To Log or Not to Log: Bootstrap as an Alternative to the Parametric Estimation of Moderation Effects in the Presence of Skewed Dependent Variables

CRAIG J. RUSSELL
University of Oklahoma

MICHELLE A. DEAN
University of North Texas

When gross deviations from parametric assumptions are observed, conventional data transformations are often applied with little regard for substantive theoretical implications. One such transformation involves using the logarithm of positively skewed dependent variables. Log transformations were shown to severely decrease estimates of true moderator effects using moderated regression procedures in a Monte Carlo simulation. Estimates of moderator effect sizes were substantially better estimates of the true latent moderator effect (i.e., larger by a multiple of 2.6 to 534) when estimated using a simple percentile bootstrap procedure in the original, positively skewed data. Conclusions with regard to the presence or absence of a true moderator effect using a simple bootstrap procedure were unaffected by the violation of parametric assumptions in the original, positively skewed data. In contrast, moderated regression analysis performed on a log-transformed dependent variable severely increased Type-II error. Implications are drawn for applied psychological and management research.

At one time or another, almost all investigators in applied psychological and management research have been concerned by the assumptions required of common parametric statistical tests. Investigators typically assume that their samples were drawn from a single population and rely on the power of the central limit theorem and other parametric assumptions to draw inferences about latent relationships within that population. When violations of parametric assumptions are severe, investigators often use some data transformation designed to minimize the violation. For example, all three empiri-
cal studies reported in a recent special *Academy of Management Journal* forum on managerial compensation performed log transformations on compensation data (Conyon & Peck, 1998; Finkelstein & Boyd, 1998; Sanders & Carpenter, 1998) with no mention of the purpose or rationale behind these transformations. Presumably, the log transformations were done to address the presence of heteroskedasticity, that is, the lack of independence between the mean of $Y$ given $X$ ($\hat{\mu}_Y|X$) and the variance of $Y$ given $X$ ($\sigma^2_Y|X$) that coincides with extreme positive outliers or severe positive skew (Winer, 1974). Winer’s 1974 text has had a pervasive influence on organizational research as reflected in the fact that it is the most highly cited publication in the *Social Science Citation Index* (Institute for Scientific Information, 1999) between 1957 and 1997 (J. L. Bennett, personal communication, July 12, 1999)—it is difficult to underestimate the effect that Winer’s text (and its subsequent updates) has had on organizational researchers. It could be argued that performing log transformations on positively skewed dependent variables has become a convention within applied psychology and management research.

One of the following characteristics is required of studies using parametric ordinary least squares (OLS) procedures to examine linear relationships between variables $X$ and $Y$: (a) $X$ and $Y$ are random, bivariate normal or (b) $X$ is fixed and $e$ is normal, where $e_i = Y_i - b_1x_i - b_0$. In the former case, $X$ is random in the sense that investigators do not specify or control levels of $X$ treatment effects in advance. Instead, $X$ values observed occur at a frequency dictated by the population probability distribution for $X$. Common survey methods employed in research examining voluntary employee turnover (e.g., Mobley, Griffeth, Hand, & Meglino, 1979), job satisfaction (Smith, Kendall, & Hulin, 1969), performance prediction (Bray, Campbell, & Grant, 1974; Owens & Schoenfeldt, 1979), and executive compensation (Finkelstein & Boyd, 1998) provide examples of random-effects designs. Importantly, when $X$ and $Y$ are distributed bivariate normal, probabilistic inferences (e.g., conducting hypothesis test of $H_0: \rho = 0$ or estimating confidence intervals can be drawn due to $\rho$’s presence in the bivariate normal density function described in Equation 1:

\[
f(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp\left\{\frac{1}{2(1-\rho^2)} \left[ \frac{(x - \mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] \right\}
\]

If $X$ is not normally distributed, as in the latter case, one may still use the central limit theorem to assume $\sigma^2_Y|X$ is normally distributed in order to test hypotheses about $\rho$. In these circumstances, $X$ is often a fixed effect that takes on values occurring in some known frequency other than what one would have expected if values of $X$ were drawn at random from the population (e.g., values of $X$ that the investigator selected for purposes of manipulation). Importantly, regardless of study design, traditional parametric procedures cannot be used in conducting hypothesis tests or estimating CIs if the true probability density function for prediction error ($e$) is unknown.

As noted above, one common violation of parametric assumptions occurs when the variance of $Y$ given $X$ ($\sigma^2_Y|X$, $\sigma^2_Y|X$) is a function of the conditional mean ($\hat{\mu}_Y|X$). Efforts examining severely positively skewed $Y$ distributions routinely occur in applied psychological research, particularly in compensation research. Skewed compensation distributions are caused by a number of factors, including the increasing span of pay
ranges as the pay range midpoint increases (England & Pierson, 1990) and the extreme levels of executive compensation typically reported in U.S. corporations. Both factors result in a lack of independence between $Y_i$ and $s_i$, violating the homoskedasticity assumption (Winer, 1974).

Winer (1974, pp. 398-401) described a number of transformations that correct or at least lessen violations of some parametric assumptions. Log transformations of variables demonstrating highly positive or negative skew yield a more bell-shaped frequency distribution, where $Y_i$ and $s_i$ are relatively uncorrelated. Winer (1974, p. 400) noted that log transforms are particularly effective in stabilizing the conditional variance of $Y$ given $X$ when the independence of error terms is violated due to $s_i^2 = k_i^2 Y_i$ or when $Y$ has a great deal of positive skew (Olds, Mattson, & Oldeh, 1956).

The usual effect of such transforms is to lessen the prediction error for values of $Y$ occurring at the extreme tail of the positively skewed dependent variable, consequently increasing $r_{xy}^2$ in additive models used to predict log $Y$. The SYSTAT6.0 for Windows: Statistics (SPSS, 1996, pp. 252-257) manual described one such example in which gross domestic product (GDP) per capita ($X$) was used to predict military spending ($Y$) in a sample of 57 countries. In this example, $r_{xy}^2$ goes from .417 to .734 in the presence of log transformation.

Importantly, the resultant model using the transformed data is $Y_{mil} = 10^{b_0 + b_1 \log_{10} X_{GDP} + b_{10} e}$, which does not technically adhere to OLS characteristics (e.g., unbiased, minimum-variance parameter estimates). This model is perfectly serviceable if prediction is the investigator’s main concern—inferences about the accuracy of prediction can be drawn from $r_{xy}^2$. Note that probabilistic inferences cannot be drawn for $r_{xy}$, $b_0$, $b_1$, or $\hat{Y}$ unless one assumes that the log$_{10}$ term in $Y_{mil} = 10^{b_0 + b_1 \log_{10} X_{GDP} + b_{10} e}$ is normally distributed. We are aware of no research stream (theory based or otherwise) that holds that log$_{10}$ of $e$ is normal. Regardless, the model must have some theoretical meaning if explanation is the investigator’s main concern. For example, it is unclear what theory or policy implications should be drawn from finding that the log of salary is differentially related to organizational tenure for men and women. The authors are unaware of any studies examining interactive models providing a theoretical rationale justifying nonlinear (monotonic or nonmonotonic) transformations in applied psychological or management research (although concepts like the diminishing marginal utility of money may provide such a rationale in the future). Enhanced statistical elegance achieved via nonlinear transformations has not been accompanied by theory-based rationale justifying its use.

Nonlinear transformations can cause more uncertainty in interpreting tests of moderation than they resolve. Investigators generally need to estimate sample sizes required for replications and extensions of past research. Investigators examining previously reported data on the GDP–military spending relationships will solve

$$r_{xy} = t_{\alpha/2} \left( \frac{1 - r_{xy}^2}{N - 2} \right)$$

for $N$, estimating that they need a sample size, when $r_{xy}^2 = .417$, that is approximately four times as large as the sample required when $r_{xy}^2 = .734$ at $\alpha = .05$. Again, absent theoretical rationale, arguments can be mounted for either estimate.
Importantly, nonlinear (monotonic and nonmonotonic) transformations of original data create a number of problems for OLS applications used to detect moderator effects. Busemeyer and Jones (1983) demonstrated monotonic transformations could be found that cause Y values generated from a truly additive model (e.g., \( Y = b_0 + b_1 X_1 + b_2 X_2 \)) to yield support for a multiplicative model (e.g., \( Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 \)) and vice versa. Hence, reports of significant and nonsignificant interaction effects after having performed log transformation on Y remain open to alternative interpretation (cf. Henderson & Frederickson, 1996; Sanders & Carpenter, 1998).

In sum, investigators often face circumstances in which data are clearly not bivariate normal, \( e \) is not normal, and/or heteroskedasticity is present. Nonlinear transformations generate unknown levels of distortion in the many estimates of moderator effects required to test theories in management and applied psychology (Busemeyer & Jones, 1983; Russell & Bobko, 1992). Investigators’ continued use of nonlinear transforms to test moderator effects (e.g., Henderson & Frederickson, 1996; Kuhn & Sweetman, 1998; Sanders & Carpenter, 1998) will result in literatures characterized by mixed findings containing frequent Type-I and -II errors. This will be especially true when other investigators do not use nonlinear transformations in studying the same phenomena (e.g., Gomez-Mejia, Tosi, & Hinkin, 1987). Severe consequences for theory development will result.

The bootstrap is a relatively new method of empirically estimating characteristics of population distributions from sample data (Efron, 1979) that holds remarkable implications for these applied research issues. Unfortunately, Mooney and Duval (1993) noted that “the bootstrap is . . . foreign to most social scientists schooled in the traditional parametric approach to inference” (p. 27). This study briefly reviews the bootstrap literature and reports the results of a Monte Carlo simulation demonstrating how log transformations can yield spuriously low estimates of moderator effect sizes (i.e., \( \Delta R^2 \)). Finally, a bootstrap approach to detect interaction effects when authors would otherwise employ log transformations and traditional OLS techniques is presented, and implications for applied psychological and management research are offered.

**Bootstrap Estimation Procedures**

Bootstrapping holds promise as a statistical estimation technique yielding precise estimates of population distributions from sample data. Bootstrapping estimates the population distribution of a statistic (e.g., \( r_{xy} \)) by iteratively resampling cases from a set of observed data. Basically, \( B \) bootstrap samples of size \( N \) are taken with replacement from the original sample of size \( N \) and saved to a file. An investigation using \( B = 1,000 \) bootstrap samples of size \( N \) is able to approximate the actual sampling distribution that would have been obtained if multiple independent samples of size \( N \) were drawn from the population (Efron & Tibshirani, 1993).

There are many advantages to using the bootstrap technique. First, it is not restricted by the normality assumptions of parametric tests. The percentile bootstrapping method (Efron & Tibshirani, 1993, chapter 13) generates information about the latent population distribution, permitting estimation of CIs directly from the bootstrapped sampling distribution (e.g., if \( B = 1,000 \) bootstrap samples are taken, the bootstrap correlations \( r_b \) representing the 25th and 975th largest values constitute the lower
and upper points of a 95% CI). The graphical interpretation of \( r_b \) frequency distributions also yields insight into characteristics of the latent population distribution (Efron & Tibshirani, 1993). When the sample is drawn from a population with a single value of \( \rho \), the central limit theorem dictates that the \( r_b \) frequency distribution will rapidly approximate the normal distribution as \( B \) and \( N \) increase. A multimodal \( r_b \) frequency distribution would suggest that the sample was drawn from multiple populations, each with its own value of \( \rho \). Second, information about the form of the original sample is retained with no loss of distributional information. Rasmussen (1987) noted that loss of information does occur when nonparametric techniques convert data to ranks. Lunneborg (1985) described bootstrapping as falling between parametric and nonparametric procedures for making probabilistic inferences.

Rasmussen (1987) presented the following simple example of a bootstrap procedure. Suppose a researcher wants to test the null hypothesis that \( \rho_{xy} = 0 \) between first-year grade point averages (GPA) and Graduate Record Exam (GRE) scores using data obtained from 10 graduate students \( (H_0: \rho_{GPA,GRE} = 0) \). First, an initial bootstrap sample \( (B_1) \) is randomly drawn with replacement from these 10 observations, yielding the possibility of some observations being represented more than once in the bootstrap sample, whereas other observations may not be included. A single bootstrap sample may include the following cases: 5, 2, 8, 6, 2, 7, 9, 6, 1, and 2. Note that due to random sampling with replacement, Case 2 was included more than once, whereas Case 3 was not included in this first bootstrap sample \( (B_1) \). The 10 cases may result in a correlation of, say, .59. This procedure is repeated a large number of times \( (e.g., B = 1,000) \), and each \( r_b \) is saved to a separate file. Second, the bootstrap correlations \( (r_b) \) are rank ordered with the 25th and 975th correlations representing 95% CI end points. Finally, the null hypothesis \( H_0: \rho_{GPA,GRE} = 0 \) is tested by determining whether zero falls within the CI (Rasmussen, 1987).

Studies examining similarities in results obtained from bootstrap and normal theory approaches when parametric assumptions are met test the bootstrap’s ability to estimate true latent population distributions \( (e.g., \text{Diaconis \& Efron, 1983; Efron, 1985, 1986; Lunneborg, 1985}) \). These studies resulted in bootstrap statistics \( (e.g., \text{estimates of CIs}) \) that were extremely close to those generated from parametric approaches. Bickel and Freedman (1981; Freedman, 1981) demonstrated that the bootstrap was asymptotically valid for many statistics with known population probability distributions \( (e.g., t \text{ and OLS regression statistics}) \). However, the procedure is perhaps of most value in drawing inferences about statistics with unknown population probability distributions \( (e.g., \text{medians or mixed samples drawn from multiple populations}) \).

Some issues remain unresolved in using bootstrapping to conduct hypothesis testing, most revolving around the relative accuracy of parametric versus bootstrap procedures in estimating probability intervals at the extreme tails of known \( (i.e., \text{normal}) \) distributions. However, the simple percentile bootstrap method of estimating CIs described above provides “good theoretical coverage properties as well as reasonable stability in practice” (Efron & Tibshirani, 1993, p. 169). “Good theoretical coverage” refers to CIs that \( (a) \) accurately estimate the probability of the population parameter falling within the CI and \( (b) \) divide coverage error equally across the two tails. Empirical comparisons of bootstrap and traditional OLS regression procedures’ abilities to detect moderator effects when the dependent variable is positively skewed are presented below.
Monte Carlo Simulation

Design

In typical random-effects designs, investigators do not know how independent variables and prediction error are distributed. In fixed-effects designs, investigators typically control or specify independent variable levels, although the dependent $Y$ distribution will be a function of the independent variable(s) and prediction error ($e$) distributions. Classical measurement theory presumes $Y_i = Y_{Ti} + e$, where $Y_{Ti}$ is the true latent value of $Y$ for person $i$. When $Y_{Ti}$ is a function of some $X$ (e.g., $Y = b_0 + b_1X_1 + e$ or $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + e$), $X_1$, $X_2$, or $e$ must be nonnormal in order for the observed $Y_i$ to be nonnormally distributed. Consequently, to simulate the kinds of data that investigators might encounter in either random- or fixed-effects designs, data were generated in nine Monte Carlo simulations in which independent variables $X_1$ and $X_2$ and prediction error ($e$) systematically varied across normal, uniform, and chi-square distributions. Normal distributions were selected to simulate multivariate normal conditions in random-effects designs. Uniform distributions were selected to simulate fixed-effects experimental designs. Chi-square distributions for $X$ and $e$ simulated positively skewed $Y$ distributions such as those found in compensation research.

Sample

Simulation data were generated for combinations of $X_1$, $X_2$, and $e$ distributions using the SYSTAT9 (SPSS, 1996) computer package. Five thousand samples of $N = 113$ paired $X_1$, $X_2$ observations were drawn at random from all possible combinations of normal, uniform, and chi-square population distributions $X_1$, $X_2$, and $e$ (Guzzo, Jette, & Katzell [1985] reported a mean $N = 113$ across studies in a meta-analysis of compensation-based intervention programs). Results are only reported for conditions in which $X_1$ and $X_2$ were drawn from identical population distributions, although the results when $X_1$ and $X_2$ were drawn from different population distributions were consistent with those reported below. Note that 5,000 samples of $N = 113$ were drawn for every combination of $X$, $Y$, and $e$ distributions described below as per Mooney’s (1997) suggestions for conducting Monte Carlo simulations, resulting in nine sets of 5,000 samples of $N = 113$. All aspects of the Monte Carlo simulation were replicated using 5,000 samples of $N = 226$ and $N = 56$ (i.e., using samples twice and one half as large as $N = 113$). Identical patterns of results emerged and are available from the first author on request.

When $X_1$ and $X_2$ observations were drawn at random from a normal population distribution, $\mu$ and $\sigma$ were set at $\mu = 3$ and $\sigma = 1$. Variables $X_1$ and $X_2$ within each data set were then rounded to the nearest integer (yielding values ranging from 1 to 5, i.e., 5-point Likert scales) in order to simulate measurement circumstances commonly encountered in applied psychological and management research. Uniform $X_1$ and $X_2$ data sets were drawn from a population containing integer values between 1 and 5, inclusive. Additional $X_1$ and $X_2$ data sets were drawn from chi-square distributions with three degrees of freedom. These steps resulted in nine Monte Carlo data sets when the three possible $X$ distributions (normal, uniform, chi-square) were combined with the three possible $e$ distributions (normal, uniform, chi-square).
Three dependent variables were generated within each data set to reflect large, medium, and small effect sizes. Equations 2, 3, and 4 were used to generate values for $Y_1$, $Y_2$, and $Y_3$ within each of the nine data sets:

$$Y_1 = .75X_1X_2 + .25\epsilon$$  \hspace{1cm} (2) \\
$$Y_2 = .50X_1X_2 + .50\epsilon$$  \hspace{1cm} (3) \\
$$Y_3 = .25X_1X_2 + .75\epsilon$$  \hspace{1cm} (4)

Under the $\epsilon$-equals-normal condition, prediction error $\epsilon$ was drawn from a normal population with a mean and standard deviation set equal to the mean and standard deviation of the $X_1X_2$ product term with which it was paired. Under the $\epsilon$ = uniform condition, $\epsilon$ was randomly drawn from a uniform population distribution ranging from 1 to 20. Under the $\epsilon$ = chi-square condition, $\epsilon$ was randomly drawn from a chi-square population distribution with 9 degrees of freedom (where 9 is the mean population value for all $X_1X_2$ product terms regardless of sample $X_1$ and $X_2$ distribution characteristics). Hence, three dependent variables, $Y_1$, $Y_2$, and $Y_3$, reflecting large, medium, and small moderator effect sizes were available to be examined within each of the nine data sets.

Analyses

All tests of interaction effects used moderated regression analysis (Bobko, 1995; Darlington, 1968; Saunders, 1955, 1956). The $F$ test of $H_0: \Delta R^2 = 0$, where $\Delta R^2 = R^2_{\text{multiplicative}} - R^2_{\text{additive}}$, for the equations $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$ and $\hat{Y} = b_0 + b_1X_1 + b_2X_2$, respectively, constitutes the test of an interaction effect when $X_1$ and $X_2$ are interval scale measures. The strategy and organizational theory literatures commonly refer to this as the Chow test (Chow, 1960).

To provide a point of reference, samples of $N = 50,000$ for each combination of $X_1$ and $X_2$ distribution were generated separately for purposes of estimating $E(\Delta R^2)$ when $\epsilon = 0$. When $X_1$ and $X_2$ were normal, uniform, and chi-square, $E(\Delta R^2) = .057$, .077, and .256, respectively. These values should be considered asymptotes or what would occur under circumstances of perfect, error-free prediction. The addition of prediction error will slowly decrease $E(\Delta R^2)$, for example, if when $X_1$ and $X_2$ are distributed as chi-square the true prediction model is $Y = .1X_1X_2 + .9\epsilon$, then clearly $E(\Delta R^2) \neq .256$. Regardless, it should be noted that these are expected values of $\Delta R^2$ and actual values observed might be larger or smaller when $Y$ does or does not include prediction error (e.g., Russell & Bobko [1992] observed $\Delta R^2$ to be greater than $E[\Delta R^2]$ for some subjects).

Results

Table 1a reports results of moderated regression analyses performed on the three effect sizes ($Y_1$, $Y_2$, and $Y_3$) in the nine different combinations of $X$ and $\epsilon$ distributions (i.e., normal, uniform, and chi-square $X_1$ and $X_2$ distributions paired with normal, uniform, and chi-square $\epsilon$ distributions). Moderator effect sizes are captured by the
median $\Delta R^2$ column, containing the 2,500th largest value of $\Delta R^2$ obtained from the 5,000 samples of $N = 113$. Although $F$ statistics testing $H_0: \Delta R^2 = 0$ can be derived for median $\Delta R^2$ values, only the ones derived for normally distributed prediction error meet parametric assumptions and are interpretable (i.e., statistics reported in bold in Table 1a). Regardless, the 2.5 and 97.5 percentile values of $\Delta R^2$ were identified within the set of 5,000 Monte Carlo $N = 113$ samples. As the expected value of the $F$ statistic testing $H_0: \Delta R^2 = 0$ is $F = 1.0$, one would reject $H_0$ using logic that underlies simple percentile bootstrap applications when the $F$ statistic

\[
(i.e., F_{1,109} = \frac{\Delta R^2 / (4 - 3)}{(1-R_{\text{multiplicator}}^2) / (113 - 4)})
\]

for the moderator effect cutting off the lower 2.5% of the 5,000 $\Delta R^2$ is greater than 1 (i.e., when $F = 1.0$ falls outside of the 95% $\Delta R^2$ CI). Median values of $\Delta R^2$ reported in Table 1a for which the lower 2.5 percentile values generated $F$ greater than 1.0 are indicated in italic.

Interestingly, profiles of $\Delta R^2$ for large, medium, and small effect sizes for interpretable equations in Table 1a (i.e., those meeting OLS assumptions) are .047, .024, .006; .067, .041, .009; and .221, .191, .087 for $X_1$ and $X_2$ distributions drawn from normal, uniform, and chi-square populations of $X_1$ and $X_2$, respectively. Not surprisingly, smaller values of $\Delta R^2$ are observed as the effect size decreases across $Y_1$, $Y_2$, and $Y_3$. The pattern of effect sizes across normal, uniform, and chi-square distributions is consistent with McClelland and Judd’s (1993) demonstration that multiplicative effect sizes are maximized in designs using extreme values of $X_1$ and $X_2$. Normally distributed $X_1$ and $X_2$ will have the fewest extreme $X_1X_2$ observations due to low probabilities in the extreme tails of the normal distribution. Uniform and chi-square distributions for $X_1$ and $X_2$ will have increasingly more frequent extreme observations in the tails of an $X_1X_2$ distribution, respectively.

If $X_1$, $X_2$, or $e$ are highly positively skewed, as they are when drawn from $\chi^2_{c}$ populations, $Y$ will demonstrate some skewness. Investigators following Winer’s (1974) convention would perform a log transform on $Y$ in hope of permitting probabilistic inferences that are possible when parametric assumptions are met. Table 1b reports moderated regression results when $Y_1$, $Y_2$, and $Y_3$ were subjected to a log10 transformation for the five $X_1$, $X_2$, and $e$ combinations involving skewed chi-square distributions (when $X_1$, $X_2$, or $e$ are positively skewed, $Y$ will be positively skewed). Moderated regression effect sizes for the nontransformed $Y_1$, $Y_2$, and $Y_3$ (Table 1a) are 2.7 to 15 times larger than effect sizes observed for log-transformed $Y_1$, $Y_2$, and $Y_3$ (Table 1b). Perhaps most interestingly, effect sizes for the one data set that meets parametric assumptions ($X_1$ and $X_2$ distributed as chi-square, $e$ distributed normally) go from $\Delta R^2 = .221$ to $\Delta R^2 = .032$ for $Y_1$ and log$_{10} Y_1$, respectively; from $\Delta R^2 = .191$ to $\Delta R^2 = .041$ for $Y_2$ and log$_{10} Y_2$, respectively; and from $\Delta R^2 = .087$ to $\Delta R^2 = .025$ for $Y_3$ and log$_{10} Y_3$, respectively. Hence, moderated regression effect sizes are 3.5 to 6.9 times larger and more likely to correctly detect the true latent population moderator effect when estimated from the nontransformed data, although investigators following convention would have log-transformed $Y_1$, $Y_2$, and $Y_3$ before conducting the analyses. The stronger the moderator effect, the larger the difference between effect sizes derived from nontransformed versus log-transformed $Y$s.
Table 1a
Monte Carlo Moderated Regression Analyses for Different X and e Distributions: 5,000 Random Samples of $N = 113$
($Y_1 = .75X_1X_2 + .25e$, $Y_2 = .5X_1X_2 + .5e$, and $Y_3 = .25X_1X_2 + .75e$)

| $X_1$ and $X_2$ Distribution Characteristic | e Distribution Characteristic | Median ($R^2_{\text{multiplicative}} - R^2_{\text{additive}} = \Delta R^2$) |
|---------------------------------------------|-------------------------------|----------------------------------|
| Normal: $\mu = 3$, $\sigma = 1$, rounded to nearest integer | Normal: $\bar{e} = 9$, $\sigma = 5^a$ | $Y_1$ | $Y_2$ | $Y_3$ |
| | | .943 | .918 | .047 |
| | | .786 | .771 | .024 |
| | | .298 | .288 | .006 |
| Uniform random number between 1 and 18 | | $Y_1$ | .919 | .893 | .047 |
| | | $Y_2$ | .726 | .709 | .025 |
| | | $Y_3$ | .293 | .281 | .007 |
| Chi-square: $df = 9^b$ | | $Y_1$ | .943 | .918 | .047 |
| | | $Y_2$ | .732 | .712 | .029 |
| | | $Y_3$ | .383 | .373 | .008 |
| Uniform random number: distribution from 1 to 5, rounded to nearest integer | Normal: $\bar{e} = 9$, $\sigma = 5^a$ | $Y_1$ | .956 | .920 | .067 |
| | | $Y_2$ | .803 | .778 | .041 |
| | | $Y_3$ | .360 | .347 | .009 |
| Uniform random number between 1 and 20 | | $Y_1$ | .960 | .924 | .068 |
| | | $Y_2$ | .648 | .616 | .041 |
| | | $Y_3$ | .431 | .420 | .009 |
| Chi-square: $df = 9^b$ | | $Y_1$ | .969 | .933 | .070 |
| | | $Y_2$ | .713 | .680 | .046 |
| | | $Y_3$ | .496 | .484 | .012 |
| Chi-square: $df = 3$, rounded to nearest integer | Normal: $\bar{e} = 9$, $\sigma = 5^a$ | $Y_1$ | .991 | .873 | .221 |
| | | $Y_2$ | .910 | .798 | .191 |
| | | $Y_3$ | .561 | .477 | .087 |
| Uniform random number between 1 and 18 | | $Y_1$ | .992 | .874 | .220 |
| | | $Y_2$ | .924 | .815 | .190 |
| | | $Y_3$ | .658 | .589 | .086 |
| Chi-square: $df = 9^b$ | | $Y_1$ | .992 | .874 | .221 |
| | | $Y_2$ | .939 | .827 | .197 |
| | | $Y_3$ | .647 | .561 | .104 |

Note. $N = 113$ is the average $N$ across $k = 330$ effect sizes reported in a meta-analysis of Guzzo, Jette, and Katze (1985). Data presented in bold met parametric assumptions and are interpretable. Results presented in italics had the 2.5 percentile value of $F$ for $\Delta R^2$ greater than 1, hence $H_0$: $F = 1.0$ did not fall in the 95% CI.

a. $M$ and $SD$ for all normally distributed error terms were set to be equal to the $M$ and $SD$ for the product $X_1X_2$.

b. The expected value of the chi-square distribution is equal to its $df$. Hence, with $df = 9$, the expected midpoint of the error distribution is equal to the mean of the $X_1X_2$ product term.
In sum, moderated regression effect sizes derived from a Monte Carlo simulation of 5,000 \( N = 113 \) samples drawn from normal, uniform, and chi-square \( e \) and \( X \) distributions are 2.7 to 15 times more likely to detect true latent moderator effects (i.e., reject \( H_0: \Delta R^2 = 0 \)) when the dependent variable has not been subjected to a log transformation. The final portion of this study demonstrates how primary researchers would apply a simple bootstrap procedure in analyzing data obtained from a single sample and confirming implications of the Monte Carlo results (i.e., that inferences drawn from bootstrap-generated CIs about moderator effects are expected to exhibit less Type-II error).

### Bootstrap Demonstration

#### Samples

As a rule, researchers generally face circumstances in which they have data gathered from a single sample, not 5,000 samples. Hence, to simulate what individual researchers typically encounter, nine samples of \( N = 113 \) paired \( X_1, X_2 \) observations...
were created at random from normal, uniform, and chi-square population distributions using the SYSTAT9 computer package. When \( X_1 \) and \( X_2 \) observations were drawn at random from a normal population distribution, \( \mu \) and \( \sigma \) were set at \( \mu = 3 \) and \( \sigma = 1 \). As in the Monte Carlo simulation and consistent with measurement circumstances commonly encountered in applied psychological and management research, \( X_1 \) and \( X_2 \) data sets were rounded to the nearest integer, yielding values from 1 to 5. Uniform \( X_1 \) and \( X_2 \) data sets were drawn from a population containing integer values between 1 and 5, inclusive, yielding \( \bar{X}_1 = 3.012, \sigma_{X_1} = 1.438 \), and \( \bar{X}_2 = 2.889, \sigma_{X_2} = 1.394 \), respectively. Finally, \( X_1 \) and \( X_2 \) data sets were drawn from chi-square distributions with 3 degrees of freedom, yielding \( \bar{X}_1 = 2.986, \sigma_{X_1} = 2.344 \), and \( \bar{X}_2 = 3.008, \sigma_{X_2} = 2.660 \), respectively. Three dependent variables were generated within each sample using Equations 2, 3, and 4 described in the Monte Carlo simulation above. Error terms (\( e \)) were drawn from the same populations as described in the Monte Carlo simulation above, with their means and standard deviations set equal to the \( X_1X_2 \) product term means and standard deviations.

**Analyses**

Tests of interaction effects using moderated regression analysis were performed using dependent variables \( Y_1, Y_2, Y_3, \log Y_1, \log Y_2, \) and \( \log Y_3 \) in each of the nine samples. In addition, \( B = 1,000 \) bootstrap estimates of \( D\hat{R}^2 \) were derived for all dependent variables in each of the nine samples using the percentile bootstrap method described above.9

**Results**

Table 2a reports results of moderated regression analyses performed on the nine samples of \( N = 113 \) containing different combinations of \( X_1, X_2, \) and \( e \) distributions (i.e., normal, uniform, and chi-square distributions of \( X_1 \) and \( X_2 \) paired with normal, uniform, and chi-square \( e \) distributions). Although \( F \) statistics are reported for moderation effects in all nine combinations, only the three derived for normally distributed prediction error meet parametric assumptions and are interpretable (i.e., statistics reported in bold in Table 2a). Moderator effect sizes are captured by the \( D\hat{R}^2 \) column (the \( F \) statistic tests \( H_0: \Delta\hat{R}^2 = 0 \) (Bobko, 1995; Darlington, 1968).

\( D\hat{R}^2 \) for \( Y_1 \) in the three interpretable equations in Table 2a are .049, .068, and .216 for \( X_1 \) and \( X_2 \) distributions drawn from normal, uniform, and chi-square populations, respectively. This profile of effect sizes is again consistent with the observation that normal \( X_1 \) and \( X_2 \) will have the fewest extreme \( X_1X_2 \) observations due to low probabilities in the extreme tails of the normal distribution and results reported in the Monte Carlo study reported above.

Figure 1 demonstrates when \( X_1, X_2, \) or \( e \) were highly positively skewed, as they are when drawn from \( \chi^2_{1} \) populations, \( Y \) exhibited some positive skewness. Investigators following convention would perform a log transform on \( Y \) hoping to permit the probabilistic inferences that are possible when parametric assumptions are met. Table 2b reports moderated regression results when \( Y \) was subjected to a \( \log_{10} \) transformation for the five \( X_1, X_2, \) and \( e \) combinations with skewed chi-square distributions (i.e., skewed \( Y \) distributions appear only when \( X_1, X_2, \) or \( e \) distributions were positively skewed). Consistent with the Monte Carlo findings reported above, moderated regres-
sion effect sizes for the original nontransformed data were two to seven times larger than effect sizes observed for log-transformed data. Effect sizes for the one data set that met parametric assumptions (\(X_1\) and \(X_2\) distributed as chi-square, \(\epsilon\) distributed normally) went from \(\Delta R^2 = .216\) to \(\Delta R^2 = .030\) when \(Y\) was subjected to log transformation. Hence, moderated regression effect size was 7.2 times larger when it was (correctly) estimated from nontransformed data.

However, CIs around \(\Delta R^2\) can be derived via bootstrapping procedures regardless of how \(X_1\), \(X_2\), \(\epsilon\), or \(Y\) are distributed. Table 3 reports bootstrap estimates of the 2.5 percentile values of the moderated regression effect size \(\Delta R^2\) taken from \(B = 1,000\) bootstrap samples of size \(N = 113\) for the five situations in which \(Y\) is positively skewed, that is, those that are subject to log transformation using the current methodological convention. Interestingly, median effect sizes across 1,000 bootstrap samples were between 2.6 and 534 times larger than \(\Delta R^2\) effect sizes resulting from analyses conducted after \(Y\) was log transformed (see Table 2b). This suggests that to be equally likely to be detected, moderator effect sizes when \(Y\) is skewed and log transformed must be 2.6 to 534 times as large as those observed under conditions when \(Y\) is not log transformed and \(\Delta R^2\) is estimated from the median bootstrap \(\Delta R^2\) value. Alternatively, other things being equal, the sample size needed to correctly reject \(H_0: \Delta R^2 = 0\) would need to be 6.76 to 285,156 times as large when \(Y\) is skewed and log transformed in these samples. Investigators using OLS moderated regression and log-transformed dependent variables would be much more likely to fail to detect true interaction effects (Type-II error).

**Discussion**

This study demonstrated a fundamental problem in the detection of latent moderation effects when log transforms are used to correct positively skewed dependent variables. Specifically, the increased probability of Type-II error was demonstrated in a Monte Carlo simulation generating 5,000 samples from known population distributions and in a subsequent bootstrap analysis of individual simulated samples. Results suggested severe decrements in the statistical power required to test moderation regression effects (i.e., \(H_0: \Delta R^2 = 0\)) that result from log transformations. These decrements occurred when parametric assumptions were in fact met (i.e., the data reported in bold in Tables 1a and 2a) and when parametric assumptions were not met. Graphically, log transformations change the \(Y\) distribution shape, effectively decreasing \(Y\) variance by reducing the degree to which extremely positive \(Y\) values deviate from the mean. If these extreme \(Y\) values were created by an interaction between one or more positively skewed independent variables (e.g., when \(X_1\) and \(X_2\) are distributed as chi-square), log transformations of \(Y\) effectively disguise the extreme values of \(Y\) that should result from the product of extreme \(X_1\) and \(X_2\) values as less extreme values, effectively yielding a log \(Y\) variable that exhibits less variance than the original raw \(Y\) observations. Although Type-I error is always possible (cf. Aguinis & Pierce, 1998), it is clear that log transformations of positively skewed dependent variables greatly enhance the probability of Type-II error.

(text continues on p. 182)
Moderated Regression Analyses for Different X and e Distributions: N = 113 (Y₁ = .75X₁X₂ + .25e, Y₂ = .5X₁X₂ + .5e, and Y₃ = .25X₁X₂ + .75e)

| X₁ and X₂ Distribution Characteristic | e Distribution Characteristic | R² multiplicative | R² additive | ΔR² | F₁,₁₀⁹ | p Value |
|---------------------------------------|-----------------------------|------------------|------------|-----|--------|---------|
| Normal: μ = 3, σ = 1, rounded to nearest integer | Normal: e = 8.789, SD = 4.996<sup>a</sup> |                   |            |     |        |         |
|                                       | Y₁                          | .799             | .750       | .049 | 5.616  | < .01  |
|                                       | Y₂                          | .587             | .561       | .026 | 2.910  | > .05  |
|                                       | Y₃                          | .331             | .322       | .009 | 0.990  | > .05  |
| Uniform random number between 1 and 20 |                               |                   |            |     |        |         |
|                                       | Y₁                          | .880             | .821       | .059 | 6.834  | < .05  |
|                                       | Y₂                          | .601             | .580       | .021 | 2.357  | > .05  |
|                                       | Y₃                          | .291             | .282       | .009 | 0.990  | > .05  |
| Chi-square: df = 9<sup>b</sup>       |                               |                   |            |     |        |         |
|                                       | Y₁                          | .932             | .887       | .045 | 5.136  | < .05  |
|                                       | Y₂                          | .599             | .567       | .032 | 3.603  | > .05  |
|                                       | Y₃                          | .333             | .329       | .004 | 0.438  | > .05  |
| Uniform random number: distribution from 1 to 5, rounded to nearest integer | Normal: e = 8.702, SD = 6.029<sup>a</sup> |                   |            |     |        |         |
|                                       | Y₁                          | .825             | .757       | .068 | 7.953  | < .01  |
|                                       | Y₂                          | .633             | .590       | .043 | 4.898  | < .05  |
|                                       | Y₃                          | .436             | .425       | .011 | 1.110  | > .05  |
| Uniform random number between 1 and 20 |                               |                   |            |     |        |         |
|                                       | Y₁                          | .831             | .787       | .044 | 4.561  | < .05  |
|                                       | Y₂                          | .599             | .580       | .019 | 2.111  | > .05  |
|                                       | Y₃                          | .395             | .383       | .012 | 1.324  | > .05  |
| Chi-square: df = 9<sup>b</sup>       |                               |                   |            |     |        |         |
|                                       | Y₁                          | .813             | .792       | .071 | 9.269  | < .01  |
|                                       | Y₂                          | .701             | .651       | .050 | 5.737  | < .05  |
|                                       | Y₃                          | .521             | .479       | .042 | 4.779  | < .05  |
Chi-square: $df = 3$, rounded to nearest integer

Normal: $\bar{e} = 8.982$, $SD = 6.163^a$

|     | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $Y_6$ | $Y_7$ | $Y_8$ | $Y_9$ | $Y_{10}$ |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------|
|     | .846  | .630  | .216  | .3030 | .003  |       |       |       |       |           |
| $Y_1$ | .620  | .460  | .160  | .12976| .001  |       |       |       |       |           |
| $Y_2$ | .461  | .355  | .106  | .12924| .001  |       |       |       |       |           |
| Uniform random number between 1 and 20
| $Y_1$ | .906  | .688  | .214  | .29677| .01   |       |       |       |       |           |
| $Y_2$ | .782  | .642  | .140  | .17744| .01   |       |       |       |       |           |
| $Y_3$ | .506  | .446  | .060  | .06957| .01   |       |       |       |       |           |
| Chi-square: $df = 9^b$
| $Y_1$ | .949  | .799  | .150  | .19235| .01   |       |       |       |       |           |
| $Y_2$ | .682  | .581  | .101  | .12246| .01   |       |       |       |       |           |
| $Y_3$ | .466  | .426  | .040  | .04542| .05   |       |       |       |       |           |

*Note. Only statistics appearing in bold are interpretable under parametric assumptions and $F$ statistics test $H_0: \Delta R^2 = 0$. $N = 113$ is the average $N$ across $k = 330$ effect sizes reported in a meta-analysis of Guzzo, Jette, and Katzell (1985).

a. $M$ and $SD$ for all normally distributed error terms were set to be equal to the $M$ and $SD$ for the product $X_1X_2$.

b. The expected value of the chi-square distribution is equal to its $df$. Hence, with $df = 9$, the expected midpoint of the error distribution is equal to the mean of the $X_1X_2$ product term.
Figure 1: Frequency Distribution for $X_1, X_2$ and $Y$ ($X_1$ and $X_2$ are chi-square distributed and $Y = .75X_1X_2 + .25e$)

Note. The variable $e$ is drawn from random normal or random uniform distributions with $e$ and $\sigma_e$ set equal to $X_1X_2$ and $\sigma_{X_1X_2}$, respectively.
Table 2b
Moderated Regression Results for $\log_{10}(Y_1) = .75X_1X_2 + .25e$, $\log_{10}(Y_2) = .5X_1X_2 + .5e$, and $\log_{10}(Y_3) = .25X_1X_2 + .75e$

| $X_1$ and $X_2$ Distribution Characteristic | $e$ Distribution Characteristic | $R^2_{\text{multiplicative}}$ | $R^2_{\text{additive}}$ | $\Delta R^2$ | $F_{1, 109}$ | $p$ Value |
|---------------------------------------------|---------------------------------|-------------------------------|-------------------------|--------------|------------|-----------|
| Normal                                      | Chi-square: $df = 9^a$          |                               |                         |              |            |           |
| $\log_{10} Y_1$                            |                                 | .757                          | .736                    | .021         | 2.338      | > .05     |
| $\log_{10} Y_2$                            |                                 | .511                          | .500                    | .011         | 1.100      | > .05     |
| $\log_{10} Y_3$                            |                                 | .409                          | .402                    | .007         | 0.768      | > .05     |
| Uniform random number                      | Chi-square: $df = 9^a$          |                               |                         |              |            |           |
| $\log_{10} Y_1$                            |                                 | .880                          | .877                    | .003         | 0.328      | > .05     |
| $\log_{10} Y_2$                            |                                 | .567                          | .566                    | .001         | 0.109      | > .05     |
| $\log_{10} Y_3$                            |                                 | .399                          | .399                    | .000         | 0.087      | > .05     |
| Chi-square: $df = 3$                       | Normal: $\chi^2 = 8.808$, $SD = 6.029^b$ |                               |                         |              |            |           |
| $\log_{10} Y_1$                            |                                 | .553                          | .523                    | .030         | 3.371      | > .05     |
| $\log_{10} Y_2$                            |                                 | .399                          | .385                    | .014         | 1.548      | > .05     |
| $\log_{10} Y_3$                            |                                 | .278                          | .269                    | .009         | 0.990      | > .05     |
| Uniform random number between 1 and 20     |                                 |                               |                         |              |            |           |
| $\log_{10} Y_1$                            |                                 | .611                          | .550                    | .061         | 7.081      | < .05     |
| $\log_{10} Y_2$                            |                                 | .456                          | .425                    | .031         | 3.487      | > .05     |
| $\log_{10} Y_3$                            |                                 | .342                          | .331                    | .011         | 1.100      | > .05     |
| Chi-square: $df = 9^a$                     |                                 |                               |                         |              |            |           |
| $\log_{10} Y_1$                            |                                 | .678                          | .588                    | .090         | 10.780     | < .01     |
| $\log_{10} Y_2$                            |                                 | .444                          | .399                    | .045         | 5.136      | < .05     |
| $\log_{10} Y_3$                            |                                 | .367                          | .348                    | .019         | 2.111      | > .05     |

a. The expected value of the chi-square distribution is equal to its $df$. Hence, with $df = 9$, the expected midpoint of the error distribution is equal to the mean of the $X_1X_2$ product term.
b. $M$ and $SD$ for all normally distributed error terms were set to be equal to the $M$ and $SD$ for the product $X_1X_2$. 
Fortunately, results also indicated that bootstrapping procedures provide a viable alternative to traditional, parametric statistical procedures for detecting moderator effects regardless of how $X_1$, $X_2$, and $e$ are distributed. In fact, in situations in which convention dictates that $Y$ should be subjected to log transformation, log transformations caused extremely severe decrements in statistical power for parametric OLS procedures relative to bootstrap procedures. Data simulated here are commonly found in compensation research, where parametric procedures are commonly used after $Y$ has been subjected to a routine log transformation (e.g., Henderson & Frederickson, 1996; Sanders & Carpenter, 1998).

Of course, log transformations could be justified on some theoretical basis. The authors are unaware of any theoretical rationale put forth by compensation theory or any other area of applied psychological or management research to justify such a transformation in the presence of a multiplicative model. Furthermore, the authors have never seen any discussion of the theoretical underpinnings of latent models that result from such a transformation, such as $Y = 10^{b_0 + b_1 X_1 + b_2 X_2}$ (SPSS, 1996). As a result, any
gains in statistical elegance and predictive power (i.e., for additive models) stemming from log transformations are not currently matched by gains in theoretical insight. Null results for tests of moderation in studies employing log transformations are expected to frequently reflect Type-II errors when a true latent moderation process is present.

In sum, when hypothesized models involve interaction effects, applied psychological and management research would benefit from a routine application of the bootstrap procedures. Although they do not replace common parametric procedures, bootstrap applications are appropriate when parametric assumptions are not viable (e.g., when heteroskedasticity is present due to a positively skewed dependent variable). Nonlinear monotonic transformations may achieve necessary statistical conditions for parametric inferences in OLS applications to additive models (Busemeyer & Jones, 1983; Winer, 1974). The results indicated that nonlinear monotonic transformations erode investigators’ capacity to assess theory-based predictions of moderation effects (e.g., estimates of moderation effect $\Delta R^2$). Importantly, bootstrapping provides an alternative method of assessing theory-based inferences of moderation effects from data that cannot be assessed with comparable statistical power by conventional procedures.

**Notes**

1. Note that if $X$ is not normally distributed and $Y$ is a linear function of $X$, $Y$ will also likely not be normally distributed.
2. Bergman, Scarpello, and Hills (1998) and Milkovich and Newman (1996) noted how pay ranges are generally a constant or increasing percentage of the range midpoint. Hence, as the pay range midpoint (and mean) increases, the variation in observed salaries around the midpoint increases.
3. Of course, weighted least square (WLS) procedures would also resolve the heteroskedasticity problem. However, it would do so by migrating what was a nonlinear transformation paired with ordinary least squares into the internal optimal weighting procedures that are characteristic of WLS.
4. We thank an anonymous reviewer for bringing this to our attention.
5. We thank an anonymous reviewer for this example.
6. See Efron and Tibshirani (1993), chapter 14 (pp. 178-201) for a discussion of alternatives to the simple bootstrap. Specifically, the bias-corrected and accelerated (BC$_a$) and the approximate bootstrap confidence (ABC) interval methods are marginally more complex techniques that overcome most shortcomings associated with the simple bootstrap.
7. The first author will provide these results on request.
8. $\Delta R^2$ for the 2.5 and 97.5 percentile values are available from the first author on request.
9. See Efron and Tibshirani (1993), chapter 14 (pp. 178-201) for more elaborate bootstrap procedures exhibiting certain statistical elegance that might yield more robust confidence intervals.

**References**

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1*, 296-305.

Bergman, T. J., Scarpello, V. G., & Hills, F. S. (1998). *Compensation decision making* (3rd ed.). San Diego, CA: Dryden Press.

Bickel, P., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics, 9*, 1196-1217.
Bobko, P. (1995). *Correlation and regression: Principles and applications of industrial/organizational psychology and management*. New York: McGraw-Hill.

Bray, D. W., Campbell, R. J., & Grant, D. L. (1974). *Formative years in business*. Malabar, FL: Robert Krieger.

Busemeyer, J. R., & Jones, J. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, 93, 549-562.

Chow, G. (1960). Test of equality between sets of coefficients in two linear regressions. *Econometrica*, 28, 591-604.

Conyon, M. J., & Peck, S. I. (1998). Board control, remuneration committees, and top management compensation. *Academy of Management Journal*, 41, 146-157.

Darlington, R. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161-187.

Diaconis, P., & Efron, B. (1983, May). Computer-intensive methods in statistics. *Scientific American*, 116-130.

Efron, B. (1979). Computers and the theory of statistics: Thinking the unthinkable. *Society for Industrial and Applied Mathematics Review*, 21, 460-480.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, 72, 45-58.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81, 461-470.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

England, J. D., & Pierson, D. A. (1990). Salary ranges and merit matrices: The time targeting approach. *Compensation and Benefits Review*, 107, 25.

Finkelstein, S., & Boyd, B. K. (1998). How much does the CEO matter? The role of managerial discretion in the setting of CEO compensation. *Academy of Management Journal*, 41, 179-199.

Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, 9, 1218-1228.

Gomez-Mejia, L. R., Tosi, H., & Hinkin, T. (1987). Managerial control, performance, and executive compensation. *Academy of Management Journal*, 30, 51-70.

Guzzo, R., Jette, Q., & Katzell, R. (1985). The effects of psychologically based intervention programs on workers' productivity: A meta-analysis. *Personnel Psychology*, 38, 275-291.

Henderson, A. D., & Frederickson, J. W. (1996). Information-processing demands as a determinant of CEO compensation. *Academy of Management Journal*, 39, 575-606.

Institute for Scientific Information. (1999). *Social Science Citation Index*. Philadelphia: Author.

Kuhn, P., & Sweetman, A. (1998). Wage loss following displacement: The role of union coverage. *Industrial and Labor Relations Review*, 51, 384-400.

Lunneborg, C. E. (1985). Estimating the correlation coefficient: The bootstrap approach. *Psychological Bulletin*, 98, 209-215.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376-390.

Milovich, G. T., & Newman, J. M. (1996). *Compensation* (5th ed.). Boston: McGraw-Hill.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin*, 86, 493-522.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Newbury Park, CA: Sage.

Olds, E. G., Mattson, T. B., & Oldeh, R. E. (1956). *Notes on the use of transformations in the analysis of variance* (WADC Tech Rep. No. 59-82). Dayton, OH: Wright-Patterson Air Force Base.

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology*, 65, 569-607.
Rasmussen, J. L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin, 101*, 136-139.

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology, 77*, 336-342.

Sanders, W. G., & Carpenter, M. A. (1998). Internationalization and firm governance: The roles of CEO compensation, top team composition, and board structure. *Academy of Management Journal, 41*, 158-178.

Saunders, D. R. (1955). The “moderator variable” as a useful tool in prediction. *Proceedings of the 1954 Invitational Conference on Testing Problems*, 54-58.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209-222.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement*. Chicago: Rand McNally.

SPSS. (1996). *SYSTAT 6.0 for Windows: Statistics* [Computer software]. Chicago: Author.

Winer, B. J. (1974). *Statistical principles in experimental design*. New York: McGraw-Hill.

Craig J. Russell is the J.C. Penney Chair of Business Leadership at the Price School of Business, University of Oklahoma and holds a joint appointment as a professor of psychology. He received his Ph.D. in business administration from the University of Iowa. He pursues research in management and leadership development, personnel selection, and research methods.

Michelle A. Dean is an assistant professor in the Department of Management, College of Business Administration, University of North Texas. She received her Ph.D. in business administration from Louisiana State University. Her research focuses on performance prediction in personnel selection systems, with recent emphasis on biodata criterion and construct validity.