hold fairly often in practice.

To illustrate the effect of the omission of merit variables that fulfill the above two conditions, Table 2 gives a correlation matrix for four variables in a sample of 947 employees. These data are described in more detail below; only a brief description will be given here. *STSAL* is starting monthly salary, which serves as the criterion variable in the direct regression analysis. *EDUC* is the educational level of the employee measured on a 0-7 scale. *AGEGRP* is a dichotomous variable based on the employee's age. Employees who are at least 40 years of age received a "1," and all other employees received a "0." *AGEGRP* serves as the demographic variable (*V*) for the analysis. Finally, *PRSAL* is the employee's monthly salary on his or her previous job. *EDUC* and *PRSAL* are the observed merit measures.

In a direct regression analysis of these data, *AGEGRP*, *EDUC* and *PRSAL* serve as the predictors. As noted earlier, the estimated regression coefficient for *AGEGRP* is used to make judgements about observed fairness. The results of the direct regression analysis give an estimated regression coefficient for *AGEGRP* of 127.90. This estimate means that on average, after adjusting for *EDUC* and *PRSAL*, the older employees make about $128 per month more than the younger employees. Suppose however that *PRSAL* is omitted from the analysis. How would the results change? If *PRSAL* is omitted and the data in Table 2 are again used to perform a direct regression using only *EDUC* as the merit measure, the estimated regression coefficent for *AGEGRP* is 370.39. The estimated age difference in monthly starting salary becomes about three times as large if *PRSAL* is omitted from the analysis.

The effect of omitting *PRSAL* from the analysis depends on the correlations among the four variables, and hence the above results illustrate only one scenario. It is useful to illustrate some additional scenarios that would arise under different correlational structures. Table 3 presents these results. In all cases, it is assumed that the intercorrelations among *STSAL*, *EDUC*, and *AGEGRP*, and the correlation between *PRSAL* and *EDUC*, remain as in Table 2. For this reason, the regression coefficent for *AGEGRP* in an equation that omits *PRSAL* is 370.39, as above. Table 3 presents the regression coefficient for *AGEGRP* in the full equation that

Table 2. Correlation Matrix for the Omitted Variable Example

	STSAL	*EDUC*	*AGREGRP*	*PRSAL*
STSAL	1.0			
EDUC	.1719	1.0		
AGREGP	.2927	.1747	1.0	
PRSAL	.5743	.1042	.3415	1.0

Notes: STSAL = Starting salary, EDUC = Education, AGEGRP = age of employee (under 40 = "0", 40 and over = "1"), PRSAL = Previous salary

Table 3. Results of the Omitted Variable Analysis for AGEGRP Regression Weight

Correlation		
STSAL/PRSAL	*AGEGRP/PRSAL*	*AGEGRP Regression Weight*
.5743	.3415	127.90
.70	.3415	62.88
.80	.3415	37.02
.40	.3415	218.06
.5743	.40	79.67
.5743	.50	−15.52
.5743	.60	−140.76
.5743	.20	234.27

Notes: STSAL = Starting salary, PRSAL = previous salary, AGEGRP = age of employee (under 40 = "0", 40 and over = "1").

includes *PRSAL*, under varying assumptions about the correlations between *PRSAL* and either *AGEGRP* or *STSAL*.

The main trend to be noted in Table 3 is that as the correlation between *PRSAL* and either *AGEGRP* or *STSAL* increases, the effect of omitting *PRSAL* becomes larger. For example, if the correlation between *PRSAL* and *AGEGRP* is .50, the regression coefficient for *AGEGRP* would be -15.52 in a full equation that includes *PRSAL*. In this case, the full equation would indicate a low level of observed unfairness toward the older employee group. The amount of unfairness here is probably insignificant in practical terms. But as noted above, when *PRSAL* is omitted from the analysis, the *AGEGRP* regression coefficient reverses its sign and becomes 370.39. This estimate indicates substantial unfairness toward the younger employee group, a conclusion that is entirely at odds with the conclusion suggested by the full regression analysis. These results demonstrate that the omission of observed merit variables can have important consequences for conclusions drawn from a direct regression analysis.

Latent Variable Models

The difficulties created by measurement error and omitted variables in regression analysis have led researchers to consider more complex models for salary discrimination (Birnbaum, 1979; Goldberger, 1984; McFatter, 1987). These models have usually taken the form of path models, often including unobserved or latent variables. Path analysis is the simultaneous analysis of a set of regression equations. These equations correspond to the network of hypothesized relationships among the variables being studied. Measurement error is incorporated in the path model by including latent variables to represent the "error-free" constructs that are measured by the fallible observed variables. A full description of path analysis is not given here, as many references are available (e.g., Bollen, 1989).

An important motivation for considering path models in the salary discrimination problem has been to understand the conditions under which ordinary regression analysis, such as the direct regression discussed earlier, will correctly indicate true fairness, or unfairness. Each path model has implications for how ordinary regression analysis would perform if applied to data that follow the proposed model. Regression analysis can be shown to perform adequately under some path models, while giving misleading results under others. The results given by a regression analysis depend heavily upon the model that underlies the variables under study.

The path model represents the hypothesized relationships among the salary, merit, and group membership variables. A latent variable is often included in the model to represent true merit. Group membership may be represented in the model by a group-membership indicator, or may be implicitly represented by specifying separate models in each group, with simultaneous analyses across groups. Different path models for the salary problem are distinguished by which variables in the model are given direct path linkages, and which variables are linked indirectly. Two path models that have appeared in the literature for the salary discrimination problem are the models proposed by Birnbaum (1979) and McFatter (1987).

Birnbaum's (1979) Model

The "one-mediator" model presented by Birnbaum (1979) was the first path model proposed for the salary problem. A path diagram of the model is given in Figure 1. In this model, true merit is a latent variable that acts as a common factor underlying both salary and the observed merit measures. True merit is represented by the circle in Figure 1, with direct paths to both salary and observed merit, each represented by boxes. Measurement error is represented by unique factors that directly influence the observed variables (the ε in Figure 1). Because of these similarities to ordinary factor analysis, the model will here be denoted the *common factor model*. Group membership has a direct path to true merit in the model, but may or may not be directly related to salary. The absence of any direct path from group membership to salary implies that true fairness holds for the salary distribution. Conversely, a direct path indicates unfairness in the salary distribution: group differences in salary remain after controlling for true merit.

This common factor model has two essential features that distinguish it from other path models:

1. The relationship between salary and the observed merit measures arises solely through the common influence of true merit. The observed merit measures have no direct or independent influence on salary.
2. Group differences in the observed merit measures arise solely from group differences in true merit. The observed merit measures are unbiased measures of merit (Millsap & Meredith, 1992).

The first feature implies that salary is determined by true merit, rather than by the employee's status on the measures of observed merit. The apparent relationship between observed merit and salary is produced by the common influence of true merit. This feature would rule out any explicit determination of salary on the basis of observed merit, for example. The second feature prohibits any measurement bias in observed merit as an indicator of true merit. The possibility of bias in the observed merit measures is sometimes a point of dispute in discrimination litigation (Finkelstein & Levin, 1990).

Under the common factor model, ordinary regression analysis such as direct regression can give misleading results. Suppose that the model holds with no direct path between group membership and salary. True fairness holds in this situation. It can be shown mathematically that regression analysis will indicate that observed fairness does not hold for the salary distribution in this case (Meredith & Millsap, 1992; Millsap, 1995). This problem was illustrated earlier in the simulation results given in Table 1. The data for this simulation were generated under the common factor model with true fairness in the salary distribution. Birnbaum (1979) further illustrates the problem with examples of both direct and reverse regression analysis under the common factor model.

McFatter's (1987) Model

A second path model for the salary discrimination problem was proposed by McFatter (1987), and is shown in Figure 2. In this model, true merit again serves as a latent variable or common factor that underlies measured merit. Unlike the common factor model however, true merit bears no direct relationship to salary. Instead, salary is influenced directly by observed merit. True fairness in the salary distribution is defined by the absence of any direct path between group-membership and salary.

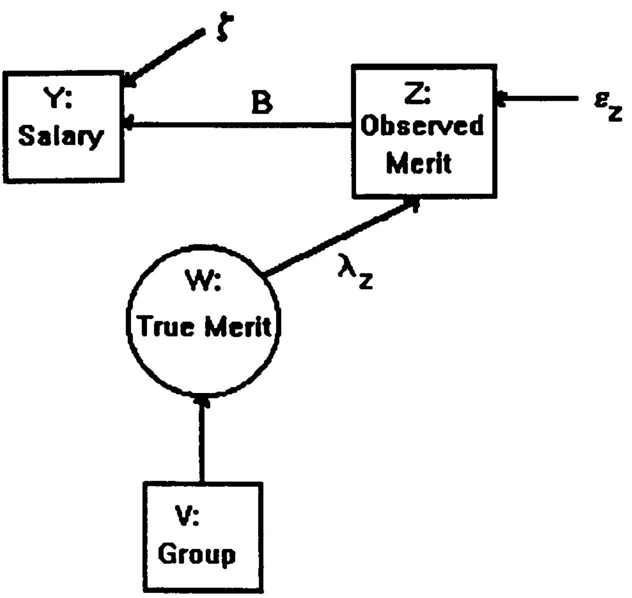

Figure 2. The McFatter (1987) Model

McFatter's model has two essential features that distinguish it from other path models:

1. No direct paths are present in the model between true merit and salary. Although the observed merit measures may be fallible indicators of true merit, salary is influenced by the observed merit measures directly.
2. The observed merit measures are unbiased as measures of true merit. The model shares this feature with the common factor model.

McFatter (1987) argued in favor of the first feature by noting that salary allocations are usually based on available information, even if this information is unreliable. For example, unreliable supervisory evaluations may still be considered in determining salary. McFatter's model assumes that the observed merit measures include all of the information relevant to the salary allocation. This assumption stands in direct contrast to the common factor model in which the observed merit information has no direct role in salary determination.

If McFatter's (1987) model holds, it can be shown that a direct regression analysis will correctly identify true fairness or unfairness in the salary distribution. Ordinary regression analysis is diagnostic in this case for the presence or absence of a direct path between group membership and salary. Under McFatter's model, fallible measures of observed merit may be used successfully in a regression analysis to study fairness. Omitted variables may still create problems, however, because the model assumes that the observed measures of merit contain all of the information about true merit that is relevant to salary. The conditions under which omitted variables will create problems for the model are identical to those discussed earlier in the context of ordinary regression analysis.

Both of the above path models require statistical assumptions similar to those required in ordinary regression analysis. Linearity, homoscedasticity, and conditional normality are required, for example. On the other hand, path analysis has at least two advantages in comparison to ordinary regression analysis. First, measurement error can be incorporated in the model, as already noted. Secondly, path analysis permits both direct and indirect relationships to be studied. Ordinary regression analysis studies only direct relationships. These advantages provide the motivation for the application of path analysis to salary problems.

Neither of the above models was originally specified in sufficient detail to permit implementation in available software such as LISREL (Jöreskog & Sörbom, 1989). The Appendix describes the LISREL specification for both models. Further information can be found in Millsap and Meredith (1994). The path models must be sensitive to group differences in mean salary levels. It is therefore necessary to include latent mean structures in the path model. Group membership is handled by using multiple group models with equality constraints across groups as needed. We next illustrate the use of the two path models in real data.

An Example

The data to be analyzed were provided by a large investment firm with offices nationwide. Salary and merit data were collected on 947 newly-hired investment brokers. The salaries are starting salaries (*STSAL*). The merit measures to be used are education (*EDUC*) and previous salary (*PRSAL*). As noted earlier, education is measured on a 0-7 scale. Previous salary is the employee's salary on his or her previous job. Both of these variables were considered by the employer in setting starting salaries, although no explicit formula was used. According to company representatives, the starting salary was typically anchored "near" the applicant's previous salary, and was then adjusted up or down depending on additional information. The grouping variable to be considered is based on age. Age is measured in years, but will be dichotomized as described earlier to create the grouping variable *AGEGRP*. The threshold of 40 is chosen to correspond to the typical legal cutpoint in age discrimination litigation.

Table 4 gives descriptive statistics for both age groups. Starting and previous salaries are expressed as monthly salaries in this Table. In terms of annual salary, the older employees average about $4800 more per year than do the younger employees. The older employees also have more education on average, and tend to have made higher salaries on their previous jobs. It is not obvious from the descriptive statistics whether the gap in current starting salaries can be attributed to the gaps in education and previous salaries.

Regression Analyses

The regression analysis was performed using the monthly salary metric for both current and previous salaries. A direct regression was performed as in Equation 3, with the two merit variables (*PRSAL* and *EDUC*) and the grouping variable *AGEGRP* as predictors. The resulting estimated regression equation was:

$$\hat{Y} = 1633.85 + .18(PRSAL) + 86.69(EDUC) + 127.91(AGEGRP).$$

All regression coefficients in this equation were statistically significant ($p < .05$), with a multiple correlation of .59. The coefficent estimate for *AGEGRP* indicates that after adjusting for observed merit, older employees make about $1536 more

Table 4. Descriptive Statistics for Salary and Merit Variables: Monthly Salary Metric

		Old (N = 207)				
		Correlations				
		STSAL	*EDUC*	*PRSAL*	*Mean*	*SD*
	STSAL	1.0	.1569	.3373	2865.13	585.39
Young	*EDUC*	.1169	1.0	.0667	3.24	.84
N = 740	*PRSAL*	.5993	.0393	1.0	4623.49	1827.40
	Mean	2464.99	2.96	3228.14		
	SD	527.70	.58	1515.59		

Notes: STSAL = Starting salary, EDUC = Education, PRSAL = Previous salary. Correlations, means, and standard deviations for employees in the "40 and over" group are above the diagonal; correlations, means, and standard deviations for the "under 40" group are below.

annually in salary than do the younger employees. Observed fairness does not hold in these data, as younger employees receive lower salaries than would be expected on the basis of observed merit.

Latent Variable Path Analyses

All path analyses were performed using the LISREL VII program (Jöreskog and Sörbom, 1989). Both monthly salary variables (*PRSAL* and *STSAL*) were transformed using natural logarithms prior to their use in the LISREL analyses. Table 5 gives the descriptive statistics for the log-tranformed salary variables (*LPRSAL* and *LSTSAL*) and *EDUC*.

Models were evaluated for goodness-of-fit using the fit indices provided by LISREL, and two other indices. The relative noncentrality index (RNI) (McDonald and Marsh, 1990) is a comparative fit index that measures the fit of the model under study relative to a baseline or null model. The null model used here is described in the Appendix. This null model gave $\chi^2 = 566.14$, df = 12. RNI values above .90 are generally considered to indicate a good fit. The second fit index used here is the root-mean-square error of approximation (RMSEA) (Steiger, 1990). This index estimates the root-mean-square error in approximating the true population covariance matrix by the "best-fitting" covariance matrix under the proposed model, adjusted for degrees of freedom (Browne & Cudeck, 1993). Browne and Cudeck (1993) suggest that values of the RMSEA below .05 indicate "close" fit, and values between .05 and .08 indicate adequate fit. Further information on these fit indices (and many others) can be found in Bollen and Long (1993).

The path analyses began by fitting the common factor model that assumes true fairness, using a multiple-groups analysis as described in the Appendix. This model did not fit well ($\chi^2 = 85.91$, df = 7, RNI = .86, RMSEA = .11). As an alternative model, the version of the McFatter model that assumes true fairness was fit to the data. This model also failed to fit adequately ($\chi^2 = 107.19$, df = 7, RNI = .82, RMSEA = .12).

Table 5. Descriptive Statistics for Salary and Merit Variables: Ln(salary) Metric

		Old (N = 207)				
		Correlations				
		LSTSAL	*EDUC*	*LPRSAL*	*Mean*	*SD*
	LSTSAL	1.0	.1568	.2863	7.94	.20
Young	*EDUC*	.1233	1.0	.0419	3.24	.84
N = 740	*LPRSAL*	.5543	.0465	1.0	8.33	.51
	Mean	7.79	2.96	7.96		
	SD	.20	.58	.50		

Notes: LSTSAL = Starting monthly salary in ln units, EDUC = Education, LPRSAL = previous salary in ln units. Correlations, means, and standard deviations for employees in the "40 and over" group are above the diagonal; correlations, means, and standard deviations for the "under 40" group are below.

One obvious explanation for the lack of fit in these models is that true fairness does not hold. To investigate this possibility, the common-factor model was modified to permit different intercept parameters for *STSAL*, as described in the Appendix. This model failed to reach convergence after 500 iterations, indicating poor fit. As an alternative, the McFatter model was modified to permit different intercepts. This model converged, but did not fit well (χ^2 = 84.85, df = 6, RNI = .86, RMSEA = .12).

Finally, we returned to the common-factor model that specifies true fairness, and modified this model in response to the modification indices in the LISREL output. One modification was strongly indicated: the invariance restriction on the unique variance for *EDUC* was too restrictive. We eliminated this restriction, allowing this parameter to assume different values for the two groups. This modified model still included true fairness. The model gave a good fit (χ^2 = 32.13, df = 6, RNI = .95, RMSEA = .07). The implication is that true fairness holds, but that education functions differently in the two groups as a measure of merit, and is biased in this sense (Millsap & Meredith, 1992).

The parameter estimates for this final model are given in Table 6, and a path diagram is given in Figure 3. Figure 3 displays the model within a single age group, with subscripted parameters permitted to vary in value across groups. The triangles in the diagram represent constant terms that create the means or intercepts for both latent and observed variables. In Table 6, the estimate for the latent mean κ is higher in the older group, as is the unique variance for education. These results imply that the regression equation for predicting salary from previous

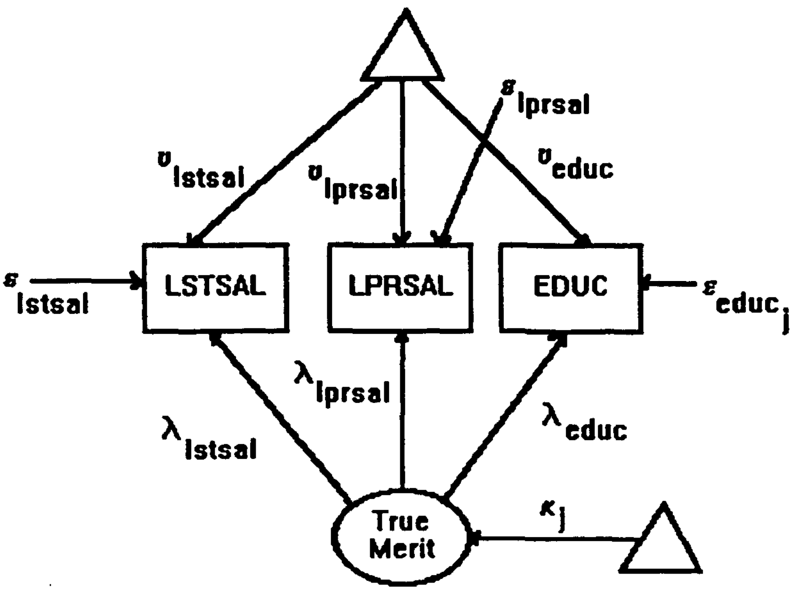

Figure 3. Final Common Factor Model for Example Data

Table 6. Parameter Estimates for a Common-Factor Model of Starting Salary, Previous Salary, and Education

Parameter	*"Old" estimate*		*"Young" estimate*
υ_{LSTSAL}		4.257	
υ_{LPRSAL}		0*	
υ_{EDUC}		.555	
λ_{LSTSAL}		.443	
λ_{LPRSAL}		1.000*	
λ_{EDUC}		.304	
Θ_{LSTSAL}		.018	
Θ_{LPRSAL}		.142	
Θ_{EDUC}	.692		.326
κ_j	8.325		7.966
σ^2_{wj}	.086		.122

Notes: *Fixed for identification. LSTSAL = Starting monthly salary in ln units, EDUC = Education, LPRSAL = Previous monthly salary in ln units.

salary and education is not the same for the two groups. This condition violates an assumption of direct regression, and suggests that direct regression is inappropriate for these data.

Discussion

The problems created by measurement error and omitted variables in statistical investigations of salary fairness have led researchers to consider path models with latent variables. We have described two of these models, and illustrated their use in real data. The models serve at least two purposes. First, they can be used to test the conditions under which true fairness may be investigated using tests of observed fairness. These conditions were described by Millsap and Meredith (1994). Second, the models may provide direct tests of true fairness.

The two models described here are not the only possible latent variable models for the salary fairness problem. A third model was presented by Goldberger (1984), and was discussed by Millsap and Meredith (1994). This model can be specified as a MIMIC model, with the intervening latent variable being true merit. Salary is the measured outcome variable in this model. The observed merit measures are exogenous. A unique feature of this model is that it does not require unbiased measures of merit. This model will not be fully identified in most salary investigations because salary is usually a scalar measure. Although this model cannot be fully tested, Millsap and Meredith (1994) describe how portions of the model might be tested.

Two practical problems will be encountered in using latent variable models in salary investigations. One problem concerns sample size. Latent variable models generally require substantial sample sizes for adequate estimation and testing. Salary fairness applications do not always meet these requirements, as protected groups may be underrepresented in the employee population. In small samples, latent variable models are more likely to yield improper parameter estimates. A second problem concerns the number of available merit measures. As discussed in the Appendix, the number of such measures places limitations on the variety of models that may be tested. If only a single merit measure is available, it may be difficult to reject either of the two models described here. Fortunately, multiple measures of merit are available in most salary investigations.

The problem of omitted variables is not fully solved by the use of path models in latent variables. In the common-factor model for example, the introduction of additional measures of merit will create no difficulty if these measures load on the same "true merit" factor as those already present in the model. Presumably, the additional measures would not alter the relationship of salary to true merit. Complications ensue if the additional measures load on a separate "true merit" factor, however. The relationship of salary to any additional merit factors may differ across groups, violating true fairness. The McFatter model is more vulnerable to the omission of important merit indicators. Even if the additional indicators load on the merit factor along with the existing measures, the model requires invariant paths between the additional indicators and salary. The unique variances for these additional indicators must also be invariant.

Latent variable models represent a fully parametric approach to the salary fairness problem. Distributional assumptions are required if the models are to have implications for true fairness in Equation 2, for example. An important direction for future research lies in the development of semi-parametric methods for the investigation of salary fairness. The urn model approach of Levin and Robbins (1983) is a good example. This model does not provide for measurement error or omitted merit variables, however. An ideal method would require few parametric assumptions, yet would be robust in the presence of measurement error or omitted merit variables.

The use of latent variable models should not be a substitute for careful thought and scrutiny of the data. The researcher should be familiar with the variety of models that have been proposed for salary fairness applications. These models are potentially useful tools for investigating salary fairness, but their limitations must be recognized. The models may answer the question of whether the data are consistent with a pattern of discrimination. The models do not answer questions about intent, and do not provide causal explanations for the patterns found in the data.

Acknowledgement: This article is based on a talk given by the first author during the RMD Conference on Causal Modeling at Purdue University, West Lafayette, Indiana, on March 3-5, 1994.

Appendix

The path models for the salary discrimation problem are specified as multiple-group latent variable models with structured means. Groups are defined by V, the demographic indicator(s) under study. Equality constraints that operate across groups are used to impose invariance restrictions on parameters that should be invariant under the model.

Common-factor models. Let the number of observed merit measures in Z be p, and let salary Y be a scalar measure. Let true merit W also be a scalar (generalizations with multivariate W are possible). The common-factor model is specified by the following equations:

$$Y_j = \nu_y + \lambda_y W_j + \varepsilon_y, \tag{4}$$

$$Z_j = \nu_z + \lambda_z W_j + \varepsilon_z, \tag{5}$$

$$E(Y_j) = \nu_y + \lambda_y \kappa_j, \tag{6}$$

$$E(Z_j) = \nu_z + \lambda_z \kappa_j, \tag{7}$$

$$Cov(\varepsilon_z, \varepsilon_y) = Cov(W_j, \varepsilon_y) = Cov(W_j, \varepsilon_z) = 0, \tag{8}$$

$$Var(W_j) = \sigma_{wj}^{2}, \tag{9}$$

$$Var(\varepsilon_y) = \Theta_y, \tag{10}$$

$$Var(\varepsilon_z) = \Theta_z. \tag{11}$$

Equations 4 and 5 give the factor analytic model for salary Y and observed merit Z. The subscript j denotes group membership. The factor loadings λ and the intercepts ν are not subscripted, as these parameters are invariant across groups. The unique factor scores ε have covariance matrices in Equations 10 and 11 that are also invariant. The expected values for salary and observed merit are given in Equations 6 and 7. The latent mean κ may vary across groups. Group differences in the variance of true merit in Equation 9 are also permitted.

Under standard normality assumptions on W and (ε_y, ε_z), the common-factor model represented in Equations 4–11 can be shown to imply true fairness. Some additional constraints are needed to achieve identification. For this purpose, one factor loading may be fixed to a nonzero value, with the corresponding intercept fixed to zero (Millsap & Everson, 1991). Adopting these constraints, and assuming that two groups are compared, there will be $3p + 5$ independent parameters to be estimated. The degrees of freedom will be $p^2 + 2p - 1$.

The true fairness assumption requires that ν_y, λ_y and Θ_y be invariant across groups. This assumption can be examined by weakening any of these restrictions, and observing the improvement in fit. The conditional mean for salary, given true merit, is given in Equation 6 by replacing κ_j with W_j. The conditional variance for salary is the variance in Equation 10. The number of observed merit measures may limit the number of restrictions that can be eliminated. For example, with a single measure of merit ($p = 1$), the model with true fairness has two degrees of freedom. In this case, only one restriction can be eliminated if the resulting model is to be testable.

The McFatter Model. Using notation similar to that used earlier, the equations that describe the McFatter model are:

$$Y_j = \alpha + \beta' Z_j + \delta, \tag{12}$$

$$Z_j = \nu + \lambda W_j + \varepsilon, \tag{13}$$

$$E(Y_j) = \alpha + \beta'\nu + \beta'\lambda\kappa_j, \tag{14}$$

$$E(Z_j) = \nu + \lambda\kappa_j, \tag{15}$$

$$Cov(\varepsilon, \delta) = Cov(\varepsilon, W_j) = Cov(\delta, W_j) = 0, \tag{16}$$

$$Var(W_j) = \sigma_{wj}^2, \tag{17}$$

$$Var(\delta) = \Theta_\delta, \tag{18}$$

$$Var(\varepsilon) = \Theta_\varepsilon. \tag{19}$$

Equation 13 gives the factor structure for observed merit, and is similar to Equation 5. Equation 12 specifies the regression of salary on observed merit. The intercept α and the slope β are invariant across groups, as is the residual variance in Equation 18. As in the common-factor model, the mean κ and variance α_w^2 of true merit may vary across groups.

Under standard normality assumptions on W and (ε, δ), the model specified in Equations 12–19 can be shown to imply true fairness. Some additional constraints are needed to identify the model. In Equation 13, one loading may be fixed to a nonzero value, with the corresponding intercept fixed to zero. If $p = 1$, one further constraint is required. One choice in this case is to fix the variance of true merit in Equation 17 to a nonzero value in one group. For $p > 1$, and assuming two groups with the above identification constraints, the model will have $p^2 + p$ degrees of freedom.

The assumption of true fairness in this model requires invariance in all parameters in Equations 12, 13, 18, and 19. Note that group differences in the unique variances for the observed merit measures will violate true fairness by creating group differences in the conditional variance of salary given true merit. The expressions for the conditional mean and variance for salary, given true merit, are

$$E(Y_j | W_j) = \alpha + \beta'(\nu + \lambda W_j) \tag{20}$$

$$Var(Y_j | W_j) = \beta'\Theta_\varepsilon\beta + \Theta_\delta. \tag{21}$$

The invariance restrictions on any of these parameters may be eliminated to study the true fairness assumptions. If group differences in the conditional mean salary are to be introduced, the invariance restrictions on α, β, ν, *or* λ could be eliminated. Eliminating the restrictions on β will produce group differences in the conditional variance of salary, as will the removal of restrictions on the variances in Equation 21.

It is often useful to create a null model that can be used as a baseline for the computation of comparative fit indices (Bollen & Long, 1993). The inclusion of mean structures in the model produces a variety of choices for the null model (Millsap & Everson, 1991). One useful choice is a null model that requires all salary and merit measures to be mutually uncorrelated, with identical means and variances across groups. This null model is used in the analyses reported earlier. The invariance constraints are required because the path models to be studied impose invariance restrictions in some form. This null model is more restrictive than any of the path models to be considered. Less restrictive null models could be considered for particular applications.

Two other practical issues are important in applying the path models described here. First, when implementing the models in software programs such as LISREL (Jöreskog & Sörbom, 1989), the user may need to provide start values for the iterations. The automatic start values produced by the program are often poor for these models. Secondly, as in other applications of structural equation modeling, improper parameter estimates (e.g., negative variances) will be encountered when sample sizes are small or the proposed model fits poorly. Unless these improper estimates can be attributed solely to sampling error, the proposed model should be rejected in such cases.

References

Ali, M.M. & Silver, J.L. (1985). Tests for equality between sets of coefficients in two linear regressions under heteroscedasticity. *Journal of the American Statistical Association, 80:* 730-735.

Birnbaum, M.H. (1979). Procedures for the detection and correction of salary inequities. Pp. 121-144 in T.R. Pezzullo & B.E. Brittingham (Eds.), *Salary equity.* Lexington,MA: Lexington Books.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K.A. & Long, J.S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. Pp. 136-162 in K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models.* Newbury Park, CA: Sage.

Conway, D.A. & Roberts, H.V. (1983). Reverse regression, fairness, and employment discrimination. *Journal of Business and Economic Statistics, 1:* 75-85.

Dempster, A.P. (1988). Employment discrimination and statistical science. *Statistical Science, 3:* 149-185.

Draper, N. & Smith, H. (1981). *Applied regression analysis,* (2nd ed.) New York: Wiley.

Finkelstein, M.O. & Levin, B. (1990). *Statistics for lawyers.* New York: Springer-Verlag.

Gastwirth, J.L. (1988). *Statistical reasoning in law and public policy.* Boston: Academic Press.

Goldberger, A.S. (1984). Reverse regression and salary discrimination. *Journal of Human Resources, 19:* 293-318.

Gollob, H.F. (1984). Detecting sex bias in salaries. *American Psychologist, 39:* 448-451.

Gregory, K.L. (1991). *A reconsideration of bias in employment testing from the perspective of factorial invariance.* Unpublished doctoral dissertation, University of California, Berkeley.

Holland, P. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association, 81:* 945-970.

Jöreskog, K.G. & Sörbom, D. (1989). *LISREL VII user's reference guide.* Mooresville, IN: Scientific Software.

Levin, B. & Robbins, H. (1983). Urn models for regression analysis, with applications to employment discrimination studies. *Law and Contemporary Problems, 46:* 247-267.

Marascuilo, L.A. & Levin, J.R. (1983). *Multivariate statistics in the social sciences: A researcher's guide.* Monterey, CA: Brooks-Cole.

Mauro, R. (1990). Understanding L.O.V.E. (Left Out Variables Error): A method for estimating the effects of omitted variables. *Psychological Bulletin, 108:* 314-329.

McDonald, R.P. & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin, 107:* 247-255.

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

Meredith, W. & Millsap, R.E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57:* 289-311.

Millsap, R.E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research, 30:* 577-605.

Millsap, R.E. & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research, 26:* 479-497.

Millsap, R.E. & Meredith, W. (1994). Statistical evidence in salary discrimination studies: Nonparametric inferential conditions. *Multivariate Behavioral Research, 29:* 339-364.

Millsap, R.E. & Meredith, W. (1992). Inferential conditions in the statistical detection of measurement bias. *Applied Psychological Measurement, 16:* 389-402.

Peterson, D.W. (1986). Measurement error and regression analysis in employment cases. Pp. 107-131 in D.H. Kaye & M.H. Aicken (Eds.), *Statistical methods in discrimination litigation.* New York: Dekker.

Rosenbaum, P.R. (1995). *Observational studies.* New York: Springer-Verlag.

Schafer, D.W. (1987). Measurement-error diagnostics and the sex discrimination problem. *Journal of Business and Economic Statistics, 5:* 529-537.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25:* 173-180.