MEASUREMENT ERROR IN JOB EVALUATION
AND THE GENDER WAGE GAP

Shih-Neng Chen^a, Peter F. Orazem^b, J. Peter Mattila^b, and Jeffrey J. Greig^c

published in *Economic Inquiry* 37 (April 1999): 181-194

^aWorld College of Journalism and Communications, Taipei and Iowa State University.

^bIowa State University.

^cBurns and McDonnell.

Partial support from NSF grant number 8909479 is gratefully acknowledged. We received helpful comments from Yasuo Amemiya, Wayne Fuller, and Wally Huffman on earlier drafts of this paper. Donna Otto prepared the manuscript.

Abstract

Job evaluation is used to establish pay for more than half of the workers in the United States. It is also the key tool used in establishing the extent of pay bias in firm pay systems. However, job evaluations are subject to measurement error that can bias estimates of the magnitude of pay discrimination. A practical procedure for making these corrections is outlined. Using computed reliability ratios to adjust for measurement error in a study of state government jobs, we find that measurement errors exaggerate the implied extent of discrimination against predominantly female jobs by 34% to 44%. Measurement errors also exaggerate the number of independent job factors which affect pay.

Wage equations are widely used to estimate the size of the pay gap between male- and female-dominated jobs. When jobs are the unit of observation, job wage rates are regressed on job attributes (such as skill, effort, responsibility, and working conditions) and on the percent of female incumbents in the job. The coefficient on the percent female variable estimates the size of the gender pay gap, which is often taken as a measure of pay discrimination against women.¹ These estimates frequently become the metric for upward adjustment of female-dominated job pay to redress the discrimination. Summarizing the research which uses jobs as the unit of observation, Ehrenberg (1989, p. 93) concluded that female-dominated jobs are underpaid by 15% to 34% relative to male-dominated jobs. Similar findings have been reported in studies which use individuals instead of jobs as the unit of observation.

Many criticisms have been raised of these two types of studies. One common criticism is that not all relevant control variables for job attributes or individual productivity may be measured (or even measurable) so that specification error arises. If these omitted variables are correlated with the gender variable, then the estimated gender pay gap incorporates legitimate differences in productivity and/or job attributes. However, few studies have analyzed the impact of omitted factors bias empirically.

Another common criticism is that there is measurement error in the control variables which in turn leads to biases in estimating the size of the pay gap. Measured job attributes are highly subjective in nature, making measurement error a major problem. Even human capital measures of education and training are subject to error. Although one may easily measure years of *completed* education and experience, it is much more difficult to measure years of *effective* learning or achievement. Only a few studies have analyzed the size of these biases empirically.

This paper provides the first estimates of the size of this measurement error bias in the context of job level (job attribute) wage equations. To our knowledge, only two previous studies have estimated measurement error bias using individual level, human capital wage equations, and both are tangential to the major question that we address.

In particular, we find that the estimated gender pay gap is overstated by 34% to 44% when measurement error bias is ignored in job level wage equation analysis. Wage adjustments to correct the gender gap in pay which adhere strictly to such biased estimates may lead to millions of dollars in over-compensation. In the future, job evaluation should correct for this bias and we suggest appropriate procedures. In addition, we provide unique evidence that previous attempts to add measures of seemingly "omitted" job attributes need not have reduced the bias in the measured pay gap. We find that the number of job attributes which enter the wage regression is artificially increased because measurement errors reduce the measured collinearity among job factors.

The next section explains factor-point pay plans and summarizes the few related studies. Section II reviews the statistical theory of regression analysis when there is measurement error in the explanatory variables. We analyze the effect of independent and correlated measurement errors between job factors on the estimated magnitude of pay discrimination. Section III discusses the data and reports the empirical results from implementing various measurement error corrections. The final section summarizes our findings and suggests how this study could be used to improve future factor-point pay analysis.

I. BACKGROUND AND RELATED RESEARCH

More than half of the workers in the United States are paid according to a pay system at least partly based on job evaluation. These systems establish hierarchies of jobs and pay levels using information on a common set of job attributes such as skill, effort, responsibility, and working conditions. Most frequently, a factor-point system is used. For each job, points are assigned to each job factor. Higher levels of compensable factors receive higher points. Weights are assigned to each factor to reflect its relative importance to the employer. The sum of the weighted job factor points creates a ranking of jobs by which relative pay is set.² Such plans are widely used in both the private and public sector.

A comparable worth pay plan is a factor-point system which sets pay according to the relative value of jobs to the firm without explicitly incorporating market wage information. Market wages are ignored because comparable worth advocates view them as distorted by gender discrimination. Since factor-point systems claim to establish the relative value of jobs without the direct influence of potentially discriminatory market wages, they have become the leading mechanisms for implementing comparable worth pay plans. A set of job factors (attributes) are defined, each having a set of levels and factor points. Teams of evaluators are given the task of assigning points to each job for each factor. Typically, each team is given information about the jobs (including job descriptions) and asked to reach an agreement on assigned points in a process that is inherently very subjective. Sometimes only one team evaluates each job. However, if two or more teams independently evaluate a set of jobs, then measurement error bias may be evaluated and corrected as emphasized in this paper.

After all jobs have been evaluated and factor points have been assigned, wage equations and the size of the gender pay gap can be estimated. The typical procedure is to regress job pay (y) on a vector of job factors (X) and the percentage of female incumbents (f) in a job classification. The regression is of the form $y = X\beta + f\gamma + v$.³ A finding that γ is negative is interpreted to imply that the firm's pay structure undercompensates predominantly female jobs relative to male dominated jobs. If $\gamma = 0$, then pay is determined entirely by the vector of job

attributes (X). Therefore, if $\gamma = 0$, a gender neutral pay structure can be created by setting pay equal to $X\beta$ or, equivalently, setting pay equal to $[y - f\gamma]$. For example, such a policy would leave pay for 100 percent male jobs unchanged but would raise pay for 100 percent female jobs by -100γ .⁴

The consensus of numerous studies is that, at least in the public sector where most research has focused, $\gamma < 0$ so that female-dominated jobs are underpaid.⁵ Although pay has seldom been adjusted by the full amount ($-f\gamma$) in practice due to compromises made as part of the associated political process, the adjustment costs can be substantial.⁶ Hence it is important that the estimates be as accurate as possible. As a consequence, the possibility of measurement error and specification error raise create major concern.

The problems of measurement error are well-recognized in the literature, with several papers questioning the validity and reliability of job factor measurements.⁷ Nevertheless, England (1992, p.213) argued that presumed random errors "cannot explain the systematic tendency for job evaluation studies to find that female jobs have lower pay lines." However, such views are necessarily speculative because no studies have explicitly corrected for the measurement error problem in the context of job evaluation or comparable worth. As a consequence, the magnitude of the bias has not been established.

A few studies have examined the effect of measurement error on measured discrimination in other contexts. In an early paper, Hashimoto and Kochin (1980) pointed out that human capital model estimates of race discrimination are biased if, as is likely, the reported levels of non-white schooling are greater than the actual levels of attained knowledge. They illustrated the problem by reestimating a human capital based earnings function using a grouped data. They found that measurement error overstates the size of racial pay discrimination. However, they caution that the magnitude of the overstatement is uncertain due to data limitations.

More recently, Rapaport (1995) concluded that measurement error in survey responses concerning years of work experience may explain most of the pay gap between male and female teachers in two California school districts. She uses survey responses concerning educational attainment and work experience to predict the level of actual earnings of individual teachers paid according to a standard pay plan. Although there should be little room for sex or racial discrimination in such "impersonal" pay plans, conventionally estimated wage equations imply that females and blacks are paid less. Through a process of elimination, Rapaport concludes that this finding is likely due to measurement error, although she can't rule out the possibility of discrimination. Unlike Rapaport's study, we explicitly observe and control for measurement error in the control variables. Our study also differs in that we use job level data and control for

job evaluation factors, which are likely to be more prone to measurement error than is education and experience. Like Rapaport, we show that the potential impact of measurement error can be substantial.

Disagreement on the methodology of pay analysis has not been restricted to issues of measurement error. Job evaluation analysts have also disagreed about which or how many job factors should be included. The presumption has been that adding some control for potentially important job attributes, even if subject to measurement error, will yield better estimates of γ than would result from ignoring these job attributes.⁸ Our paper shows that adding noisy estimates of job factors may increase rather than decrease the bias in γ .

II. SUMMARY OF STATISTICAL THEORY

A. Measurement Error Biases the Coefficient on Percent Female

To illustrate how measurement error in job factors can affect the coefficient on percent female incumbents, consider the linear regression model⁹:

$$(1) \quad \begin{aligned} y_j &= \beta_0 + \beta_1 x_{1j} + \gamma x_{2j} + e_j & j = 1, 2, \dots, n \\ X_{1j} &= x_{1j} + u_{1j} & j = 1, 2, \dots, n \end{aligned}$$

where β_0 , β_1 and γ designate true parameters and e_j is a random error with mean zero and variance σ_e . Lower-case variables represent true variables and upper-case variables are observed values. In equation (1), X_{1j} measures x_{1j} with error u_{1j} , where $u_{1j} \sim NI(0, \sigma_u)$. The variables x_{2j} and y_j are measured without error.¹⁰

Assuming x_{1j} , u_{1j} , and e_j are mutually uncorrelated for $i = 1, 2$, model (1) can be rewritten as

$$(2) \quad \begin{aligned} y_j &= \beta_0 + \beta_1(X_{1j} - u_{1j}) + \gamma x_{2j} + e_j & j = 1, 2, \dots, n \\ &= \beta_0 + \beta_1 X_{1j} + \gamma x_{2j} + v_j \end{aligned}$$

where $v_j = e_j - \beta_1 u_{1j}$. Note that in model (2), X_{1j} and v_j are not independent. The ordinary least squares (OLS) estimator is $(X'X)^{-1}(X'y)$, where $X = (\mathbf{1}, X_1, x_2)$, and y , $\mathbf{1}$, X_1 , and x_2 are all $n \times 1$ vectors.

The least squares estimator is inconsistent. It can be shown¹¹ that in the limit, the bias approaches

$$(3) \quad \text{plim} \begin{bmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \\ \hat{\gamma} - \gamma \end{bmatrix} = - \begin{bmatrix} (\sum X_{1j}/n) \beta_1 \sigma_{u1} \\ (\sum X_{1j}^2/n) \beta_1 \sigma_{u1} \\ (\sum X_{1j}x_{2j}/n) \beta_1 \sigma_{u1} \end{bmatrix}$$

where σ_{u1} is the measurement error variance for X_{1j} . From (3) one can see that, in general, every parameter estimated by OLS will be biased, whether or not the variable is estimated with error. In particular, the regression coefficient on the explanatory variable which is measured without error, $\hat{\gamma}$, is also biased by the presence of measurement error in X_{1j} . Because $\Sigma X_{1j}^2 > 0$, and $\sigma_{u1} > 0$, $\hat{\gamma}$ is biased toward zero if $\beta > 0$. The direction of bias for $\hat{\gamma}$ depends on the sign of $\Sigma X_{1j} x_{2j}$ and cannot be established in general.

The implications of these results for pay analysis are clear. Random measurement error in a job factor (i.e. x_{1j}) will bias the coefficient on percent female incumbents (i.e. x_{2j}), provided female incumbency and the measured job characteristic are correlated ($\Sigma X_{1j} x_{2j} \neq 0$), and provided the true job characteristic affects pay ($\beta_1 \neq 0$). The bias exists, even though female incumbency is observable without error. However, even in this simplest of cases with only one job factor, the direction of bias for γ is ambiguous.

B. Measurement Error in Job Factors Reduces Correlation Between Factors

The true population squared correlation between x_{1j} and x_{2j} is defined by

$$(R_{x1x2})^2 = (\sum x1x2)^2 / (\sum x1^2 \sum x2^2)$$

while the observed population squared correlation between X_{1t} and X_{2t} is

$$(R_{X1X2})^2 = (\sum X1X2)^2 / (\sum X1^2 \sum X2^2) = x1 [(\sum x1x2)^2 / (\sum x1^2 \sum x2^2)] = x1 (R_{x1x2})^2$$

where $x1 = (\sum x1^2 / \sum X1^2)$ is the reliability ratio for X_1 . Because $0 < x1 < 1$, measurement error causes the observed correlation, R_{X1X2} , to be less than the true correlation R_{x1x2} . Similarly, if both explanatory variables are measured with mutually independent random errors, the observed population squared correlation between X_{1j} and X_{2j} is

$$(4) \quad (R_{X1X2})^2 = (\sum X1X2)^2 / (\sum X1^2 \sum X2^2) = x1 x2 [(\sum x1x2)^2 / (\sum x1^2 \sum x2^2)] \\ = x1 x2 (R_{x1x2})^2$$

The existence of random measurement error means that observed correlation among factors is less than the true correlation among factors. This can lead to more job factors being used in the pay analysis than would be possible were the job factors measured without error.

For example, suppose that only one job factor entered the true model of wages, but there are two measures of this job factor. If the two measures were perfectly correlated, only one could be used in the regression. Inclusion of both would lead to a singular moment matrix. However, if both measures included random errors, the two factors would no longer be perfectly correlated and both could be used as regressors. The measurement error provides an artificial identification of the ordinary least squares coefficients, and artificially increases the dimensionality of the empirical model from one to two.¹²

C. Generalization to Many Job Factors Measured with Error

Let us now consider the specific linear regression model used in this paper.

$$(5) \quad y_j = \alpha_0 + \alpha_1 x_{1j} + \alpha_2 x_{2j} + \dots + \alpha_k x_{kj} + f_j + e_j \quad j = 1, 2, \dots, n$$

$$(6) \quad X_{ij} = x_{ij} + u_{ij} \quad i = 1, 2, \dots, k \quad j = 1, 2, \dots, n$$

where y_j is pay in j^{th} job classification (measured without error); X_{ij} is "measured" i^{th} evaluation factor in the j^{th} job classification; x_{ij} is "true" i^{th} evaluation factor in the j^{th} job classification; u_{ij} is measurement error with $u_{ij} \sim \text{NI}(0, \sigma_{ui})$; f_j is the percentage of female incumbents in the j^{th} job classification (measured without error); and e_j is a random disturbance term with $e_j \sim (0, \sigma_e)$.

Combining equations (5) and (6), one obtains

$$(7) \quad y_j = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 X_{2j} + \dots + \alpha_k X_{kj} + f_j + v_j \quad j = 1, 2, \dots, n$$

$$= W_j' \alpha + v_j$$

where $W_j = (1, X_{1j}, X_{2j}, \dots, X_{kj}, f_j)$

$$= (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k, f_j)'$$

$$v_j = e_j - \alpha_1 u_{1j} - \alpha_2 u_{2j} - \dots - \alpha_k u_{kj}.$$

Under the assumption that x_{ij} , u_{ij} , f_j , and e_j are mutually uncorrelated for all i and j , it can be shown that for large samples, the inconsistency from applying OLS to regression (7) is

$$(8) \quad \text{plim}(\hat{\delta}_{\text{OLS}} - \alpha) = \text{plim}[(W'W)^{-1} n^{-1} W'(e - \alpha_1 u_1 - \alpha_2 u_2 - \dots - \alpha_k u_k)]$$

$$= \text{plim}[(W'W)/n]^{-1} [0, -\alpha_1 \sigma_{u1}, -\alpha_2 \sigma_{u2}, \dots, -\alpha_k \sigma_{uk}, 0]'$$

where $\hat{\delta}_{\text{OLS}} = (W'W)^{-1} W'y$.

As before, the regression coefficient for f_j , $\hat{\alpha}_k$, is inconsistent due to the presence of measurement error of the X 's despite the fact that f_j is measured without error. The only difference between (8) and (3) is that the direction of bias cannot be established for any of the parameters in (8).

D. Generalization to Correlated Measurement Errors

All previous results are based upon the assumption of independent measurement errors. However, the measurement errors associated with evaluating job factors may be correlated. An individual evaluator may consistently give low scores or high scores to all job factors. Or evaluators may not maintain independence in their scoring, creating common errors across evaluators. Correlated measurement errors complicates the analysis considerably. Assume the previous general case in equations (5) and (6) but now let measurement errors associated with the job evaluation factors be positively correlated. The correlation can be expressed as follows:

$$(9) \quad X_{ij} = x_{ij} + u_{ij} \quad j = 1, 2, \dots, n$$

$$X_{kj} = x_{kj} + u_{kj} \quad j = 1, 2, \dots, n$$

where X_{ij} is measured job evaluation factor i ; u_{ij} is measurement error in the i^{th} job evaluation factor; X_{kj} is measured job evaluation factor k ; u_{kj} is measurement error in the k^{th} job evaluation factor; and x_{ij} , u_{kj} are uncorrelated for all i and j . Error terms u_{ij} and u_{kj} are correlated with covariance σ_{uiuk} . The observed population squared correlation between X_i and X_j under model (9) is:

$$\begin{aligned}
 (10) \quad (R_{XiXk})^2 &= (\bar{x}_{iXj})^2 / (\bar{x}_i \bar{x}_k) \\
 &= (\bar{x}_{ixk} + \bar{u}_{iuk})^2 / [(\bar{x}_i \bar{x}_k)] \\
 &= \bar{x}_i \bar{x}_k \{[(\bar{x}_{ixk})^2 + 2 \bar{x}_{ixk} \bar{u}_{iuk} + (\bar{u}_{iuk})^2] / (\bar{x}_i \bar{x}_k)\} \\
 &= \bar{x}_i \bar{x}_k (R_{xixk})^2 + \{[2 \bar{x}_{ixk} A + A^2] / \bar{x}_i \bar{x}_k\}
 \end{aligned}$$

where $A = \rho_{uiuk} \sqrt{(1-\kappa_{xi})(1-\kappa_{xk})} \sigma_{xi} \sigma_{xk}$, and ρ_{uiuk} is the correlation in measurement error between x_i and x_k . If the measurement errors are uncorrelated so that $\rho_{uiuk} = 0$ in (10), the second expression on the right-hand-side vanishes and the equation collapses to the form in equation (4). However, if the measurement errors are correlated, then the second term on the right-hand-side of (10) is not zero. The second term in (10) can be positive or negative, so the true correlations can be larger or smaller than the observed correlations.

III. DATA AND ESTIMATION PROCEDURES

A. Data and Empirical Methodology

In 1984, Arthur Young Company conducted a factor point pay analysis of 758 jobs in the State of Iowa Merit Pay System. Arthur Young trained nine teams of state employees to conduct the evaluations. Each team was composed of two men and two women with one personnel specialist on each team. Most jobs were evaluated by only one team, but two or more teams evaluated 90 of the 758 jobs. The multiple observations for specific jobs allowed Arthur Young to compute inter-rater reliability ratios for each of the 13 job factors.

To show how these reliability ratios are computed in practice, suppose we have raters, each of whom measure the i^{th} factor for J different jobs. All raters are unbiased by assumption.¹³ The reliability ratio for factor i can be estimated as the R^2 from a regression of the form:

$$X_{ijm} = \sum_{j=1}^J \bar{x}_{ij} D_j + u_{ijm}$$

$\bar{x}_{ij}$ is the average estimate across raters for the i^{th} factor for the j^{th} job, estimated as the coefficient on the j^{th} of J dummy variables for the job being evaluated. The m^{th} rater's deviation from the mean estimate for factor i and job j is u_{ijm} . Because of the unbiasedness assumption, $E(u_{ijm}) = 0$. If there is no uncertainty among the raters about the measure of factor i across the J

jobs, $u_{ijm} = 0 \forall j$ and m , so that the R^2 (and the reliability ratio) will equal one. More typically, there will be disagreement among the raters so that $u_{ijm} \neq 0 \forall j$ and m , and the R^2 (reliability ratio) will be between zero and one. These reliability ratios can be treated as estimates of the κ_{xi} in equations (4) and (10). The sample statistics and reliability ratios for the 13 job factors are reported in Table 1.

Reliability ratios provide the information needed to extract true correlation coefficients from observed data. Using (4) and the independence assumption, the true correlations $(R_{x_1x_2})^2$ can be extracted by dividing observed population correlation coefficients by the appropriate reliability ratios. With correlated measurement errors, the true correlations can be extracted using (10), the reliability ratios, and information on correlation coefficients between measurement errors.

Reliability ratios can also be used to extract the true coefficients by correcting the covariance matrix for the presence of measurement error. The estimator is

$$(11) \quad \begin{aligned} \beta &= (X'X - D'\Lambda D)^{-1} X'y \\ D &= \text{diag}(\sigma_{x1}, \sigma_{x2}, \dots, \sigma_{xn}) \\ \Lambda &= \text{diag}(1-\kappa_{x1}, 1-\kappa_{x2}, \dots, 1-\kappa_{xn}) \end{aligned}$$

for the uncorrelated measurement error case. If measurement errors are correlated, the Λ matrix has off-diagonal elements of the form $\rho_{uiuk} (\sigma_{ui} \sigma_{uk} / \sigma_{xi} \sigma_{xk})^{1/2}$.¹⁴ A computer program (EVCARP) developed by Schnell, Park and Fuller (1987) was used to estimate the regression equations reported in this study.¹⁵

B. Estimates

The original regression analysis conducted by Arthur Young Company (1984) regressed pay grade and maximum salary for 758 jobs on the thirteen job factors and a variable measuring percent female incumbents. The analysis herein concentrates on predicting maximum salary since pay grade is difficult to interpret.¹⁶ Use of maximum biweekly pay also has the advantage of holding constant incumbent step level across jobs. Variation in pay across jobs is therefore due solely to differences in how the job is rewarded by the pay system and not by differences in incumbent job tenure across jobs. The ordinary least squares regression is reported in the first column of Table 2.

The functional form uses level of salary as the dependent variable. A Box-Cox regression strongly rejected the log specification of the dependent variable.¹⁷ The results appear quite reasonable on the surface. The regressors explain 92 percent of the variation in maximum salary. Most job factors raise pay, eight of them significantly. Most importantly for this study, the coefficient on percent female is negative and significant. A ten point increase in percentage female incumbents reduces biweekly pay by \$7.19.

The estimates correcting for reliability of job factors (column 2 labeled “EVCARP” of Table 2) are very different. The standard errors explode so that all coefficients become insignificant. In addition, the test of singularity of the moment matrix easily fails to reject the null hypothesis of singularity.¹⁸ Computation of the true correlation matrix using equation (4) reveals why. Several of the true job factors are perfectly correlated, as shown in Table 3. The only reason these variables had estimable coefficients in the first column of Table 2 was that they were measured with error and the measurement error was sufficient to create independent variation in the measured factors. Measurement errors in the factors increased the apparent dimensionality of the job factor space.

The five factors which are perfectly or nearly perfectly correlated are difficult to distinguish on a conceptual as well as an empirical basis. The factors include required knowledge from experience (KFE), a job’s complexity, judgment and problem-solving (JPS); supervision required on the job (GSA); the scope and effect of the job on the institution (SE); and the impact of errors on the institution (IE). All these factors deal with job requirements and skills that come with increased experience and responsibility.

The five collinear factors were combined into a single factor (JOINT) by taking the simple average of the five. Because the combined factor is a linear combinations of the original five factors, the reliability ratio for the combined factor can be constructed from the five individual reliability ratios.¹⁹

The ordinary least squares regression using JOINT in place of the five factors is reported in column three of Table 2. The linear restrictions implied by using the combined factor in place of the five separate factors was tested and could not be rejected at standard significance levels. Most importantly, the coefficient on percent female incumbents is nearly the same as before (- .74 versus -.72) so that our use of the combined factor does not of itself affect the outcomes. Therefore, all the remaining regressions use JOINT in place of the five highly correlated factors.

The reliability corrected regressions are reported in Table 4. For comparison, the uncorrected regression is reported in the first column. The next four columns contain regressions which correct for measurement errors under progressively higher assumed error correlations across factors. In each of these columns, ρ_{uiuk} is assumed to equal a constant ρ for all job factors, i and k . Four specifications are reported, $\rho = 0$, $\rho = .1$, $\rho = .2$ and $\rho = .3$. The specifications with $\rho > .1$ create moment matrices that approach singularity. Therefore, the discussion will concentrate on the specifications with $\rho < .2$.

Correcting for measurement error in the factors sharply decreases the impact of the percent female variable, although the coefficient remains significant. When we assume uncorrelated errors, this reduces the absolute value of the coefficient by 34 percent of the OLS

estimate. Allowing measurement error correlation across factors further reduces the impact of percent female incumbents. With $\rho = .1$, the coefficient is reduced by 44 percent relative to the OLS coefficient. Clearly, the extent of implied pay bias against predominantly female jobs is sensitive to measurement error in job factors.

One way to correct the pay structure for bias against predominantly female jobs is to increase pay per job by $-\gamma f$ (from equation 7). Because the coefficient (γ) on percent female incumbents (f) is negative, this implies raising the pay for jobs having female incumbents. At sample means of percent female incumbents, the OLS coefficient implies an average increase of \$645 per year. The measurement error corrected regression with $\rho = .1$ implies an average increase of only \$359. For 100% female jobs, the dollar difference associated with measurement error is even larger: \$1,924 versus \$1,071.

The coefficients on job factors are also sensitive to measurement error corrections. The coefficient on "JOINT", which is based on indicators of experience and responsibility, increases as measurement error corrections are introduced. Mental and visual demands, an indicator of coordination required on the job, also increases in importance. Supervisory responsibility and personal contacts lose importance.

Table 5 provides additional insights on the impact of measurement error in the job factors on the outcomes of comparable worth pay analysis. For each worker in Iowa state government, we predicted pay using the indicated regression equation from Table 2 or 4, setting the percent female variable (PF) equal to zero. We then grouped workers into various occupational, educational, and market wage rate level categories and computed the average group biweekly state government pay as reported in Table 5.²⁰ The first two columns represent two baseline sets of pay structures (one using 13 job factors and the other using 8 factors plus "JOINT") where no adjustment is made using the reliability coefficients. The average wages vary little between these two baseline pay structures.²¹

The other three columns of Table 5 report the pay distributions using wage equations (8 factors plus "JOINT") corrected for measurement error. Correcting for measurement error makes little difference for highly educated and highly paid jobs. However, it does make a large difference for the less educated and lower paid jobs. Measurement corrected equations imply that lower biweekly pay is appropriate for these jobs than would be implied by the OLS estimates. Put another way, lower skilled and lower paid workers would see smaller upward comparable worth pay adjustments using measurement error adjusted wage equations.

The magnitude of these adjustments is not trivial. For those with less than a high school education, the measurement error corrected biweekly pay was \$24.20 to \$37.40 lower than the uncorrected pay, a difference of \$629 to \$972 per year. Aggregating over approximately 20,000

state employees in Iowa, the measurement error corrected pay would have resulted in \$4.2 million to \$5.6 million lower adjustment in pay per year than the uncorrected pay based on the OLS estimates.

IV. CONCLUSIONS AND IMPLICATIONS

This study evaluates the impacts of measurement errors on “policy capturing” regression coefficient weights using data from the State of Iowa's comparable worth system. Corrections for measurement error and multicollinearity in the original Arthur Young's job evaluation factors are used to examine the sensitivity and statistical robustness of these factor weights. The empirical findings can be summarized as follows:

1. The presence of measurement error caused upward bias on the absolute value of the estimated coefficient on percent female incumbents in Iowa. The adverse impact of percent female on pay is reduced by thirty-four to forty-four percent when the problem of measurement error is corrected. If this coefficient is taken to be a measure of discrimination against predominantly female jobs, then measurement errors caused the implied discrimination to be overstated. As a consequence, proposed comparable worth wage adjustments necessary to bring female jobs to parity with male jobs in Iowa were too large. While it would be inappropriate to generalize from one study, measurement errors also led to greater implied discrimination in the Hashimoto and Kochin (1980) and Rapaport (1995) studies.
2. Measurement error reduces the collinearity among job factors, allowing too many factors to be included in the pay analysis. Using reliability ratios, the "true" correlation matrix was estimated. Five factors out of thirteen were found to be perfectly or nearly perfectly correlated once we correct for measurement error. The OLS coefficients for these five factors were only identified because of measurement error.
3. The magnitude of the bias in the Iowa case suggests that other studies of pay discrimination which assume no measurement error may also be seriously flawed. However, the current study also demonstrates that measurement error correction models are tractable in pay analysis. It is straightforward for consultants to estimate reliability ratios for factors by conducting more than one independent evaluation per job. Measurement error variances, covariances and reliability coefficients can be estimated and applied to obtain unbiased estimates of factor weights using equation (11).

Endnotes

1. See Treiman and Hartmann (1981).
2. Schwab (1984) estimated the share of the labor force covered by job evaluation. England (1992, Chapter 4) has a good review of job evaluation methods and related research.
3. It is common for analysts to set the coefficients, β , on the basis of a priori presumed relative importance of the various job factors to the firm. In essence, this means establishing the β 's without appealing to statistical analysis. The Hay system and its variants are examples of a priori weighted systems in which the consultant reports an aggregate job index, $I = \hat{X}\beta$, where β is the consultants weights on the vector of job factors. Nevertheless, one can still regress y on I and f to get an estimate of γ .
4. We should note that the pay equation does not require that the firm literally set pay according to the pay equation. The equation is often interpreted as "policy capturing," meaning that the β and γ coefficients capture the implicit returns to job attributes and female incumbency in the firm. In fact, given federal and state legislation, it is unlikely that there is ever an explicit policy to underpay predominantly female jobs, so the pay equation is unlikely to be a literal statement of firm official pay policy. The analogous case is that the standard earnings function does not require a conscious policy that wages increase with experience and education, but the earnings function captures market returns which demonstrate those tendencies.
5. In addition to the Ehrenberg (1989) review cited earlier, Sorenson (1987) provides estimates that strict adherence to job evaluation would raise female pay an average of 11 percent. Ames (1995) found that job evaluation raised female pay in Ontario and Manitoba. England (1992, p. 205) concluded that job evaluation, "nearly always gives women's jobs higher wages relative to men's ..."
6. See Ames (1995) or Orazem and Mattila (1990) for studies of how the implementation process may divert the results away from the proposed gender-neutral pay structures.
7. See Greig et al (1989), Madigan (1985), Collins and Muchinsky (1993) or Schwab (1985) for discussions of these issues.
8. Examples include Lucas (1977), Filer (1989) and Schuman, Ahlburg and Mahoney (1994).
9. Judge et al. (1985), Chapter 13 contains an introduction to measurement error problems. See Fuller (1987) for a detailed discussion.
10. In the context here, job factors are not observable but are estimated by evaluators. If evaluators are unbiased, the mean evaluation of the same job factor across evaluators will

be a consistent estimator of x_{ij} , and deviations from the mean can be used to estimate the measurement error variance, $\frac{\sum (X_{ij} - x_{ij})^2}{n-1}$.

11. Derivation provided on request.
12. A likely example of this phenomenon is Filer's (1989) study which includes 225 regressors in a model explaining variation in average occupational wages.
13. Unbiased raters are a maintained assumption in job analysis because if one assumed the raters were biased, the job analysis would not be undertaken in the first place. To reduce potential bias, raters are sent through a training program to sensitize them to possible sources of bias and to disqualify bad raters. In addition, teams are chosen to represent diverse points of view to reduce the possibility of strategic ratings.
14. Note that when $i = k$, $\rho_{uiuk}(\sigma_{uiuk}/\sigma_{xi}\sigma_{xk})^{1/2} = \frac{\sigma_{ui}}{\sigma_{xi}} = 1 - \kappa_{xi}$, the same diagonal elements of Λ as in (11). If there is no measurement error, all reliability ratios equal one, and all error variances are zero. In that case, Λ becomes a null matrix, so that $\beta = \beta_{OLS}$.
15. The program, EVCARP, has a routine designed to correct for measurement error, given information on reliability ratios.
16. Chen (1995) reports results using pay grade as the dependent variable. Qualitative results are similar to those reported herein.
17. The Box-Cox regression (Judge et al., pp. 308-311) finds the value of λ which best transforms y into a normally distributed variable. The transformation $(y^\lambda - 1)/\lambda$ approaches $\ln(y)$ as $\lambda \rightarrow 0$ and approaches $y - 1$ as $\lambda \rightarrow 1$. The maximum likelihood estimate of λ was .78, so the linear specification was preferred. Logarithmic specifications yielded similar qualitative results. See Chen (1995) for details.
18. Fuller (1987, p. 201) shows that the test statistic for the null hypothesis that the rank of the true moment matrix is $k - 1$ when there are k regressors is $\left(\frac{n}{n - k + 1}\right) \hat{F}$ where $\hat{F}$ is the smallest characteristic root of the matrix $X'X - D'\Lambda D$ in (12). The test statistic is distributed $F(n - k + 1, n - 1)$.
19. Derivation available on request.
20. State jobs were matched to market jobs using job titles and job qualification requirements. Details on the market wage data are contained in Orazem and Mattila

(1989).

21. While the magnitudes of the differences in means between OLS13 and OLS9 are small, most were statistically significant. The larger differences between the EVCARP predicted salaries and predicted salaries using OLS9 are also generally significant.

REFERENCES

- Ames, Lynda J. (1995). "Fixing Women's Wages: The Effectiveness of Comparable Worth Policies." Industrial and Labor Relations Review 48(4): 709-725.
- Arthur Young Company. (1984). Study to Establish an Evaluation System for State of Iowa Merit Employment System Classifications on the Basis of Comparable Worth: Final Report and Statistical Supplement. Milwaukee: Arthur Young Company.
- Chen, S. (1995). Two Applications of Measurement Error Correction in the Economics of Human Resources. Unpublished Ph.D. dissertation. Ames: Iowa State University.
- Collins, Judith M. and Paul M. Muchinsky. (1993). "An Assessment of the Construct Validity of Three Job Evaluation Methods: A Field Experiment." Academy of Management Journal 36(4): 895-904.
- Ehrenberg, Ronald. (1989). "Empirical Consequences of Comparable Worth," in Comparable Worth: Analysis and Evidence, ed. Hill, M. Anne and Mark R. Killingsworth, 90-106. Cornell, NY: ILR Press.
- England, Paula. (1992). Comparable Worth: Theories and Evidence. New York: Aldine de Gruyter.
- Filer, Randall K. (1989). "Occupational Segregation, Compensating Differentials, and Comparable Worth," in Pay Equity: Empirical Inquiries, ed. Michael, R. T., Hartmann, H. I., and O'Farrell, B., 153-170. Washington, DC: National Academy Press.
- Fuller, W. A. (1987). Measurement Error Models. New York: John Wiley & Sons, Inc.
- Greig, J. J., Orazem, P. F., and Mattila, J. P. (1989). "Measurement Errors in Comparable Worth Pay Analysis: Causes, Consequences, and Corrections." Journal of Social Issues, 45: 135-151.

- Hashimoto, Masanori and Levis Kochin. (1980). "A Bias in the Statistical Estimation of the Effects of Discrimination," Economic Inquiry 18(3): 478-486.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T. C. (1985). The Theory and Practice of Econometrics. 2nd ed. New York : John Wiley & Sons, Inc.
- Lucas, Robert E. B. (1977). "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," American Economic Review 67(4): 549-557.
- Madigan, Robert M. (1985). "Comparable Worth Judgments: A Measure Properties Analysis." Journal of Applied Psychology 70(1): 137-147.
- Mount, Michael M. and Rebecca A. Ellis. (1989). "Sources of Bias in Job Evaluation: A Review and Critique of Research." Journal of Social Issues 45(4): 153-168.
- Orazem, P. F., and Mattila, J. P. (1989). "Comparable worth and the Structure of Earnings: The Iowa Case," in Pay Equity: Empirical Inquiries, ed. Michael, R. T., Hartmann, H. I., and O'Farrell, B., 179-199. Washington, DC: National Academy Press.
- _____. (1990). "The Implementation Process of Comparable Worth: Winners and Losers." Journal of Political Economy 98(1): 134-152.
- Rapaport, Carol. (1995). "Apparent Wage Discrimination when Wages are Determined by Nondiscriminatory Contracts." American Economic Review 85(5): 1263-1277.
- Schnell, D., Park, H. J., and Fuller, W. A. (1987). EV CARP. Ames: Iowa State University, Statistical Laboratory.
- Schuman, Paul L., Dennis A. Ahlburg, and Christine Brown Mahoney. (1994). "The Effects of Human Capital and Job Characteristics on Pay." Journal of Human Resources 29(2): 481-503.
- Schwab, Donald P. (1985). "Using Job Evaluation to Obtain Pay Equity." in Comparable Worth: Issues for the 80's: A Consultation of the U.S. Commission on Civil Rights. 83-92.

Sorenson, Elaine. (1987). "Effect of Comparable Worth Policies on Earnings." Industrial Relations 26(3): 227-239.

Treiman, Donald J., and Heidi I. Hartmann, eds. (1981). Women, Work and Wages: Equal Pay for Jobs of Equal Value. Washington, DC: National Academy Press.

Table 1. Variable definitions and sample statistics.

Variable	Definition	Mean	Std Dev	Reliability Ratio
MS	Maximum salary (Biweekly)	700.68	200.01	1.00
JOB FACTORS				
KFFTE	Knowledge from education	49.62	34.90	0.92
KFE	Knowledge from experience	36.74	21.30	0.75
CJPS	Complexity, judgement and problem-solving	23.90	17.77	0.85
GSA	Guidelines and supervision available	17.85	9.22	0.73
SE	Scope and effect	32.49	17.99	0.73
IE	Impact of errors	18.41	9.40	0.74
PC	Personal contacts	47.38	17.89	0.78
PD	Physical demands	17.76	9.78	0.84
MVD	Mental or visual demands	12.62	4.79	0.55
SEN	Supervision exercised	10.09	12.85	0.92
WE	Working environment	17.16	7.43	0.71
UH	Unavoidable hazards	9.51	5.99	0.86
WI	Work pace and interruptions	30.99	7.76	0.55
JOINT	Average of KFE, CJPS, GSA, SE and IE	25.88	13.49	0.94
PF	Percentage of female incumbents	33.52	38.40	1.00

Source of all data is Arthur Young (1984) except for the market wage which was compiled by Orazem and Mattila (1989).

Table 2. Regression analysis of Iowa State maximum biweekly pay with and without correction for measurement error.

JOB FACTOR	OLS	EVCARP	OLS
INTERCEPT	307 (17)	330 (237)	311 (16)
KFFTE	1.80*** (0.09)	5.1 (762)	1.78*** (0.08)
KFE	2.16*** (0.17)	11.0 (196)	
CJPS	1.80*** (0.24)	8.0 (195)	
GSA	2.26*** (0.49)	-80.0 (225)	
SE	1.43*** (0.24)	50.0 (115)	
IE	2.64*** (0.41)	31.0 (879)	
PC	.629*** (.173)	1.7 (65.3)	5.63*** (.169)
PD	0.071 (.316)	2.5 (82.8)	0.012 (.313)
MVD	1.68*** (0.46)	10.0 (19.9)	1.71*** (0.46)
SEN	-.225 (.222)	1.4 (11.3)	-.203 (.220)
WE	.383 (.407)	3.7 (68.0)	.378 (.408)
UH	-.07 (.425)	8.0 (20.4)	.118 (.410)
WI	.451 (.331)	0.6 (18.6)	.531 (.330)
PF	-.719*** (.062)	-0.88 (9.08)	-.740*** (0.061)
JOINT			9.87*** (0.29)
SINGULARITY F-TEST		.058	
R ²	0.92	0.98	0.92

Standard errors in parentheses. (***, **, *, significant at 1%, 5%, 10% level)

Table 3. Observed and estimated true correlation coefficients for a subset of factors.

	CJPS	GSA	SE	IE
KFE	.67 [.84]	.73 [.98]	.71 [.95]	.68 [.91]
CJPS		.81 [1.00]	.78 [.99]	.74 [.93]
GSA			.83 [1.00]	.74 [1.00]
SE				.78 [1.00]

Estimated true correlation coefficients using (5) are in brackets.

Table 4. Regression analysis of Iowa State maximum biweekly pay under different measurement error correlation assumptions.

JOB FACTOR	OLS	EVCARP $\rho_{uij} = 0$	EVCARP $\rho_{uij} = 0.1$	EVCARP $\rho_{uij} = 0.2$	EVCARP $\rho_{uij} = 0.3$
INTERCEPT	311 (16)	280*** (55)	12.7 (80.3)	-39.5 (68.1)	-51.6 (73.2)
KFFTE	1.78*** (0.08)	1.83*** (0.13)	1.91*** (0.15)	1.99*** (0.14)	2.01*** (0.14)
PC	.563*** (.169)	-.256 (.604)	.908 (.647)	1.43*** (0.46)	1.56*** (.043)
PD	.012 (.313)	-1.28 (1.75)	1.74 (1.32)	2.60*** (0.78)	2.78*** (0.74)
MVD	1.71*** (0.46)	2.97** (1.41)	8.10*** (1.75)	8.63*** (1.41)	8.61*** (1.50)
SEN	-.203 (.220)	-1.14*** (0.43)	-1.41*** (0.50)	-.969** (.433)	-.771* (.411)
WE	.378 (.408)	4.00 (3.05)	3.22* (1.84)	2.69*** (0.96)	2.68*** (0.83)
UH	.118 (.410)	-0.65 (1.26)	-.108 (.930)	.298 (.722)	.470 (.711)
WI	.531 (.330)	-1.26 (2.13)	2.12 (1.79)	2.97*** (1.01)	3.10*** (0.87)
PF	-.740*** (.061)	-.491*** (.113)	-.412*** (.107)	-.419*** (0.099)	-.410*** (.102)
JOINT	9.87*** (0.29)	12.8*** (1.10)	.126*** (0.011)	11.7*** (7.0)	11.4*** (0.6)
SINGULARITY F-TEST		1.17**	1.21***	1.10*	0.95
R ²	0.92	0.96	0.99	0.99	0.99

Standard errors in parentheses. (***, **, *, significant at 1%, 5%, 10% level)

Table 5. Group means of predicted maximum biweekly salary.

Subgroup	OLS13 ^a	OLS9 ^b	EVCARP $\rho_{uij} = 0$	EVCARP $\rho_{uij} = 0.1$	EVCARP $\rho_{uij} = 0.2$
Managerial	965.1*	967.3	967.6	973.6	976.5*
Professional	820.2*	821.3	816.1*	814.8*	814.4*
Technical	640.5	640.3	639.8	640.9	639.8
Clerical	592.0*	593.8	570.1*	562.5*	560.5*
Service	552.3*	554.0	538.9*	529.6*	531.3*
<High school	496.9*	499.6	475.4*	462.2*	462.4*
High school	601.6*	602.9	589.5*	585.7*	585.3*
Technical school	675.4	675.2	668.7*	667.5*	667.2*
>College	871.8*	873.1	871.2	870.7	870.7
\$0-5 MW	511.7*	514.8	489.9*	470.7*	471.1*
\$5.01-7.5 MW	588.2*	589.9	569.9*	563.6*	562.9*
\$7.51-10 MW	696.8*	699.0	683.1*	679.6*	681.8*
>\$10 MW	836.3	836.2	841.1*	842.8*	841.4*
Aggregate	724.3*	725.4	717.4*	714.7*	714.5*

Notes: a. OLS regression of original Arthur Young 13 job evaluation factors.

b. OLS regression of 9 job evaluation factors using JOINT.

* Difference between OLS9 and predicted salary in the cell is significant at the .05 level.

Appendices

Table A.1. Observed correlation matrix.

	KFFTE	KFE	CJPS	GSA	PC	PD	MVD	SEN	SE	IE	WE	UH	WI	PF	MW
KFFTE	1.00	0.31	0.65	0.59	0.56	-0.49	0.12	0.22	0.55	0.56	-0.35	-0.19	0.30	-0.17	0.66
KFE	0.31	1.00	0.67	0.73	0.43	-0.32	-0.15	0.64	0.71	0.68	-0.22	-0.12	0.50	-0.33	0.50
CJPS	0.65	0.67	1.00	0.81	0.60	-0.39	-0.13	0.51	0.78	0.74	-0.30	-0.19	0.47	-0.24	0.60
GSA	0.59	0.73	0.81	1.00	0.61	-0.44	-0.15	0.57	0.83	0.74	-0.32	-0.19	0.48	-0.27	0.59
PC	0.56	0.43	0.60	0.61	1.00	-0.50	-0.26	0.32	0.64	0.58	-0.26	-0.14	0.43	-0.11	0.40
PD	-0.49	-0.32	-0.39	-0.44	-0.50	1.00	0.02	-0.20	-0.43	-0.36	0.63	0.36	-0.32	-0.10	-0.26
MVD	-0.12	-0.15	-0.13	-0.15	-0.26	0.02	1.00	-0.17	-0.20	-0.16	-0.07	-0.02	-0.16	0.15	-0.09
SEN	0.22	0.64	0.51	0.57	0.32	-0.20	-0.17	1.00	0.56	0.51	-0.12	0.01	0.45	0.21	0.31
SE	0.55	0.71	0.78	0.83	0.64	-0.43	-0.20	0.56	1.00	0.78	-0.30	-0.18	0.49	-0.26	0.55
IE	0.56	0.68	0.74	0.74	0.58	-0.36	-0.16	0.51	0.78	1.00	-0.19	0.00	0.50	-0.32	0.63
WE	-0.35	-0.22	-0.30	-0.32	-0.26	0.63	-0.07	-0.12	-0.30	-0.19	1.00	0.53	-0.12	-0.21	-0.14
UH	-0.19	-0.12	-0.19	-0.19	-0.14	0.36	-0.02	0.01	-0.18	0.00	0.53	1.00	-0.06	-0.18	-0.06
WI	0.30	0.50	0.47	0.48	0.43	-0.32	-0.16	0.45	0.49	0.50	-0.12	-0.06	1.00	-0.13	0.33
PF	-0.17	-0.33	-0.24	-0.27	-0.11	-0.10	0.15	-0.21	-0.26	-0.32	-0.21	-0.18	-0.13	1.00	-0.40
MW	0.66	0.50	0.60	0.59	0.40	-0.26	-0.09	0.31	0.55	0.63	-0.14	-0.06	0.33	-0.40	1.00

Table A.2. Reliability corrected correlation matrix assuming $\rho_{uuj} = 0$.

	KFFTE	KFE	CJPS	GSA	PC	PD	MVD	SEN	SE	IE	WE	UH	WI	PF	MW
KFFTE	1.00	0.38	0.74	0.72	0.66	-0.55	-0.17	0.23	0.68	0.68	-0.43	-0.22	0.42	-0.18	0.69
KFE	0.38	1.00	0.84	0.98	0.56	-0.40	-0.23	0.77	0.95	0.91	-0.31	-0.14	0.77	-0.39	0.58
CJPS	0.74	0.84	1.00	1.00	0.74	-0.46	-0.18	0.57	0.99	0.93	-0.38	-0.22	0.69	-0.26	0.65
GSA	0.72	0.98	1.00	1.00	0.81	-0.56	-0.24	0.70	1.00	1.00	-0.44	-0.24	0.76	-0.32	0.69
PC	0.66	0.56	0.74	0.81	1.00	-0.62	-0.40	0.38	0.85	0.77	-0.35	-0.17	0.65	-0.12	0.46
PD	-0.55	-0.40	-0.46	-0.56	-0.62	1.00	0.03	-0.22	-0.54	-0.45	0.81	0.42	-0.47	-0.11	-0.29
MVD	-0.17	-0.23	-0.18	-0.24	-0.40	0.03	1.00	0.24	0.31	-0.26	-0.11	0.04	0.29	0.21	-0.12
SEN	0.23	0.77	0.57	0.70	0.38	0.22	-0.24	1.00	0.68	0.62	-0.15	0.01	0.63	-0.22	0.32
SE	0.68	0.95	0.99	1.00	0.85	-0.54	-0.31	0.68	1.00	1.00	-0.41	-0.22	0.77	-0.30	0.64
IE	0.68	0.91	0.93	1.00	0.77	-0.45	0.26	0.62	1.00	1.00	-0.27	0.00	0.78	-0.37	0.74
WE	-0.43	-0.31	-0.38	-0.44	-0.35	0.81	-0.11	-0.15	-0.41	-0.27	1.00	0.68	-0.19	-0.25	-0.16
UH	-0.22	-0.14	-0.22	-0.24	-0.17	0.42	-0.04	0.01	-0.22	0.00	0.68	1.00	-0.09	-0.19	-0.06
WI	0.42	0.77	0.69	0.76	0.65	-0.47	-0.29	0.63	0.77	0.78	-0.19	-0.09	1.00	-0.18	0.45
PF	-0.18	-0.39	-0.26	-0.32	-0.12	-0.11	0.21	-0.22	-0.30	-0.37	-0.25	-0.19	-0.18	1.00	-0.40
MW	0.69	0.58	0.65	0.69	0.46	-0.29	-0.12	0.32	0.64	0.74	-0.16	-0.06	0.45	-0.40	1.00

Table A.3. Student's t-tests for subgroup means of predicted maximum salary between various OLS and EVCARP regressions.

Subgroup	OLS13 ^a	OLS9 ^b	OLS9 ^b	OLS9 ^b
	OLS9 ^b	EVCARP $\rho_{uij} = 0$	EVCARP $\rho_{uij} = 0.1$	EVCARP $\rho_{uij} = 0.2$
81-100 PF job	-6.61 ^{***}	9.50 ^{***}	6.60 ^{***}	6.12 ^{***}
21-80 PF job	-3.24 ^{***}	10.45 ^{***}	8.28 ^{***}	7.29 ^{***}
0-20 PF job	-1.55	-0.94	-0.35	-0.10
Union	-5.86 ^{***}	7.92 ^{***}	6.10 ^{***}	5.91 ^{***}
Not Union	-1.90 [*]	2.83 ^{***}	3.12 ^{***}	2.87 ^{***}
Managerial	-2.44 ^{**}	-0.10	-1.34	-1.97 ^{**}
Professional	-2.79 ^{***}	2.95 ^{***}	2.75 ^{***}	2.78 ^{***}
Technical	0.41	0.19	-0.14	0.10
Clerical	-3.93 ^{***}	10.53 ^{***}	6.14 ^{***}	6.20 ^{***}
Service	-4.64 ^{***}	6.10 ^{***}	5.99 ^{***}	4.93 ^{***}
<High school	-8.89 ^{***}	9.05 ^{***}	7.25 ^{***}	6.55 ^{***}
High school	-3.04 ^{***}	5.86 ^{***}	4.33 ^{***}	4.05 ^{***}
Tech. school	0.49	2.95 ^{***}	2.29 ^{***}	2.23 ^{**}
>College	-3.56 ^{***}	1.22	1.04	1.04
\$0-5 MW	-8.13 ^{***}	8.07 ^{***}	6.01 ^{***}	5.47 ^{***}
\$5.01-7.5 MW	-4.48 ^{***}	10.68 ^{***}	7.46 ^{***}	6.97 ^{***}
\$7.51-10 MW	-5.61 ^{***}	9.24 ^{***}	7.94 ^{***}	6.15 ^{***}
>\$10 MW	0.13	-2.93 ^{***}	-2.64 ^{***}	-2.00 ^{**}
Aggregate	-4.93 ^{***}	7.40 ^{***}	6.39 ^{***}	6.09 ^{***}

Notes: a. OLS regression using 13 factors.

b. OLS regression using 9 factors.

c. ***, **, and * denote 0.01, 0.05, and 0.10 level of significance, respectively.

Appendix

In the limit, the bias discussed in Section II.A approaches

$$(1-A) \quad \text{plim} (\hat{\delta}_{OLS} - \delta) = \text{plim} \{[(X'X)/n]^{-1}(X'v)/n\} \\ = \text{plim}[(X'X)/n]^{-1} \text{plim}[(X'v)/n]$$

where $v = e - \beta_1 u_1$, $\delta = (\beta_0, \beta_1, \gamma)'$ and $u_1 = (u_{11}, u_{12}, \dots, u_{1n})'$. Assume the following limits exist:

$$\text{plim} \left[\frac{(X'X)}{n} \right]^{-1} = \text{plim} \begin{bmatrix} \frac{n}{n} & \frac{\sum X_{1j}}{n} & \frac{\sum x_{2j}}{n} \\ \frac{\sum X_{1j}}{n} & \frac{\sum X_{1j}^2}{n} & \frac{\sum X_{1j}x_{2j}}{n} \\ \frac{\sum x_{2j}}{n} & \frac{\sum X_{1j}x_{2j}}{n} & \frac{\sum x_{2j}^2}{n} \end{bmatrix}^{-1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

$$\begin{aligned} \text{plim}[(X'v)/n] &= \text{plim}[X'(e - \beta_1 u_1)/n] \\ &= \text{plim}[\mathbf{1}'(e - \beta_1 u_1)/n, (X_1)'(e - \beta_1 u_1)/n, (x_2)'(e - \beta_1 u_1)/n]' \\ &= [0, -\beta_1 \sigma_{u1}, 0]' \end{aligned}$$

where σ_{u1} is the measurement error variance for X_1 . The expression (3) can then be written as

$$(2-A) \quad \text{plim}(\hat{\delta}_{OLS} - \delta) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} 0 \\ -\beta_1 \sigma_{u1} \\ 0 \end{bmatrix} \\ = - \begin{bmatrix} a_{12} \beta_1 \sigma_{u1} \\ a_{22} \beta_1 \sigma_{u1} \\ a_{23} \beta_1 \sigma_{u1} \end{bmatrix}$$

In
Se
cti

on III.B, assuming independent measurement errors, the new reliability ratio is

$$\kappa_{\text{JOINT}} = \frac{\sum_{i=1}^5 \kappa_{xi} \sigma_{xi}}{\sum_{i=1}^5 \sigma_{xi}}$$

When measurement errors are assumed to be correlated, the new joint reliability ratio is

$$\kappa_{\text{JOINT}} = \frac{\sum_{i=1}^5 \kappa_{xi} \sigma_{xi} + 2 \sum_{k \neq i} \sum_{i=1}^5 \rho_{uiuk} [(1 - \kappa_{xi})(1 - \kappa_{xk}) \sigma_{xi} \sigma_{xk}]^{1/2}}{\sum_{i=1}^5 \sigma_{xi}}$$