INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University
Microfilms
International

300 N. Zeeb Road
Ann Arbor, MI 48106

Fleischman, Lynn Ellen

AN ANALYTICAL INVESTIGATION OF THE ROBUSTNESS OF THE RESTRICTION OF RANGE CORRECTION PROCEDURE

City University of New York PH.D. 1985

AN ANALYTICAL INVESTIGATION OF THE ROBUSTNESS OF THE RESTRICTION OF RANGE CORRECTION PROCEDURE

by

LYNN ELLEN FLEISCHMAN

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1985

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

3/22/85
date

Chairman of Examining Committee

3/22/85
date

Executive Officer

Dr. Alan L. Gross, Chairman

Dr. David M. Rindskopf

Dr. Phillip Ramsey
Supervisory Committee

The City University of New York

Abstract

AN ANALYTICAL INVESTIGATION OF THE ROBUSTNESS OF
THE RESTRICTION OF RANGE CORRECTION PROCEDURE

by

LYNN ELLEN FLEISCHMAN

Adviser: Prof. Alan L. Gross

The use of test scores for selection purposes is continually under legal scrutiny. An organization must be able to substantiate the validity of the test. Often this task is complicated by the problem of missing data, i.e., whereas test scores (x) are available for all applicants, criterion measures (y) are available only for selected cases. There is currently a statistical procedure that supposedly "corrects" for restriction of range, that is, it estimates the x-y correlation for the total population. The correction for restriction of range yields a statistical estimate of the unrestricted correlation coefficient.

It is often assumed that this adjustment procedure is effective, i.e., will produce an improved estimate. However, this result is

highly dependent upon a set of strong assumptions (linearity, homoscedasticity, and selection on the predictor alone). In practice, these assumptions are often violated. Furthermore, empirical research has shown that departures from these assumptions can lead to significant errors in estimating the unrestricted population correlation.

The primary goal of this research was to investigate analytically the robustness of the restriction of range correction procedure to violations in the assumption of linearity. This analytical investigation derived expressions for the bias, standard error, and expected mean square error of the squared correlation in the selected group and the squared corrected correlation where the regression of the criterion on the predictor is both linear and nonlinear.

The findings of the present investigation suggest that the correction formula is of limited value when sampling variability and violations of the linearity assumptions are considered. Only under certain conditions is it advantageous to correct for restriction of range. These cases occur for large sample sizes, liberal selection strategies, and high x-y relationships in the total group. Recommendations for both researchers and practitioners are discussed. In addition, potential areas for future research are suggested.

ACKNOWLEDGEMENTS

It is my pleasure to acknowledge Dr. Alan L. Gross, who served as my dissertation advsior. Without his supervision and expertise, this investigation would not have been possible. His incisive thinking, careful guidance, and patience helped sustain my efforts during the preparation of this manuscript. Dr. Gross' brilliance in this area of investigation is already well documented in the literature. It was my good fortune to have trained and worked with Dr. Gross and to have known him as a fine mentor.

I would also like to thank the other members of my dissertation committee, Dr. David M. Rindskopf and Dr. Phillip Ramsey, both of whom made invaluable suggestions that have added much to this study.

CONTENTS

LIST OF FIGURES AND TABLES

Figure

Table

Chapter I

INTRODUCTION

Whenever a test score (x) is used as a basis for selection, certain applicants will be excluded. Those applicants who are not accepted will then be unavailable for future evaluation on some criterion variable (y). As a result, data are available only from selected applicants to support or refute validity claims about the relationship of y to x. The problem arises, for example, in validating the Medical College Admissions Test (MCAT) (x) as a predictor of the criterion measurement, first-year medical school grades (y). Paired x-y data are only available for selected applicants. This major and unavoidable statistical problem, known as "restriction of range," poses a dilemma for practitioners and researchers. The use of test scores for selection purposes is continually under legal scrutiny. An organization must be able to substantiate the validity of the test in question, even though there are missing data. The search for this needed evidence is hindered by the problem of restriction of range.

Given the best of all possible worlds, two approaches can, in theory, be employed in overcoming the restriction of range problem. In one method, the admissions committee selects all applicants. In the second approach, the admissions committee selects applicants in a purely random fashion. Although both of these procedures attempt to create an unrestricted range of scores, i.e., a complete (x,y) data set, it is obvious that such selection strategies are not only

unrealistic, but impractical as well. The researcher, therefore, must deal with the problem of statistically estimating test validity when y scores are not available for all examinees.

There is currently a statistical procedure that supposedly "corrects" for restriction of range, that is, it estimates the x-y correlation for the total population. The standard formula to correct for range restriction in the two-variable case is (Lord & Novick, 1968, p. 143):

$$r^2_{xy_c} = \frac{r^2_{xy_s}}{(r^2_{xy_s} + \frac{s^2_{x_s}}{s^2_{x_t}}(1 - r^2_{xy_s}))}, \qquad (1)$$

where $r^2_{xy_s}$ = observed squared correlation between x and y in the selected group; $s^2_{x_s}$ = variance of x in the selected group; $s^2_{x_t}$ = variance of x in the total group; $r^2_{xy_c}$ = squared correlation between x and y in the total group corrected for restriction of range. This correlation takes into account the actual correlation coefficient obtained from the restricted group, as well as the relationship between the variances of the restricted group scores and total group scores. The correction procedure is based on a type of extrapolation procedure where the unknown total sample correlation is inferred from various sample statistics computed in the selected group. The correlation procedure is based on a set of assumptions. The first assumption, linearity, requires that the regression of the criterion (y) on the predictor (x) is described by a linear relationship. The second assumption, homoscedasticity, requires that the conditional variance of y around the regression line be constant for all values of the predictor. The third

assumption states that selection is based on the test score alone, or, if based on x and additional variables, these additional variables are conditionally independent of y given x.

The regression, distribution, and selection assumptions are important considerations when using the restriction of range correction procedure. Previous research, most of which is empirical, has suggested that the procedure will yield accurate estimates only when these assumptions are violated to a minor extent (Brewer & Hills, 1969; Greener & Osburn, 1979, 1980; Gross, 1982; Gross & Fleischman, 1983; Linn, 1968; Linn & Dunbar, 1982; Linn, Harnisch, & Dunbar, 1981; Novick & Thayer, 1969; Roe, 1979). In practice, though, deviations from these assumptions are often noted. It has been shown that departures from these assumptions can lead to significant errors in estimating the unrestricted population correlation. For example, Gross and Fleischman (1983) simultaneously violated all three premises. Their results suggested that the correction formula is not robust with respect to violations in the assumptions.

At present, the most commonly used solution to the restriction of range problem relies on the correction formula. The use of the correction formula still needs further investigation. The bulk of the work done in this area has been empirical. Very little systematic analytical study of the accuracy of the method has been presented. The primary goal of the present research is to investigate analytically the robustness of the restriction of range correction procedure. More specifically, we analytically study the effect of the violation of the linearity assumption. The analytical investigation will yield

generalizable results, whereas an empirical study is typically more limited.

The expected value of the squared correlation coefficient in the total group $E(r^2_{xy_t})$ defines a parameter which can be taken to represent the validity of x as a predictor of y in the applicant group. If we denote an estimate of $E(r^2_{xy_t})$ as $\hat{E}(r^2_{xy_t})$, the bias of the estimate is defined as the deviation of the expected value of the estimate, i.e., $E(\hat{E}(r^2_{xy_t}))$ from the expected value $E(r^2_{xy_t})$:

$$\text{BIAS} = E(\hat{E}(r^2_{xy_t})) - E(r^2_{xy_t}). \tag{2}$$

The standard error (SE) of the estimate can be expressed in the following manner:

$$\text{SE} = [E[\hat{E}(r^2_{xy_t}) - E(\hat{E}(r^2_{xy_t}))]^2]^{\frac{1}{2}}. \tag{3}$$

The expected mean square error (EMSE) can be defined as the expected value of the squared difference between the estimate and $E(r^2_{xy_t})$:

$$\text{EMSE} = E[\hat{E}(r^2_{xy_t}) - E(r^2_{xy_t})]^2. \tag{4}$$

The bias, SE, and EMSE indices are used to measure the accuracy of $\hat{E}(r^2_{xy_t})$ as an estimate of $E(r^2_{xy_t})$. In the present research, two estimates of $E(r^2_{xy_t})$ are considered: $r^2_{xy_s}$ (the squared xy correlation in the selected group) and $r^2_{xy_c}$ (the squared corrected correlation). Expressions are derived for the bias, SE, and EMSE of each estimator for the situations where the underlying regression model is linear, and also for the case where it is nonlinear. Varying degrees of violations from the underlying linearity assumption are systematically

investigated via data sets that approximate real life situations. The behavior of $r^2_{xy_s}$ and $r^2_{xy_c}$ in such circumstances is noted as different degrees of restrictions, distributions of x scores, nonlinearity, sample sizes, and squared correlation coefficients in the total group are considered.

The study of the robustness of the restriction of range correction procedure from an analytical vantage point should result in not only a greater understanding of the process, but also a practical guide to determine under what conditions it will yield accurate estimates. This analytical procedure will yield results that are more precise than previous studies, which have basically been empirical in nature.

The relevant research is reviewed in Chapter II. In Chapter III, the method employed in the investigation is described. The results are presented in Chapter IV. Finally, Chapter V discusses and summarizes the findings.

Chapter II

REVIEW OF THE LITERATURE

This review presents and summarizes analytical and empirical studies concerning the restriction of range correction procedure. It is important to note that the analytical studies have focused on the development of statistical correction procedures, whereas the empirical research has most typically investigated the robustness of the correction procedures.

Analytical Studies

Pearson (1903) derived formulas for expressing the total population variance covariance matrix as a function of the parameters from a selected subpopulation. These formulas were constructed under the assumptions of a multivariate normal population. Using these general relationships, the correction formula can be constructed. Lawley (1934) showed that normality is a sufficient but not necessary condition for developing the correction formula. The two assumptions underlying Lawley's derivation were that the regression be linear and homoscedastic. The derived formulas are based solely on parameter values, so the use of Lawley's formulas requires data sets large enough to justify ignoring sampling errors.

Birnbaum, Paulson, and Andrews (1950) provided formulas to reconstruct the means, standard deviations, and correlation coefficients of an original population. These formulas are the same as

Lawley's. The underlying assumptions of their procedure were linearity and homoscedasticity of regression. To demonstrate the use of the formulas, they randomly obtained a sample of 942 individuals. Five variables were measured on each individual. A subgroup was then selected by using two of the variables, leaving 647 individuals, or approximately 69% of the original population. Population parameters were then "reconstructed" from the selected sample statistics. The "true" population parameters were available and were compared with the estimated values. This comparison shows "agreement," but Birnbaum et al. cautioned that the degree of this "agreement" needed further investigation, such as the standard errors of the estimated values.

Cohen (1955) considered the problem of estimating parameters of a bivariate normal population from restricted samples. He derived maximum likelihood estimators to determine this population correlation. Cohen discussed censoring and truncation as two procedures resulting in selected samples. A censored sample arises from the process of filling a fixed number of openings. Typically, applicants are ranked on the selection variable and then chosen according to a particular selection strategy. In a truncated sample, a fixed cut score distinguishes between selected and unselected applicants. It was determined that the maximum likelihood estimates of the population correlation have the same form regardless of the selection procedure employed, differing only with respect to the estimation of the variance in the selected group. For the censored case, Cohen's maximum likelihood estimator is exactly the correction formula given by equation (1).

Watterson (1959) used linear least squares estimation on censored samples to estimate population parameters. The advantages of using linear least squares estimation are that it yields unbiased estimators and is simple to calculate. The major criticism of Watterson's work is that his procedure offered unbiased estimates only for functions of the population correlation coefficient.

Bobko and Rieck (1980) drew on the original work of Moran (1970). They employed Taylor series expansions as a vehicle for computing the standard errors of functions of correlation coefficients in the linear case.

A predictive probability distribution approach employed by Gross and Perry (1983) yielded confidence interval estimates for the least squares regression weights and the residual variances in the total group based on restricted data. In addition, interval estimates of the difference between the mean in the selected group and the mean in the unselected group were presented. The usefulness of these interval estimates was demonstrated using a real data set. Selected samples were generated by varying the proportion selected from .20 to .90. The confidence intervals of interest were then computed for each censored sample. When the proportion selected was at least .50, the confidence interval estimates always contained the "true" population parameter and were reasonably narrow.

Grows and Perry (1983) also considered the problem of inferring the correlation in some "future" group based on data from a selected group. The question of interest here was not in estimating the relationship in the total group, but for a second group of applicants, who,

for example, were applying for admission. This investigation derived a confidence interval estimate for the squared correlation coefficient in the future group. The derivation, also Bayesian in nature, was based upon the computation of the predictive probability distribution for the future group. Real life data showed evidence of the applicability of this technique. Specifically, a selected group of 262 and a group of 524 future applicants were considered. It was observed that as the ratio of the standard deviation of the predictor in the future group to that in the current group decreased, the probability of the squared correlation in the future group exceeding the squared correlation in the selected group increased.

Studies of Robustness

Noting that real life data often violate both linearity and homoscedasticity assumptions, other studies investigated the effect of using the correction formula when certain assumptions were violated. One such study was done by Linn (1968), who demonstrated the effects of substituting an available explicit selection variable on this standard correction formula. Explicit selection implies that subjects are selected directly on the basis of their observed scores on a variable. Implicit selection, on the other hand, implies an indirect or incidental selection procedure. Although Linn violated the selection assumption, he assumed a linear relationship. He demonstrated the errors that arose when an available predictor variable (y) was assumed to be the explicit selection variable (x) by setting $R_{xz} = R_{yz} = .60$. The correlation between the availabe and assumed variables (ρ_{xy}) ranged from

0.0 to 1.0, and the range of the selection ratio ($\sigma_x^2/\sigma_{x_s}^2$) used was from 1.1 to 2.0. ρ_{xy} was determined from the correction formula and then compared to the *a priori* value of .60. The results suggested that the correction formula underestimated ρ_{xy} when $\rho_{xz} \geq .30$ and that it overestimated ρ_{xy} when $\rho_{xz} < .30$. As the selection ratio approached 2.0, the degree of under/overestimation increased. Although Linn only looked at certain limited situations, errors were not larger than .06 correlation units in the extreme cases presented.

Linn, Harnisch, and Dunbar (1981) empirically investigated conditions resulting in conservative corrections. They studied the combined effects on corrected estimates of violations of assumptions as well as selection on an unspecified variable. They used the results from more than 700 criterion-related validity studies. If the selection variable was unspecified, then underestimates of the correction resulted. Their findings suggested conservative corrections even though underlying assumptions were violated and that these violations tended to overcorrect in certain situations. Later results (Linn & Dunbar, 1982) discussed the "drastic" effects that selection on several variables can have on the estimate. They reported the standard deviations and intercorrelations for the criterion (grade point average (GPA)) and three predictor variables (high school grade average (HSGA) and the verbal and mathematics scores on the Scholastic Aptitude Test (SAT-V and SAT-M)). The data were based on a sample of 277 college students. The predictive validity of HSGA was .22, and the predictive validities of SAT-V and SAT-M were slightly negative (-.08 and -.02, respectively). The atypical intercorrelations among the

predictor variables were assumed to be due to selection effects.

Whereas Linn's (1968) study produced small errors, a study by Novick and Thayer (1969) suggested that errors might even be more serious. The uniqueness of Novick and Thayer's research was the large sample sizes they worked with. Real data sets of approximately 20,000 subjects made it feasible to simulate extreme selection yet remain with a large sample size in the selected group. In addition, sampling errors can be disregarded when the sample size is large. When the production correlation was relatively large ($\rho \geqslant .40$) and the selection was not extreme ($P_s \geqslant .40$), errors were all smaller than .06 correlation units. For smaller ($\rho < .40$) population correlations and more extreme selection schemes, corrections for restriction of range were unsatisfactory. Both overestimations and underestimations were noted. Novick and Thayer tried to relax the homoscedasticity assumption in an attempt to improve the accuracy of the correction. One technique discarded the assumption of constant error variance, while the other assumed that the error variances have a general linear form. These attempts did little to increase the accuracy of the estimate.

Greener and Osburn considered the accuracy of the estimate when bivariate normality cannot be assumed. In one of their studies (1980), the bias of the correction for restriction of range due to explicit selection was investigated via simulated distributions that violated either the linearity or homoscedasticity of regression assumptions. A sample size of 4,000 was used to generate nine distributions, with nine cutoff points ranging from .10 to .90, and nine values of ρ from

.10 to .90. Three types of distributions were studied: sigmoid, football, and fan-shaped. The sigmoid distribution displayed a regression line that was flattened at its tails. It was used to investigate violations of linearity. The football-shaped distribution was characterized by large conditional variances of the criterion given the predictor near the mean of the predictor and smaller variances away from the mean. The fan-shaped distribution possessed the conditional variance of the criterion given the predictor as an increasing function of the predictor. The latter two distributions were generated to study the violation of homoscedasticity. The results suggested that for moderate degrees of restriction (less than 40% unselected) the corrected correlation provided a good estimate of ρ and was better than the uncorrected correlation. As the degree of restriction increased, violations of linearity or homoscedasticity resulted in deteriorating estimates of the population correlation. This bias became "unacceptable" in the direction of overestimates of ρ in the football-shaped distribution. If Greener and Osburn's work were replicated using smaller sample sizes, generalizations of their results to more realistic situations would be possible.

In another study by Greener and Osburn (1979), the accuracy of the correction formula was examined in 13 empirical bivariate distributions that contained violations of either the linearity or homoscedasticity assumption. Their results showed that for relatively small unrestricted correlations ($.10 \leqslant \rho \leqslant .25$), the corrected correlation was not more accurate than the uncorrected correlation, for all truncations considered. The uncorrected correlation tended to

underestimate ρ, while the corrected correlation tended to overestimate ρ. This conclusion held for distributions that did not violate the linearity or homoscedasticity assumptions. It was noted that for large unrestricted correlations ($.60 \leqslant \rho \leqslant .80$), the corrected value was always a better estimate of ρ than the uncorrected statistic. As the degree of restriction increased, the error incurred by using the uncorrected estimate also increased. For moderate ρ ($.30 \leqslant \rho \leqslant .55$), their results were similar to those noted for the high unrestricted correlations where the correction formula yielded a more accurate estimate of ρ than the uncorrected correlation. Generally, the correction deteriorated as truncation increased. Finally, their data indicated that the correction was sensitive to modest departures from linearity. With respect to departures from homoscedasticity, however, the correction was robust. This finding is consistent with that of Novick and Thayer's.

Gross and Kagen (1983) demonstrated the advantage of not correcting for restriction of range in certain instances. The usual assumptions of linearity, homoscedasticity, and selection were satisfied. The expected mean square error criterion was used as a measure of accuracy. Eighty-one censored distributions were simulated from a bivariate normal distribution by considering three sample sizes for the restricted group (50, 100, and 200), three cutoff points (.30, .50, and .70), and nine values of ρ from .10 to .90. Their findings suggested that the corrected correlation was negatively biased. This bias was most notable when less than half of the sample was selected, when this selected group contained less than 30 subjects, and when ρ

was less than .50. Gross and Kagen's results also showed that even though the estimate was less negatively biased than the uncorrected correlation, the expected mean square error for the estimate was considerably greater than that for the uncorrected correlation in certain situations where ρ was less than .50 and less than 20% of the subjects were selected. In other words, even though the corrected correlation yielded a more accurate estimate than the uncorrected correlation in terms of bias, it may be less accurate in terms of mean square error.

Oftentimes, selection is based on more than one variable. An approach to deal with a multivariate selection process is to construct a single selection variable. The major flaw in this method, however, is that it is assumed that all of the selected variables are known and can be quantified. Roe (1979) described a least squares regression equation procedure for reconstructing the actual selector variable. He explained that actual selection can be different from intended selection. It is this actual selector variable, he contended, that should be used in correcting for restriction of range. The limitation of Roe's method is that by using regression weights, variability of sampling must be considered. When this variability is taken into account, the reconstructed selection variable may fail to satisfactorily represent the true selection process.

Brewer and Hills (1969) evaluated the effect of various levels of positive skew of the predictor variable. Their results suggested that the correction procedure should not be used when the predictor variable is skewed, as erroneous estimates resulted. The correction procedure should be used only when it can be assumed that the distri-

bution for the predictor variable in the unrestricted group is "nearly" symmetrical. The degree of restriction, as well as the size of the correction in the total group, determined the accuracy with which ρ could be estimated. Brewer and Hills noted that the higher the correlation was in the total group, the more accurate the estimates were even for large degrees of restriction. On the other hand, in cases where $\rho < .4$, at least half of the original group had to be included in the selected group in order for reasonable estimates to occur. It should be noted, however, that skewness was confounded with the linearity issue.

Recognizing that underlying assumptions have been violated is one consideration. It is also important to investigate how these assumptions have been violated. There is some evidence that suggests that violations can offset each other. Gross (1982) has recently shown that the correction formula will yield exact values of ρ even for nonlinear heteroscedastic relationships. He described a sufficient condition for this to occur, which is:

$$Q = (S_e/s_e)/(\beta/\beta_s) = 1, \tag{5}$$

where Q = the quantity that will provide a sufficient condition for the validity of the correction formula for a given selection procedure; S_e = standard error in the total population; s_e = standard error in the selected population; β = linear slope coefficient in the total population; and β_s = linear slope coefficient in the selected subpopulation. In other words, nonlinearity of regression can be "compensated" for by heteroscedastic variances. The implication of this finding for

practitioners is significant: the correction formula can now be applied to some data sets which violate the underlying assumptions previously thought necessary. Gross further explained how to predict the direction of bias. He noted that when $Q > 1.0$ a highly positively biased estimate will result.

Overestimation of the corrected correlation has been noted in the literature (Levin, 1972). The practical implication of the seemingly paradoxidal occurrence of an increase in correlation by restriction of range is the fact that correlations in selected groups need not be underestimates. In contrast, when $Q < 1.0$ an underestimate is to be expected. Underestimation has been the more common situation.

Gross and Fleischman (1983) studied the restriction of range corrections when both the distributions and selection assumptions were simultaneously violated. Previous research considered the accuracy of the estimate when either the linearity and homoscedasticity assumptions were violated (Greener & Osburn, 1979, 1980; Novick & Thayer, 1969), or the selection assumption was violated (Linn, 1968). They simulated selection situations by choosing subsets of data from a larger data set. They then compared the uncorrected and corrected correlation coefficients based on the selected subset to the known values for the total data set. A sample of 913 students' test score data were available with information on 19 variables for each student. Six different data sets, each consisting of 913 cases, were then created from the original data matrix. Within each data set, the criterion variable (y), the selector variable (x_1), and an additional variable that may be considered in selection process (x_2) were specified.

The correlation between x_1 and x_2 was approximately .60. The correlation between the criterion and either predictor was approximately equal in the six data sets. Five different selection procedures were generated from the six data sets. Each procedure was based on the selection of cases in terms of x_1, x_2, or both. Various proportions selected from .20 to .90 were also considered. In other words, for the eight selected proportions, the five different selection procedures were applied to the six data sets. The accuracy of the estimate was judged in terms of a measure of percentage error. Their results supported the conclusion that the correction formula is not robust with respect to simultaneous violations in the underlying distribution and selection assumptions. Their findings also suggested that when the linearity and/or homoscedasticity assumptions were violated, the additional effect of the violation of the assumption that selection is based solely on x_1 was to decrease the value of the corrected coefficient. Lastly, they noted that the corrected correlation can, but need not necessarily, produce more accurate estimates than the uncorrected correlation when both the distribution and selection assumptions were violated.

Two investigations, Forsyth (1971) and Gullickson and Hopkins (1976), attempted to provide formulas for confidence interval estimates of the population correlation. Both of these studies employed simulation techniques. Forsyth generated 24 sampling distributions by considering three sample sizes for the restricted group (25, 50, and 100), four cutoff points (.10, .25, .50, and .75), and two values of the population correlation (.80 and .50). His work

demonstrated a trend for expected values of the corrected correlations to be negatively biased estimates of the population correlations. The amount of error was independent of the sample size. As the cutoff point became more stringent, the amount of error increased. This means that the greatest errors were noted for stringent selection procedures and low population correlations. The confidence interval estimates for the population correlation calculated from restricted samples were inaccurate. In a similar fashion, Gullickson and Hopkins simulated 240 distributions by considering four sample sizes (25, 50, 100, and 200), six cutoff points (.10, .20, .40, .60, .75, and .90), and ten values of the population correlation from 0.00 to .90. They then used these distributions to construct 24 confidence interval estimates for ρ using both $\alpha = .01$ and $\alpha = .05$. Several trends were suggested from Gullickson and Hopkins' research: the width of the confidence interval estimate decreased as the sample size increased; the width of the confidence interval estimate increased as the proportion selected decreased; and the slope of the regression line decreased as the sample size and the estimate increased. They were careful to caution practitioners of the "little practical value" of the estimate for small samples with large restrictions. Gullickson and Hopkins' work also provided a test of the null hypothesis: $\rho = 0$. They suggested again, though, that no test be made for small sample sizes with large restrictions.

Linn (1983) had concluded that "Karl Pearson's selection formulas have been around fora long time, but the implications of the effects he described are still only partially understood" (p. 13).

He recommended "the use of more elaborate analytical techniques" (p. 13) as a requirement for a more comprehensive understanding of the correction procedure.

The basic results of most of these studies were that (i) the greatest errors were noted for stringent selection procedures and low population correlations; (ii) the correction formula was not robust with respect to simultaneous violations in the underlying distribution and selection assumptions; and (iii) the corrected correlation can, but need not necessarily, produce more accurate estimates than the uncorrected correlation when both distribution and selection assumptions were violated. One can summarize the work in the area by noting that the analytical studies have focused on formula derivation, while the empirical investigations have questioned the robustness of the correction procedure. The present study extends the research on the question of robustness by employing analytical rather than purely empirical methods. The advantage of investigating the robustness of the restriction of range correction procedure from an analytical vantage point is the gain in generalizability of the results. This gain can give researchers a greater understanding of the validation procedure. The forthcoming results will also have a pragmatic application, for they will enable one using a test for selection purposes to make a more clearly defined decision.

Chapter III

METHOD

Consider the situation where a sample of n (Y_i, X_i) pairs is observed. The variable X is a test and Y is a criterion variable. The sample of size n might be viewed as some applicant group, e.g., n individuals applying for admission who are tested on x and after admission observed on y. The general problem is to investigate the validity of x as a predictor of y, i.e., to investigate the x-y relationship for an applicant group of size n. It is assumed that given the X_i, the Y_i are normally and independently distributed with mean $\beta_0 + X_{1_i} + \beta_2 X_{1_i}^2$ and variance σ^2. When the parameter β_2 is zero, the relationship between x and y can be said to be linear. The case where $\beta_2 \neq 0$ defines the nonlinear case.

When no selection process is operating, one can observe the total or unselected sample of n xy scores. We can describe the relationship between x and y in this total sample by the squared Pearson product moment correlation coefficient:

$$r^2_{xy_t} = \frac{s^2_{xy_t}}{s^2_{x_t} s^2_{y_t}}, \tag{6}$$

where s_{xy_t} = the xy covariance in the total group, s_{x_t} = the standard deviation of x in the total group, and s_{y_t} = the standard deviation of y in the total group. The expected value of this squared correlation over all unselected samples of size n $[E(r^2_{xy_t})]$ defines a

parameter which can be taken to represent the validity of x as a predictor of y. More specifically, $E(r^2_{xy_t})$ measures this "average" validity of x as a (linear) predictor of y in an unselected sample of size n. Since $E(r^2_{xy_t})$ is an unknown parameter, one is interested in estimating its value. If $\hat{E}(r^2_{xy_t})$ is some estimate of $E(r^2_{xy_t})$, the accuracy of the estimate $\hat{E}(r^2_{xy_t})$ can be assessed in terms of the following criteria: bias, standard error, and expected mean square error. Specifically, bias of the estimate is defined as the deviation of the expected value of the estimate, i.e., $E(\hat{E}(r^2_{xy_t}))$ from the expected value $E(r^2_{xy_t})$. The standard error of the estimate is given by equation (3). The expected mean square error is defined as the expected value of the squared difference between the estimate and $E(r^2_{xy_t})$ (see equation (4)). The derivation of expressions for these criteria will be described later.

The question might arise as to why one is interested in estimating the linear correlation when the form of the regression might actually be curvilinear. There are situations that may arise where the exact form of the regression is unknown. In such cases, one can fit a linear model as a reasonable approximation. It should be noted that when confronted with the restriction of range problems, one simply does not have enough data to ascertain whether or not the relationship is nonlinear. Furthermore, Novick and Jackson (1974, p. 82) state that "it is conceivable that we might have a joint distribution in which the regression function is not or not known to be linear, but still demand to use a linear-prediction function on the ground that the gain in simplicity outweighs any possible loss in predictive

efficiency."

Consider next the situation where some selection process is operating; for example, not all applicants are admitted. In other words, we have all x scores, but only have some y scores, and one is no longer able to observe a total sample of n xy scores. One is still interested in estimating the linear xy relationship, i.e., $E(r^2_{xy_t})$. The solution of this estimation problem is complicated by two factors. First, one must deal with a restricted group, i.e., x is observed but is observed only for selected cases. Secondly, when $\beta_2 \neq 0$ the model is actually nonlinear and it becomes less likely that a linear relationship fitted to the selected group will accurately estimate $E(r^2_{xy_t})$. Two strategies are commonly used to estimate $E(r^2_{xy_t})$ from a restricted group. One such method is to use the so-called restriction of range procedure (see equation (1)). The second method is not to employ the correction for restriction of range, but rather simply use $r^2_{xy_s}$ as an estimate:

$$r^2_{xy_s} = \frac{s^2_{xy_s}}{s^2_{x_s} s^2_{y_s}}, \tag{7}$$

where $r^2_{xy_s}$ = observed squared correlation between x and y in the selected group, s_{xy_s} = the xy covariance in the selected group, s_{x_s} = the standard deviation of x in the selected group, and s_{y_s} = the standard deviation of y in the selected group. An important question to consider is which one of the two estimates of $E(r^2_{xy_t})$ is superior. One can consider and compare the accuracy of each estimator in terms of the bias, standard error, and expected mean square error criteria

as previously described. Further, one might consider the conditions under which one estimate may be superior. These conditions can be defined in terms of the sample size n, the amount of restriction, the distribution of x scores in the total group, and the values for β_1 and β_2.

The three squared correlations $r^2_{xy_t}$, $r^2_{xy_s}$, and $r^2_{xy_c}$ can all be viewed as functions of a random variable having either a central, noncentral, or doubly noncentral F distribution. By expanding these functions in terms of Taylor series expansions, one can approximate the mean and variance of each estimator. This technique has also been used by Bobko and Rieck (1980) in the case where $\beta_2 = 0$. The method can be described in the following general way. Suppose x is some random variable having a mean μ and a variance σ^2. Suppose g(x) is some function of x. We are interested in finding the expected value, E[g(x)], and the variance of the function, VAR[g(x)]. Both of these quantities are approximated in terms of partial Taylor series expansions as follows.

The expected value of g(x), expanded about μ, is given as

$$E[g(x)] \approx g(\mu) + \frac{g''(\mu)}{2}\sigma_x^2. \tag{8}$$

The variance of g(x), expanded about μ, is given as

$$VAR[g(x)] \approx [g'(\mu)]^2\sigma_x^2, \tag{9}$$

where $g'(\mu)$ = the first derivative of g evaluated at μ, and $g''(\mu)$ = the second derivative of g evaluated at μ.

We now apply this method in obtaining the approximations for $E(r^2_{xy_t})$, $VAR(r^2_{xy_t})$, $E(r^2_{xy_s})$, $VAR(r^2_{xy_s})$, $E(r^2_{xy_c})$, and $VAR(r^2_{xy_c})$.

$E(r^2_{xy_t})$, $VAR(r^2_{xy_t})$

The squared correlation in the total group $r^2_{xy_t}$ can be expressed as follows:

$$r^2_{xy_t} = \frac{SS_{\hat{y}}}{SS_{\hat{y}} + SS_e}, \tag{10}$$

where $SS_{\hat{y}}$ = the sum of squares of estimated y, and SS_e = the sum of squares of error. If we divide the numerator and denominator through by SS_e, and divide and multiply by $n_t - 2$, it can be represented as

$$r^2_{xy_t} = \frac{SS_{\hat{y}}/(SS_e/(n_t - 2))}{SS_{\hat{y}}/((SS_e/(n_t - 2)) + (1/(n_t - 2))}, \tag{11}$$

where n_t = sample size in the total group. If we let F be defined as

$$F = SS_{\hat{y}}/(SS_e/(n_t - 2)), \tag{12}$$

then $r^2_{xy_t}$ is expressable as a function g of the F variable,

$$g(F) = \frac{F/(n_t - 2)}{1 + F/(n_t - 2)} = \frac{FW}{1 + FW}, \tag{13}$$

where $W = 1/(n_t - 2)$. Under the assumption that the Y_i are normally and independently distributed with mean $\beta_0 + \beta_1 X_{1i} + \beta_2 X^2_{1i}$ and variance σ^2, the variable F has a doubly noncentral F distribution (Searle, 1971, p. 53) with degrees of freedom 1 and $n_t - 2$ and noncentrality parameters λ_1 and λ_2, where

$$\lambda_1 = \frac{\beta'X'(Z(Z'Z)^{-1}Z' - E/n)X\beta}{2\sigma^2}, \tag{14}$$

where X = n x 3 data matrix with a leading column of ones, a second column of x scores, and a third column containing squared x scores; Z = n x 2 submatrix of X (the first two columns); and E = n x n matrix containing ones; and

$$\lambda_2 = \frac{\beta'X'(I - Z(Z'Z)^{-1}Z')X\beta}{2\sigma^2}, \tag{15}$$

where I = n x n identity matrix. It should be noted that if $\beta_2 = 0$, then $\lambda_2 = 0$, and if $\beta_1 = \beta_2 = 0$, then $\lambda_1 = \lambda_2 = 0$.

It is useful to consider an approximation to the doubly noncentral F distribution. The F variable can be viewed as the ratio of two noncentral χ^2 random variables, each divided by its degrees of freedom:

$$F = \frac{\chi_1^2(df_1, \lambda_1)/df_1}{\chi_2^2(df_2, \lambda_2)/df_2}. \tag{16}$$

Further, each noncentral χ^2 can be approximated as a central χ^2 as follows:

$$\chi_i^2(df_i, \lambda_i) \approx c_i\chi^2(d_i), \tag{17}$$

where c_i and d_i are chosen so that $\chi_i^2(df_i, \lambda_i)$ and $c_i\chi^2(d_i)$ have the same mean and variance. The doubly noncentral F variable is then approximated as (Searle, 1971, pp. 52-53):

$$F \approx (n_t - 2)\left(\frac{c_1}{c_2}\right)\left(\frac{d_1}{d_2}\right)F(d_1,d_2) = (n_t - 2)TF(d_1,d_2), \tag{18}$$

where

$$c_1 = \frac{1+4\lambda_1}{1+2\lambda_1}, \tag{19}$$

$$c_2 = \frac{(n_t - 2)+4\lambda_2}{(n_t - 2)+2\lambda_2}, \tag{20}$$

$$d_1 = \frac{(1+2\lambda_1)^2}{1+4\lambda_1}, \tag{21}$$

$$d_2 = \frac{[(n_t - 2)+2\lambda_2]^2}{(n_t - 2)+4\lambda_2}, \tag{22}$$

where $F(d_1,d_2)$ = central F variable with degrees of freedom d_1 and d_2, λ_1 = see equation (14), and λ_2 = see equation (15).

If we use the partial Taylor series expansions (8), the expected value of the squared correlation in the total group $E(r^2_{xy_t})$ is given in general as

$$E(r^2_{xy_t}) = \frac{E(F)W}{1 + E(F)W} - \frac{W^2}{[1+E(F)W]^3}VAR(F), \tag{23}$$

where $W = 1/(n_t - 2)$. Substituting E(F) into equation (23) yields the approximation for $E(r^2_{xy_t})$:

$$E(r^2_{xy_t}) = \frac{TE(F(d_1,d_2))}{1 + TE(F(d_1,d_2))} - \frac{T^2\,VAR(F(d_1,d_2))}{(1+TE(F(d_1,d_2)))^3}, \tag{24}$$

where

$$T = \left(\frac{c_1}{c_2}\right)\left(\frac{d_1}{d_2}\right),$$

and $F(d_1,d_2)$ is an ordinary F distribution and the degrees of freedom d_1 and d_2 are given by equation (18). In evaluating expression (24), one needs both the expected value and variance of $F(d_1,d_2)$. These are given as (Searle, 1971, p. 48)

$$E(F(d_1,d_2)) = d_2/(d_2-2), \tag{25}$$

$$VAR[F(d_1,d_2)] = \frac{2d_2^2}{d_1(d_2-2)(d_2-4)}\left(\frac{d_1}{(d_2-2)}+1\right). \tag{26}$$

Using the expression for the variance given by (9), one can approximate the doubly noncentral F distribution as an ordinary F distribution. The expression for the variance of $r^2_{xy_t}$ is given as

$$VAR(r^2_{xy_t}) = VAR[g(f)] = \frac{W^2}{(1+E(F)W)^4}VAR(F), \tag{27}$$

where VAR(F) is given by (26).

One observes that the expressions given by (24) and (27) can be used for both the linear and the nonlinear cases.

$E(r^2_{xy_s})$, $VAR(r^2_{xy_s})$

The expressions $E(r^2_{xy_s})$ and $VAR(r^2_{xy_s})$ are identical to those of $E(r^2_{xy_t})$ and $VAR(r^2_{xy_t})$, respectively, with the exception that n_t becomes n_s, representing the sample size in the selected group.

$E(r^2_{xy_c})$, $VAR(r^2_{xy_c})$

The squared correlation in the selected group corrected for restriction of range $r^2_{xy_c}$ can also be expressed as a function of

a variable F having a doubly noncentral F distribution. However, the function is no longer given by (13). The squared correlation coefficient in the selected group corrected for restriction of range $r^2_{xy_c}$ is given by (1).

If we again define F as follows,

$$F = \frac{SS_{\hat{y}}}{SS_{e_s}/(n_s - 2)}, \tag{28}$$

where the subscript s denotes the selected group, then

$$r^2_{xy_c} = h(F) = \frac{\frac{FA}{1+FA}}{(\frac{FA}{1+FA} + RV(1 - (\frac{FA}{1+FA})))} = \frac{FA}{RV + FA} \tag{29}$$

where

$$A = 1/(n_s - 2).$$

Using the Taylor series expansions, the expected value of the squared correlation coefficient corrected for restriction of range is approximated as

$$\begin{aligned} E(r^2_{xy_c}) &= \frac{AE(F)}{E(F)A + RV} - \frac{2RVA^2}{(AE(F) + RV)^3}VAR(F) \\ &= \frac{TE(F(d_1,d_2))}{TE(F(d_1,d_2)) + RV} - \frac{(RV)T^2VAR(F(d_1,d_2))}{(TE(F(d_1,d_2)) + (RV))^3}. \end{aligned} \tag{30}$$

The expression for the variance of $r^2_{xy_c}$ is given as

$$VAR(r^2_{xy_c}) = VAR[h(F)] = \frac{RV^2A^2}{(E(F)A+RV)^4}VAR(F). \tag{31}$$

Again, the value for $E(F(d_1,d_2))$ and $VAR(F(d_1,d_2))$ are given by (25) and (26), respectively.

It is important to again note that the expressions for the expected values and variances given by equations (24), (27), (30), and (31) are based upon two approximations. First, the doubly noncentral F variable appearing in the expressions for $r^2_{xy_t}$, $r^2_{xy_s}$, and $r^2_{xy_c}$ is approximated by a central F variable. Secondly, the truncated Taylor series expansion is used in obtaining approximations for the first two moments of the distributions for $r^2_{xy_t}$, $r^2_{xy_s}$, and $r^2_{xy_c}$. We now consider some partial checks on the accuracy of these approximations.

A partial check of the use of a truncated Taylor series expansion can be obtained by considering the accuracy of equations (24) and (27) for computing the mean and variance of $r^2_{xy_t}$ in the special case where no relationship between x and y is present, i.e., $\beta_1 = \beta_2 = 0$ and thus $\lambda_1 = \lambda_2 = 0$. In this case, the exact distribution of $r^2_{xy_t}$ can be obtained. The distribution follows a beta distribution with parameters $a = .5$ and $b = (n_t - 2)/2$ (Graybill, 1961, p. 79). In this case, where $r^2_{xy_t}$ follows a simple beta distribution, the expected value and variance for $r^2_{xy_t}$ can be expressed as

$$E(r^2_{xy_t}) = 1/(n_t - 1)$$

$$VAR(r^2_{xy_t}) = ((n_t - 2)/2)/(n_t^3 - n_t^2 - n_t + 1).$$

These exact values can be compared for different sample sizes to the approximated values obtained from the Taylor series expansion given by equations (24) and (27). These results are presented in Table 1. For the special case when $\beta_1 = \beta_2 = 0$, the results in Table 1 show that

the Taylor series expansion, i.e., equations (24) and (27) yield good approximations. Although the case where $\beta_1 \neq 0$ and $\beta_2 \neq 0$ were not considered (the exact distributions in this case is no longer a simple beta distribution), these results suggest that the use of a truncated Taylor series is an appropriate method.

Table 1

The Accuracy of the Taylor Series Approximation for $E(r^2_{xy_t})$ and $VAR(r^2_{xy_t})$

Sample size	$E(r^2_{xy_t})$		$VAR(r^2_{xy_t})$	
	Exact	Approximation	Exact	Approximation
50	.0204	.0203	.0007839	.0092640
100	.0101	.0101	.0001979	.0002148
250	.0040	.0040	.0000318	.0000330

A check on the accuracy of the central F approximation to the doubly noncentral F distribution was also investigated. This distribution is the ratio of two independent noncentral chi-square variates with degrees of freedom d_1 and d_2 and noncentrality parameters λ_1 and λ_2. The accuracy of the approximation can be studied by comparing exact percentile points for a doubly noncentral F distribution (Bulgren, 1971, p. 184) to percentile points obtained from the approximation.

Consider a doubly noncentral F variable with degrees of freedom df_1 and df_2 and noncentrality parameters λ_1 and λ_2. We denote this

variable as F". Following the procedures described by equations (16)-(22), this variable is approximated by a central F variable as follows:

$$\hat{F}'' = \frac{df_2}{df_1}\frac{c_1 d_1}{c_2 d_2} F(d_1, d_2), \tag{32}$$

where

$$c_1 = \frac{df_1 + 4\lambda_1}{df_1 + 2\lambda_1},$$

$$c_2 = \frac{df_2 + 4\lambda_2}{df_2 + 2\lambda_2},$$

$$d_1 = \frac{(df_1 + 2\lambda_1)^2}{df_1 + 4\lambda_1},$$

$$d_2 = \frac{(df_2 + 2\lambda_2)^2}{df_2 + 4\lambda_2}.$$

Thus, if f denotes the $1-\alpha$ percentile point for $F''(df_1, df_2, \lambda_1, \lambda_2)$, the accuracy of the central F approximation can be considered by computing the value

$$f/k,$$

where

$$k = \frac{df_2}{df_1}\frac{c_1 d_1}{c_2 d_2}$$

and obtaining the percentile value for this point for a central F variable with d_1 and d_2 degrees of freedom.

In Table 2 the accuracy of this approximation is studied for different cases obtained by varying the degrees of freedom (df_1, df_2) and the noncentrality parameters (λ_1, λ_2). An inspection of Table 2

shows that the approximation given by equation (32) to the doubly noncentral F distribution yields values that are reasonably close to the exact values. It should be noted that this techique is a rather stringent check on the accuracy of the approximation to the doubly noncentral F distribution. This is because the method is actually comparing one entire distribution to another, i.e., central F to the doubly noncentral F; while the present study solely uses the mean and variance of the central F distribution as approximations to the mean and variance of the doubly noncentral F distribution.

Table 2

The Accuracy of the Central F Approximation to the Doubly Noncentral F Distribution

df_1	df_2	λ_1	λ_2	f	$1-\alpha$ (exact)	$1-\alpha$ (approximate)
2	4	1.5	1.5	6.94	.93	.94
2	4	3.0	3.0	6.94	.93	.95
2	4	1.5	3.0	6.94	.96	.98
2	15	1.5	3.0	3.68	.89	.84
2	4	6.0	3.0	6.94	.87	.84

Data Analysis

A double precision FORTRAN program was written to calculate (a) the expectations $E(r^2_{xy_t})$, $E(r^2_{xy_s})$, and $E(r^2_{xy_c})$; (b) the variances $VAR(r^2_{xy_t})$, $VAR(r^2_{xy_s})$, and $VAR(r^2_{xy_c})$; and (c) the respective mean square errors $EMSE(r^2_{xy_c})$ and $EMSE(r^2_{xy_s})$. The values of the ana-

lytically derived expected values, variances, and mean square errors were then systematically investigated across various conditions.

Distribution of x Scores

Four distributions of x scores were investigated where $-3 \leqslant x \leqslant 3$. The first distribution was normal and the last three followed beta distributions:

1. Normal distribution, where $E(X) = 0$, $V(X) = 1$, skewness $(S) = 0$, and kurtosis $(K) = 3$. The following two expressions were used to calculate the skewness (S) and kurtosis (K) of the three beta distributions having parameters p and q (Johnson & Kotz, 1970, p. 40):

$$S = 2(q-p)\sqrt{p^{-1}+q^{-1}+(pq)^{-1}} \cdot (p+q+2)^{-1} \tag{33}$$

$$K = 3(p+q+1)\{2(p+q)^2+pq(p+q-6)\} \times [pq(p+q+2)(p+q+3)]^{-1}. \tag{34}$$

2. Right-skewed distribution $(p = 5, q = 10)$, where $E(X) = 1$, $V(X) = .5$, $S = 71.8751$, and $K = 1867.7646$.

3. Left-skewed distribution $(p = 10, q = 5)$, where $E(X) = 1$, $V(X) = .5$, $S = -71.8751$, and $K = 1867.7646$.

4. Uniform distribution $(p = q = 1)$, where $E(X) = 0$, $V(X) = 3$, $S(X) = 0$, and $K = 1.8$.

In the normal case, the x scores for a given data set were obtained by using the IBM Scientific Subroutine Package (IBM Technical Publications Department, 1968) routine GAUSS, which was used to generate a normally distributed variable (x), where $E(X) = 0$ and $V(X) = 1$. To obtain a sample of x scores which followed a beta distribution with

parameters p and q, the following procedure was employed:

1. Using GAUSS, generate 2p and 2q independent standard normal variables ($x_1, x_2, \ldots x_{2p}; x_{2p+1}, x_{2p+2}, \ldots, x_{2p+2q}$).

2. Construct

$$T_1 = \sum_{i=1}^{2p} X_i^2, \quad T_2 = \sum_{i=2p+1}^{2p+2q} X_i^2, \quad B = T_1/(T_1 + T_2).$$

The variable B follows a beta distribution with parameters p and q.

3. To obtain an x variable on the interval (3,-3), the following transformation is applied: $6B - 3$.

Sample Size

A sample size of $n_t = 50, 100$, and 250 was chosen.

Regression

Three regression models ($E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2$) were considered where $-3 \leq X \leq 3$ (see Figure 1):

1. Linear regression: $\beta_0 = 3$, $\beta_1 = 1$, $\beta_2 = 0$.

2. Nonlinear regression, convex: $\beta_0 = 1.5$, $\beta_1 = 1$, $\beta_2 = .166$.

3. Nonlinear regression, concave: $\beta_0 = 4.5$, $\beta_1 = 1$, $\beta_2 = -.166$.

Strength of the Relationship

The expected value of the squared correlation coefficient in the total group $E(r^2_{xy_t})$ was set at the following values: .1, .3, and .5.

Restriction

The proportion selected P_s varied as follows: .25, .50, and .75. The selection procedure consisted of selecting those cases highest on x.

Equations:

1. $E(Y|X) = 4.5 + X - .166X^2$
2. $E(Y|X) = 3.0 + X$
3. $E(Y|X) = 1.5 + X + .166X^2$

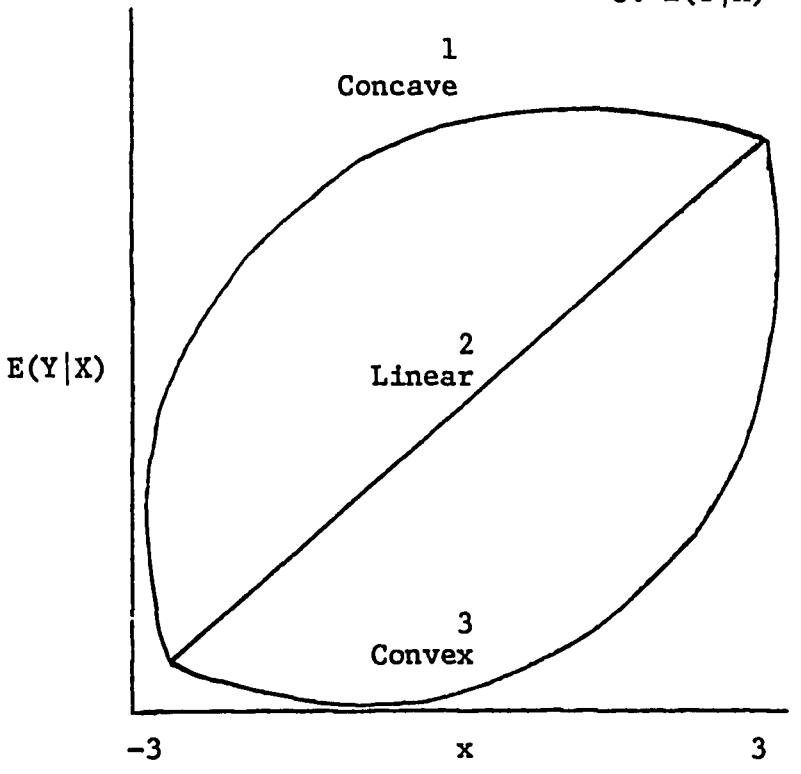

Figure 1. Schematic representation of the shape of the regression curves.

Given a specification for the values of $E(r^2_{xy_t})$, β_0, β_1, β_2, n_t, and the X data matrix, a σ^2 was chosen so as to assure a desired value (V) for $E(r^2_{xy_t})$. A second FORTRAN program was written for this purpose. The solution for σ^2 involves finding the roots of the nonlinear equation

$$E((r^2_{xy_t})|X,\beta_0,\beta_1,\beta_2,n_t;\sigma^2) - V = 0,$$

where x, β_0, β_1, β_2, and n_t are all fixed.

The following specific questions were considered:

1. How will the estimates of $E(r^2_{xy_t})$ be affected by the distribution of x scores?

2. How will the sample size affect the estimates of $E(r^2_{xy_t})$?

3. Will the estimates of $E(r^2_{xy_t})$ be affected by the violation of the linearity assumption?

4. How will the estimates of $E(r^2_{xy_t})$ be affected by the value prespecified for $E(r^2_{xy_t})$?

5. How will the degree of restriction affect the estimates of $E(r^2_{xy_t})$?

6. Is there an interaction effect between the degree of nonlinearity and the form of the distribution of x scores?

Chapter IV

RESULTS

In Tables 1-5 the individual or overall effect of each of the five independent variables (distribution of x scores, number of subjects, strength of the relationship in the total group [$E(r^2_{xy_t})$], form of the regression, and proportion selected) are presented with respect to the bias, variance, and EMSE criteria. The bias for each estimator is presented. For the variance and EMSE criteria, differences are presented, i.e.,

$$VAR(r^2_{xy_s}) - VAR(r^2_{xy_c}) \text{ and } EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c}).$$

Positive differences therefore indicate that $r^2_{xy_c}$ is the better estimate, while negative values show that $r^2_{xy_s}$ is the favored estimator.

One can draw different conclusions of the superiority of one estimator over the other depending on the criterion used to assess accuracy. For example, in the majority of cases, $r^2_{xy_c}$ has a smaller bias. However, $r^2_{xy_s}$ always has the smaller variance. It should be noted that of the three criteria of accuracy it can be argued that the EMSE is the most meaningful criterion as it takes into account both the bias of an estimator as well as its variance. The EMSE criterion, then, as an overall measure of the closeness of an estimator to the parameter, is the most meaningful to interpret. Therefore, in considering Tables 3 to 7, the EMSE criterion will be discussed.

Table 3

A Comparison of Selected and Corrected Coefficients as a Function of the Distribution of x Scores[a]

Criterion	Distribution of x scores			
	Normal	Uniform	Left-skewed	Right-skewed
BIAS($r^2_{xy_s}$)	-.1003	-.0630	-.1651	-.1867
BIAS($r^2_{xy_c}$)	.0543	.1335	-.0073	-.0219
VAR($r^2_{xy_s}$) - VAR($r^2_{xy_c}$)	-.0157	-.0279	-.0251	-.0268
EMSE($r^2_{xy_s}$) - EMSE($r^2_{xy_c}$)	-.0166	-.0537	.0001	.0090

[a]Values presented are averaged over 81 different conditions.

An inspection of Table 3 shows that $r^2_{xy_s}$ is the better estimator for the symmetric distributions (normal, uniform). For the skewed distributions, there is virtually no difference between the two estimators. In Table 4, the selected coefficient $r^2_{xy_s}$ is favored for small sample

Table 4

A Comparison of Selected and Corrected Coefficients as a Function of the Number of Subjects[a]

Criterion	Number of subjects		
	50	100	250
BIAS($r^2_{xy_s}$)	-.1171	-.1284	-.1408
BIAS($r^2_{xy_c}$)	.0366	.0423	.0400
VAR($r^2_{xy_s}$) - VAR($r^2_{xy_c}$)	-.0438	-.0221	-.0057
EMSE($r^2_{xy_s}$) - EMSE($r^2_{xy_c}$)	-.0379	-.0128	.0049

[a]Values presented are averaged over 108 different conditions.

Table 5

A Comparison of Selected and Corrected Coefficients as a Function of the Strength of the Relationship in the Total Group [$E(r^2_{xy_t})$][a]

Criterion	Strength of relationship [$E(r^2_{xy_t})$]		
	.1	.3	.5
BIAS($r^2_{xy_s}$)	.0158	-.1324	-.2697
BIAS($r^2_{xy_c}$)	.1351	.0336	-.0498
VAR($r^2_{xy_s}$) - VAR($r^2_{xy_c}$)	-.0287	-.0239	-.0191
EMSE($r^2_{xy_s}$) - EMSE($r^2_{xy_c}$)	-.0663	-.0190	.0395

[a]Values presented are averaged over 108 different conditions.

sample sizes ($n \leq 100$). When the sample size increases ($n = 250$), no difference is noted between the coefficients. Table 5 shows that the use of $r^2_{xy_s}$ is advantageous when the strength of the relationship in

Table 6

A Comparison of Selected and Corrected Coefficients as a Function of the Form of the Regression[a]

Criterion	Form of the regression		
	Linear	Convex	Concave
BIAS($r^2_{xy_s}$)	-.1300	-.0597	-.1966
BIAS($r^2_{xy_c}$)	.0485	.1540	-.0835
VAR($r^2_{xy_s}$) - VAR($r^2_{xy_c}$)	-.0241	-.0185	-.0290
EMSE($r^2_{xy_s}$) - EMSE($r^2_{xy_c}$)	-.0112	-.0366	.0020

[a]Values presented are averaged over 108 different conditions.

the total group is weak [$E(r^2_{xy_t}) \leq .3$]. As the relationships in the total group strengthens [$E(r^2_{xy_t}) = .5$], $r^2_{xy_c}$ becomes the better estimator. When the linearity assumption of the restriction of range correction procedure is met (Table 6), $r^2_{xy_s}$ is slightly preferred. When this assumption is violated, $r^2_{xy_s}$ is favored for the convex regression, but not for the concave regression, where no difference in the coefficients is noted. Finally, Table 7 considers the coefficients as a function of the proportion selected. It is noted that when stringent selection schemes are used ($P_s \leq .50$), $r^2_{xy_s}$ is the superior estimator. When the proportion selected increases ($P_s = .75$), $r^2_{xy_c}$ is slightly more advantageous.

Table 7

A Comparison of Selected and Corrected Coefficients as a Function of the Proportion Selected[a]

	Proportion selected		
	.25	.50	.75
$BIAS(r^2_{xy_s})$	-.2034	-.0705	-.1124
$BIAS(r^2_{xy_c})$	-.0082	.1307	-.0036
$VAR(r^2_{xy_s}) - VAR(r^2_{xy_c})$	-.0639	-.0068	-.0010
$EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$	-.0215	-.0385	.0141

[a]Values presented are averaged over 108 different conditions.

It was of particular interest to investigate the interactions between the form of the distribution of x scores, the number of subjects, and the proportion selected (Table 8), because these are three

Table 8

A Comparison of Selected and Corrected Coefficients in Terms of $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$ as a Function of the Distribution of x Scores, the Sample Size, and the Proportion Selected[a]

Distribution of x scores	Proportion selected								
	.25			.50			.75		
	Number of subjects			Number of subjects			Number of subjects		
	50	100	250	50	100	250	50	100	250
Normal	-.0699	-.0044	.0239	-.0496	-.0418	-.0341	.0072	.0096	.0100
Uniform	-.0736	-.0579	.0114	-.1223	-.1309	-.1312	.0092	.0056	.0066
Right-skewed	-.0903	-.0101	.0459	-.0130	.0087	.0075	.0015	.0186	.0223
Left-skewed	-.0725	-.0046	.0444	-.0126	.0279	.0297	.0209	.0258	.0221

[a]Values presented are averaged over 9 different conditions.

variables that can be observed in a real life study. To reduce complexity, only EMSE values will be presented (Tables 8-12). An inspection of Table 8 shows that an interaction is present. Specifically, when the proportion selected is stringent ($P_S = .25$), $r^2_{xy_c}$ is advantageous only when the sample size is large ($n = 250$). This is true regardless of the distribution of x scores. As the proportion selected increases ($P_S = .50$), $r^2_{xy_s}$ is favored for the symmetric distributions regardless of the sample size. For the skewed distributions, $r^2_{xy_c}$ becomes favored as sample size increases. When the proportion selected in liberal ($P_S = .75$), the correction procedure is always favored, especially for the skewed distributions and less so for the symmetric distributions.

Table 9

A Comparison of Selected and Corrected Coefficients in Terms of $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$ as a Function of the Distribution of x Scores and the Form of the Regression[a]

Distribution of x scores	Form of the regression		
	Linear	Convex	Concave
Normal	-.0133	-.0458	.0095
Uniform	-.0579	-.0861	-.0171
Right-skewed	.0095	-.0145	.0053
Left-skewed	.0169	-.0001	.0102

[a]Values presented are averaged over 27 different conditions.

It was also of interest to investigate the interaction between the form of the distribution of x scores and the form of the regression (Table 9). An inspection of Table 9 shows that there is an

interaction. For the linear and convex cases, $r^2_{xy_s}$ is superior for the symmetric distributions, while no major difference between the estimates is noted for the skewed distribution types. However, for the concave distribution, major differences between $r^2_{xy_s}$ and $r^2_{xy_c}$ were absent for all distributions considered.

The interaction between the distribution of x scores and the regression form is further investigated by considering this interaction at various $E(r^2_{xy_t})$ values (Table 10). In Table 10, a three-factor interaction is observed. When the x-y relationship is weak [$E(r^2_{xy_t}) = .1$], $r^2_{xy_s}$ is the better estimator regardless of the form of the x score distribution and the regression function. When the x-y relationship is moderate [$E(r^2_{xy_t}) = .3$], $r^2_{xy_s}$ is superior only for symmetric x distributions, and there is basically no difference between $r^2_{xy_s}$ and $r^2_{xy_c}$ for the skewed distributions. However, when the strength of the relationship in the total group is high [$E(r^2_{xy_t}) = .5$], it is generally advantageous to correct for restriction of range. This preference for $r^2_{xy_c}$ is most notable for the skewed distributions.

The interaction between the x distributions and regression forms is also investigated at different sample size values. The results are presented in Table 11 and parallel those of Table 10. For a small sample size ($n = 50$), $r^2_{xy_s}$ is the better estimator regardless of the form of the x score distribution and the regression function. When the sample size increases ($n = 100$), $r^2_{xy_s}$ is superior only for the symmetric x distributions, and there is basically no difference between $r^2_{xy_s}$ and $r^2_{xy_c}$ for the skewed distributions. However, when the sample size is large ($n = 250$), it is generally better to use the correction formula.

Table 10

A Comparison of Selected and Corrected Coefficients in Terms of $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$ as a Function of the Distribution of x Scores, the Form of the Regression, and the Strength of the Relationship $[E(r^2_{xy_t})]$[a]

Distribution of x scores	Strength of the relationship								
	.1			.3			.5		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	-.0683	-.0896	-.0369	-.0113	-.0449	.0070	.0396	-.0031	.0585
Uniform	-.1462	-.1135	-.0673	-.0593	-.1024	-.0146	.0318	-.0423	.0307
Right-skewed	-.0487	-.0730	-.0276	.0058	-.0191	-.0003	.0713	.0488	.0440
Left-skewed	-.0373	-.0544	-.0329	.0083	-.0026	.0059	.0799	.0566	.0576

[a]Values presented are averaged over 9 different conditions.

Table 11

A Comparison of Selected and Corrected Coefficients in Terms of $EMSE(r^2_{xy_S}) - EMSE(r^2_{xy_C})$ as a Function of the Distribution of x Scores, the Form of the Regression, and the Number of Subjects[a]

Distribution of x scores	Number of subjects								
	50			100			250		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	-.0323	-.0593	-.0207	-.0102	-.0441	.0178	.0025	-.0342	.0315
Uniform	-.0605	-.0898	-.0363	-.0683	-.0908	-.0242	-.0449	-.0776	.0093
Right-skewed	-.0245	-.0353	-.0320	.0144	-.0062	.0091	.0386	-.0019	.0390
Left-skewed	-.0152	-.0258	-.0233	.0251	.0078	.0162	.0409	.0176	.0376

[a]Values presented are averaged over 9 different conditions.

Table 12

A Comparison of Selected and Corrected Coefficients in Terms of $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$ as a Function of the Strength of the Relationship $[E(r^2_{xy_t})]$ and the Proportion Selected[a]

$E(r^2_{xy_t})$	Proportion selected		
	.25	.50	.75
.1	-.0815	-.1169	-.0006
.3	-.0317	-.0379	.0127
.5	.0487	.0394	.0303

[a]Values presented are averaged over 36 different conditions.

It should be noted that more detailed results are presented in the three appendices. Appendix A is an exhaustive comparison of selected and corrected coefficients in terms of $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$ as a function of all five independent variables simultaneously, i.e., the distribution of x scores, the number of subjects, the strength of the relationship in the total group $[E(r^2_{xy_t})]$, the form of the regression, and the proportion selected. Appendix B gives comparisons of selected and corrected coefficients in terms of $BIAS(r^2_{xy_s})$, $BIAS(r^2_{xy_c})$, and $VAR(r^2_{xy_s}) - VAR(r^2_{xy_c})$ that correspond with the functions investigated in Tables 8-12. Appendix C presents the results of a five-factor factorial ANOVA. Appendix C was used to determine if any higher order effects other than those investigated in Tables 8-11 were present. The size of various effects can be considered in terms of the mean square error since there are zero degress of freedom for the within cell sum of squares.

This analysis showed that the effects presented in Tables 8-11 explained most of the variance within the cells of the design. Table 12, for example, is included purely for theoretical speculation, as it was not of *a priori* interest.

An interaction is observed between the strength of the relationship [$E(r^2_{xy_t})$] and the proportion selected (Table 12). The strength of the relationship affects the estimate more when the proportion selected is stringent/moderate ($P_s \leq .50$) than for liberal selection ($P_s = .75$). Specifically, when a weak/moderate relationship is observed [$E(r^2_{xy_t}) \leq .3$], $r^2_{xy_s}$ is superior for stringent and moderate selection schemes ($P_s \leq .50$), while no substantial difference is noted between the estimates when $P_s = .75$.

Chapter V

SUMMARY AND DISCUSSION

The restriction of range correction procedure is often suggested as a means for estimating the validity of a test when data are missing on the y variable. Theoretically, this adjustment procedure is effective, i.e., should produce an improved estimate (relative to $r^2_{xy_s}$) when certain underlying assumptions (linearity, homoscedasticity, and selection on x alone) have been met. In practice, though, deviations from these assumptions are often noted. Furthermore, empirical research has shown that departures from these assumptions can lead to significant errors in estimating the unrestricted population correlation. The primary goal of this research was to investigate analytically the robustness of the restriction of range correction procedure to violations in the assumption of linearity. This analytical investigation derived expressions for the bias, SE, and EMSE of $r^2_{xy_s}$ and $r^2_{xy_c}$ where the regression of y on x is both linear and nonlinear.

One of the main questions under investigation in this study is whether or not the correction procedure is robust to nonstandard situations, i.e., nonlinear regressions and/or x scores that follow various nonnormal forms.

It is observed that the correction procedure is affected by the linearity assumption. Specifically, when the linearity assumption is violated, the restriction of range correction procedure has a larger EMSE than $r^2_{xy_s}$ when the regression function is convex. However, for

the concave regression function there is no practical difference between the two coefficients. When the linearity assumption is met, $r^2_{xy_s}$ is slightly preferred (Table 6).

It can also be noted that the correction procedure is affected by the form of the distribution of x scores (Table 3). Overall, the correction procedure yields a higher EMSE when the form of the distribution of x scores is uniform. In the case of the skewed distributions, however, no substantive difference between the two estimators is seen. When the form of the distribution of x scores is "standard," i.e., normal, $r^2_{xy_s}$ is somewhat favored.

When both the distribution of x scores and the form of the regression are jointly considered (Table 9), a somewhat different finding is observed: $r^2_{xy_s}$ is favored for the convex regression function only when the x distribution is symmetric. A similar finding is noted for the linear case. No practical difference between the coefficients is noted for the concave regression. It should be noted that it is the concave regression function that shows up more in real life studies. For example, if x represents the Medical College Admissions Test (MCAT) as a predictor of the criterion measurement, first-year medical school grades (y); then we will very often observe a regression form between x and y that increases up to a certain point and then tends to level off. This particular x-y relationship is characteristic of the concave regression function. If we accept the premise that the linear and convex regression functions do not appear as frequently as does the concave form, then the use of the correction procedure, from a practical viewpoint, is not suggested.

A further examination considered the interaction between the form of the distribution of x scores and the regression form for various $E(r^2_{xy_t})$ (Table 10). It is generally noted that the correction procedure is recommended when the strength of the relationship is high ($E(r^2_{xy_t}) = .5$). Table 11 considers the interaction between the form of the distribution of x scores and the regression form for different sample size values. The correction procedure is advised when the sample size is large ($n = 250$).

Another important question to consider is how the correction procedure varies as a function of the sample size, the strength of the relationship, and the proportion selected. One of the major and general findings of this investigation is that there are a wide set of circumstances (defined in terms of these three variables) where it is not worthwhile to use the correction procedure, i.e., it is advantageous not to correct but rather to use $r^2_{xy_s}$ as an estimate. In general, these situations are described by small sample sizes ($n \leq 100$), low $E(r^2_{xy_t})$ values ($E(r^2_{xy_t}) \leq .3$), and for stringent selection schemes ($P_s \leq .50$). Only under certain conditions is it advantageous to correct for restriction of range. These cases occur for large sample sizes ($n = 250$), high $E(r^2_{xy_t})$ values ($E(r^2_{xy_t} = .5$), and liberal selection strategies ($P_s = .75$). The remaining cases do not show a notable difference between the estimators.

The results presented in Table 8 should be of particular interest to practitioners who often have immediate knowledge of the distribution of x scores, number of subjects, and the proportion selected; but not the regression function or the strength of the x-y relationship in the

total group. As a guide for practitioners, one can describe three different situations in terms of the observed distribution of x scores, the observed number of subjects, and the observed proportion selected:

1. For stringent selection ($P_S = .25$), it is advantageous to correct for restriction of range, only when the sample size is large ($n = 250$). This recommendation holds regardless of the x score distribution.

2. For moderate degrees of selection ($P_S = .50$), it is appropriate to correct for restriction of range when the sample size is moderate/large ($n \geqslant 100$), but only if the x score distribution is skewed.

3. For liberal selection schemes ($P_S = .75$), the correction procedure is favored. This applies to all sample sizes and x score distributions, especially the skewed distributions.

These results can be understood in terms of the effective sample size, i.e., the number of cases for whom both x and y are observed. When this effective sample size is small, $r^2_{xy_c}$ has a much larger sampling variance than $r^2_{xy_s}$, and thus a larger EMSE value. Consequently, $r^2_{xy_c}$ is favored only when the effective sample size is sufficiently large, i.e., when the sampling variance is reduced.

The specifics of the above recommendations are displayed in Table 13. If a competitive institution, for example, is interested in validating the test it uses for admission and the following information is known:

1. left-skewed distribution of x scores,

2. n = 50,

3. proportion selected = .25,

then it is advised that the correction procedure not be employed. On the other hand, if a less competitive institution is concerned with test validation and observes the following:

1. right-skewed distribution of x scores,

2. n = 250,

3. proportion selected = .75,

then the use of the correction procedure is suggested.

Table 13

Cases Where $r^2_{xy_c}$ Is Preferred (C) or $r^2_{xy_s}$ Is Preferred (S)[a]

Distribution of x scores	Proportion selected								
	.25			.50			.75		
	Number of subjects			Number of subjects			Number of subjects		
	50	100	250	50	100	250	50	100	250
Normal	S	S	C	S	S	S	C	C	C
Uniform	S	S	C	S	S	S	C	C	C
Right-skewed	S	S	C	S	C	C	C	C	C
Left-skewed	S	S	C	S	C	C	C	C	C

[a]Cases presented are averaged over 9 different conditions

In our opinion, the results of the present study can be applied to the field of testing. Specifically, the use of test scores for selection purposes has recently been examined from a legal point of

view. An institution must be able to substantiate the validity of the test in question, even though there is missing information. As a result of restriction of range, the validity coefficient is typically a deflated one. In the past, the correction formula was used as a way of improving the estimate of the correlation in the total group. The theoretical justification for this procedure is based on a set of strong assumptions. The findings of the present investigation suggest that the correction formula is of limited value when sampling variability and violations of the linearity assumptions are considered.

It should be noted that several questions are not completely answered by the present study. A limitation of the applicability of these results is due to the fact that the x variable was treated as fixed, i.e., $E(r^2_{xy_t})$ is considered over all samples as having the same x score distribution. Although it is of value to consider the case where x is a random variable, the treatment of the case presents some rather complex mathematical difficulties. The generalizability of this research is further constrained by the limited number of levels of the independent variables. Perhaps future research will consider a wider class of values. In addition, the two other assumptions underlying the correction procedure, homoscedasticity and selection on x alone, were not under consideration in this investigation. It would also be of interest to study the effects of the violations of these assumptions on the correction procedure.

Appendix A

AN EXHAUSTIVE COMPARISON OF $r^2_{xy_s}$ AND $r^2_{xy_c}$ AS A FUNCTION OF IDIST, NSUBJ, ER2T, IBETA, AND IPS†

---

†IDIST = the distribution of x scores
1 = normal distribution
2 = uniform distribution
3 = right-skewed distribution
4 = left-skewed distribution
NSUBJ = the number of subjects
ER2T = the expected value of the squared correlation in the total group [$E(r^2_{xy_t})$]
IBETA = the regression curve
1 = linear regression
2 = convex regression
3 = concave regression
IPS = the proportion selected

Variable $EMSE(r^2_{xy_s}) - EMSE(r^2_{xy_c})$	Code	Sum	$\bar{X}$	SD	Vari-ance	n
a	b	c	d	e	f	g
Entire population		-.49502	-0.0153	0.0790	0.0062	324
IDIST	1.00	-1.3407	-0.0166	0.0651	0.0042	81
NSUBJ	50.00	-1.0103	-0.0374	0.0695	0.0048	27
ER2T	0.1000	-0.8273	-0.0919	0.0776	0.0060	9
IBETA	1.00	-0.2514	-0.0838	0.0717	0.0051	3
IPS	0.25	-0.1331	-0.1331	0.0000	0.0000	1
IPS	0.50	-0.1167	-0.1167	0.0000	0.0000	1
IPS	0.75	-0.0016	-0.0016	0.0000	0.0000	1
IBETA	2.00	-0.3568	-0.1189	0.1054	0.0111	3
IPS	0.25	-0.1340	-0.1340	0.0000	0.0000	1
IPS	0.50	-0.2160	-0.2160	0.0000	0.0000	1
IPS	0.75	-0.0068	-0.0068	0.0000	0.0000	1
IBETA	3.00	-0.2191	-0.0730	0.0781	0.0061	3
IPS	0.25	-0.1559	-0.1559	0.0000	0.0000	1
IPS	0.50	-0.0625	-0.0625	0.0000	0.0000	1
IPS	0.75	-0.0007	-0.0007	0.0000	0.0000	1
ER2T	0.3000	-0.3703	-0.0411	0.0455	0.0021	9
IBETA	1.00	-0.1172	-0.0391	0.0487	0.0024	3
IPS	0.25	-0.0911	-0.0911	0.0000	0.0000	1
IPS	0.50	-0.0315	-0.0315	0.0000	0.0000	1
IPS	0.75	0.0054	0.0054	0.0000	0.0000	1
IBETA	2.00	-0.1893	-0.0631	0.0535	0.0029	3
IPS	0.25	-0.0729	-0.0729	0.0000	0.0000	1
IPS	0.50	-0.1100	-0.1110	0.0000	0.0000	1
IPS	0.75	-0.0054	-0.0054	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.0638	-0.0213	0.0416	0.0017	3
IPS	0.25	-0.0689	-0.0689	0.0000	0.0000	1
IPS	0.50	-0.0030	-0.0030	0.0000	0.0000	1
IPS	0.75	0.0081	0.0081	0.0000	0.0000	1
ER2T	0.5000	0.1873	0.0208	0.0213	0.0005	9
IBETA	1.00	0.0776	0.0259	0.0119	0.0001	3
IPS	0.25	0.0150	0.0150	0.0000	0.0000	1
IPS	0.50	0.0386	0.0386	0.0000	0.0000	1
IPS	0.75	0.0240	0.0240	0.0000	0.0000	1
IBETA	2.00	0.0129	0.0043	0.0088	0.0001	3
IPS	0.25	0.0144	0.0144	0.0000	0.0000	1
IPS	0.50	-0.0014	-0.0014	0.0000	0.0000	1
IPS	0.75	-0.0001	-0.0001	0.0000	0.0000	1
IBETA	3.00	0.0968	0.0323	0.0308	0.0010	3
IPS	0.25	-0.0024	-0.0024	0.0000	0.0000	1
IPS	0.50	0.0568	0.0568	0.0000	0.0000	1
IPS	0.75	0.0424	0.0424	0.0000	0.0000	1
NSUBJ	100.00	-0.3291	-0.0122	0.0629	0.0040	27
ER2T	0.1000	-0.5583	-0.0620	0.0696	0.0048	9
IBETA	1.00	-0.2302	-0.0767	0.0703	0.0049	3
IPS	0.25	-0.0960	-0.0960	0.0000	0.0000	1
IPS	0.50	-0.1354	-0.1354	0.0000	0.0000	1
IPS	0.75	0.0011	0.0011	0.0000	0.0000	1
IBETA	2.00	-0.2538	-0.0846	0.1034	0.0107	3
IPS	0.25	-0.0511	-0.0511	0.0000	0.0000	1
IPS	0.50	-0.2006	-0.2006	0.0000	0.0000	1
IPS	0.75	-0.0021	-0.0021	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.0743	-0.0248	0.0245	0.0006	3
IPS	0.25	-0.0290	-0.0290	0.0000	0.0000	1
IPS	0.50	-0.0468	-0.0468	0.0000	0.0000	1
IPS	0.75	0.0015	0.0015	0.0000	0.0000	1
ER2T	0.3000	-0.0913	-0.0101	0.0354	0.0013	9
IBETA	1.00	-0.0163	-0.0054	0.0277	0.0008	3
IPS	0.25	0.0096	0.0096	0.0000	0.0000	1
IPS	0.50	-0.0374	-0.0374	0.0000	0.0000	1
IPS	0.75	0.0116	0.0116	0.0000	0.0000	1
IBETA	2.00	-0.1157	-0.0386	0.0463	0.0021	3
IPS	0.25	-0.0215	-0.0215	0.0000	0.0000	1
IPS	0.50	-0.0910	-0.0910	0.0000	0.0000	1
IPS	0.75	-0.0032	-0.0032	0.0000	0.0000	1
IBETA	3.00	0.0406	0.0135	0.0011	0.0000	3
IPS	0.25	0.0135	0.0135	0.0000	0.0000	1
IPS	0.50	0.0124	0.0124	0.0000	0.0000	1
IPS	0.75	0.0147	0.0147	0.0000	0.0000	1
ER2T	0.5000	0.3205	0.0356	0.0374	0.0014	9
IBETA	1.00	0.1543	0.0514	0.0254	0.0006	3
IPS	0.25	0.0744	0.0744	0.0000	0.0000	1
IPS	0.50	0.0557	0.0557	0.0000	0.0000	1
IPS	0.75	0.0241	0.0241	0.0000	0.0000	1
IBETA	2.00	-0.0275	-0.0092	0.0032	0.0000	3
IPS	0.25	-0.0090	-0.0090	0.0000	0.0000	1
IPS	0.50	-0.0125	-0.0125	0.0000	0.0000	1
IPS	0.75	-0.0060	-0.0060	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.1937	0.0646	0.0176	0.0003	3
IPS	0.25	0.0696	0.0696	0.0000	0.0000	1
IPS	0.50	0.0791	0.0791	0.0000	0.0000	1
IPS	0.75	0.0450	0.0450	0.0000	0.0000	1
NSUBJ	250.00	-0.0012	-0.0000	0.0590	0.0035	27
ER2T	0.1000	-0.3679	-0.0409	0.0669	0.0045	9
IBETA	1.00	-0.1334	-0.0445	0.0761	0.0058	3
IPS	0.25	-0.0031	-0.0031	0.0000	0.0000	1
IPS	0.50	-0.1323	-0.1323	0.0000	0.0000	1
IPS	0.75	0.0020	0.0020	0.0000	0.0000	1
IBETA	2.00	-0.1955	-0.0652	0.0975	0.0095	3
IPS	0.25	-0.0166	-0.0166	0.0000	0.0000	1
IPS	0.50	-0.1774	-0.1774	0.0000	0.0000	1
IPS	0.75	-0.0015	-0.0015	0.0000	0.0000	1
IBETA	3.00	-0.0390	-0.0130	0.0228	0.0005	3
IPS	0.25	-0.0026	-0.0026	0.0000	0.0000	1
IPS	0.50	-0.0392	-0.0392	0.0000	0.0000	1
IPS	0.75	0.0028	0.0028	0.0000	0.0000	1
ER2T	0.3000	0.0186	0.0021	0.0395	0.0016	9
IBETA	1.00	0.0314	0.0105	0.0345	0.0012	3
IPS	0.25	0.0426	0.0426	0.0000	0.0000	1
IPS	0.50	-0.0259	-0.0259	0.0000	0.0000	1
IPS	0.75	0.0147	0.0147	0.0000	0.0000	1
IBETA	2.00	-0.0990	-0.0330	0.0449	0.0020	3
IPS	0.25	-0.0088	-0.0088	0.0000	0.0000	1
IPS	0.50	-0.0848	-0.0848	0.0000	0.0000	1
IPS	0.75	-0.0054	-0.0054	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.0862	0.0287	0.0047	0.0000	3
IPS	0.25	0.0323	0.0323	0.0000	0.0000	1
IPS	0.50	0.0304	0.0304	0.0000	0.0000	1
IPS	0.75	0.0234	0.0234	0.0000	0.0000	1
ER2T	0.5000	0.3480	0.0387	0.0415	0.0017	9
IBETA	1.00	0.1249	0.0416	0.0299	0.0009	3
IPS	0.25	0.0748	0.0748	0.0000	0.0000	1
IPS	0.50	0.0333	0.0333	0.0000	0.0000	1
IPS	0.75	0.0167	0.0167	0.0000	0.0000	1
IBETA	2.00	-0.0132	-0.0044	0.0050	0.0000	3
IPS	0.25	-0.0017	-0.0017	0.0000	0.0000	1
IPS	0.50	-0.0013	-0.0013	0.0000	0.0000	1
IPS	0.75	-0.0101	-0.0101	0.0000	0.0000	1
IBETA	3.00	0.2363	0.0788	0.0275	0.0008	3
IPS	0.25	0.0985	0.0985	0.0000	0.0000	1
IPS	0.50	0.0904	0.0904	0.0000	0.0000	1
IPS	0.75	0.0474	0.0474	0.0000	0.0000	1
IDIST	2.00	-4.3483	-0.0537	0.1026	0.0105	81
NSUBJ	50.00	-1.6798	-0.0622	0.1062	0.0113	27
ER2T	0.1000	-1.1237	-0.1249	0.1177	0.0139	9
IBETA	1.00	-0.4680	-0.1560	0.1650	0.0272	3
IPS	0.25	-0.1367	-0.1367	0.0000	0.0000	1
IPS	0.50	-0.3298	-0.3298	0.0000	0.0000	1
IPS	0.75	-0.0014	-0.0014	0.0000	0.0000	1
IBETA	2.00	-0.4210	-0.1403	0.1302	0.0169	3
IPS	0.25	-0.1542	-0.1542	0.0000	0.0000	1
IPS	0.50	-0.2630	-0.2630	0.0000	0.0000	1
IPS	0.75	-0.0038	-0.0038	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.2348	-0.0783	0.0785	0.0062	3
IPS	0.25	-0.1583	-0.1583	0.0000	0.0000	1
IPS	0.50	-0.0750	-0.0750	0.0000	0.0000	1
IPS	0.75	-0.0014	-0.0014	0.0000	0.0000	1
ER2T	0.3000	-0.5990	-0.0666	0.0632	0.0040	9
IBETA	1.00	-0.1982	-0.0661	0.0630	0.0040	3
IPS	0.25	-0.0846	-0.0846	0.0000	0.0000	1
IPS	0.50	-0.1177	-0.1177	0.0000	0.0000	1
IPS	0.75	0.0041	0.0041	0.0000	0.0000	1
IBETA	2.00	-0.2805	-0.0935	0.0783	0.0061	3
IPS	0.25	-0.1029	-0.1029	0.0000	0.0000	1
IPS	0.50	-0.1667	-0.1667	0.0000	0.0000	1
IPS	0.75	-0.0109	-0.0109	0.0000	0.0000	1
IBETA	3.00	-0.1203	-0.0401	0.0613	0.0038	3
IPS	0.25	-0.1049	-0.1049	0.0000	0.0000	1
IPS	0.50	-0.0324	-0.0324	0.0000	0.0000	1
IPS	0.75	0.0170	0.0170	0.0000	0.0000	1
ER2T	0.5000	0.0429	0.0048	0.0965	0.0093	9
IBETA	1.00	0.1214	0.0405	0.0899	0.0081	3
IPS	0.25	0.1329	0.1329	0.0000	0.0000	1
IPS	0.50	-0.0466	-0.0466	0.0000	0.0000	1
IPS	0.75	0.0351	0.0351	0.0000	0.0000	1
IBETA	2.00	-0.1065	-0.0355	0.1119	0.0125	3
IPS	0.25	0.0637	0.0637	0.0000	0.0000	1
IPS	0.50	-0.1568	-0.1568	0.0000	0.0000	1
IPS	0.75	-0.0135	-0.0135	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.0280	0.0093	0.1106	0.0122	3
IPS	0.25	-0.1173	-0.1173	0.0000	0.0000	1
IPS	0.50	0.0873	0.0873	0.0000	0.0000	1
IPS	0.75	0.0580	0.0580	0.0000	0.0000	1
NSUBJ	100.00	-1.6489	-0.0611	0.0961	0.0092	27
ER2T	0.1000	-1.0000	-0.1111	0.1021	0.0104	9
IBETA	1.00	-0.4533	-0.1511	0.1538	0.0236	3
IPS	0.25	-0.1476	-0.1476	0.0000	0.0000	1
IPS	0.50	-0.3066	-0.3066	0.0000	0.0000	1
IPS	0.75	0.0009	0.0009	0.0000	0.0000	1
IBETA	2.00	-0.2905	-0.0968	0.0927	0.0086	3
IPS	0.25	-0.1039	-0.1039	0.0000	0.0000	1
IPS	0.50	-0.1859	-0.1859	0.0000	0.0000	1
IPS	0.75	-0.0008	-0.0008	0.0000	0.0000	1
IBETA	3.00	-0.2562	-0.0854	0.0760	0.0058	3
IPS	0.25	-0.1191	-0.1191	0.0000	0.0000	1
IPS	0.50	-0.1387	-0.1387	0.0000	0.0000	1
IPS	0.75	0.0017	0.0017	0.0000	0.0000	1
ER2T	0.3000	-0.6229	-0.0692	0.0900	0.0081	9
IBETA	1.00	-0.2124	-0.0708	0.0911	0.0083	3
IPS	0.25	-0.0593	-0.0593	0.0000	0.0000	1
IPS	0.50	-0.1671	-0.1671	0.0000	0.0000	1
IPS	0.75	0.0140	0.0140	0.0000	0.0000	1
IBETA	2.00	-0.3528	-0.1176	0.1258	0.0158	3
IPS	0.25	-0.0885	-0.0885	0.0000	0.0000	1
IPS	0.50	-0.2554	-0.2554	0.0000	0.0000	1
IPS	0.75	-0.0088	-0.0088	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.0577	-0.0192	0.0317	0.0010	3
IPS	0.25	-0.0505	-0.0505	0.0000	0.0000	1
IPS	0.50	-0.0201	-0.0201	0.0000	0.0000	1
IPS	0.75	0.0128	0.0128	0.0000	0.0000	1
ER2T	0.5000	-0.0260	-0.0029	0.0690	0.0048	9
IBETA	1.00	0.0512	0.0171	0.0682	0.0047	3
IPS	0.25	0.0881	0.0881	0.0000	0.0000	1
IPS	0.50	-0.0479	-0.0479	0.0000	0.0000	1
IPS	0.75	0.0109	0.0109	0.0000	0.0000	1
IBETA	2.00	-0.1736	-0.0579	0.0700	0.0049	3
IPS	0.25	-0.0216	-0.0216	0.0000	0.0000	1
IPS	0.50	-0.1385	-0.1385	0.0000	0.0000	1
IPS	0.75	-0.0134	-0.0134	0.0000	0.0000	1
IBETA	3.00	0.0964	0.0321	0.0503	0.0025	3
IPS	0.25	-0.0188	-0.0188	0.0000	0.0000	1
IPS	0.50	0.0819	0.0819	0.0000	0.0000	1
IPS	0.75	0.0333	0.0333	0.0000	0.0000	1
NSUBJ	250.00	-1.0196	-0.0378	0.1072	0.0115	27
ER2T	0.1000	-0.8196	-0.0911	0.1193	0.0142	9
IBETA	1.00	-0.3948	-0.1316	0.1694	0.0287	3
IPS	0.25	-0.0743	-0.0743	0.0000	0.0000	1
IPS	0.50	-0.3222	-0.3222	0.0000	0.0000	1
IPS	0.75	0.0017	0.0017	0.0000	0.0000	1
IBETA	2.00	-0.3096	-0.1032	0.1398	0.0195	3
IPS	0.25	-0.0452	-0.0452	0.0000	0.0000	1
IPS	0.50	-0.2626	-0.2626	0.0000	0.0000	1
IPS	0.75	-0.0018	-0.0018	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.1152	-0.0384	0.0427	0.0018	3
IPS	0.25	-0.0358	-0.0358	0.0000	0.0000	1
IPS	0.50	-0.0823	-0.0823	0.0000	0.0000	1
IPS	0.75	0.0030	0.0030	0.0000	0.0000	1
ER2T	0.3000	-0.3650	-0.0406	0.1013	0.0103	9
IBETA	1.00	-0.1231	-0.0410	0.1137	0.0129	3
IPS	0.25	0.0372	0.0372	0.0000	0.0000	1
IPS	0.50	-0.1715	-0.1715	0.0000	0.0000	1
IPS	0.75	0.0111	0.0111	0.0000	0.0000	1
IBETA	2.00	-0.2886	-0.0962	0.1344	0.0181	3
IPS	0.25	-0.0289	-0.0289	0.0000	0.0000	1
IPS	0.50	-0.2510	-0.2510	0.0000	0.0000	1
IPS	0.75	-0.0087	-0.0087	0.0000	0.0000	1
IBETA	3.00	0.0467	0.0156	0.0261	0.0007	3
IPS	0.25	-0.0138	-0.0138	0.0000	0.0000	1
IPS	0.50	0.0361	0.0361	0.0000	0.0000	1
IPS	0.75	0.0244	0.0244	0.0000	0.0000	1
ER2T	0.5000	0.1651	0.0183	0.0784	0.0061	9
IBETA	1.00	0.1135	0.0378	0.1086	0.0118	3
IPS	0.25	0.1556	0.1556	0.0000	0.0000	1
IPS	0.50	-0.0583	-0.0583	0.0000	0.0000	1
IPS	0.75	0.0162	0.0162	0.0000	0.0000	1
IBETA	2.00	-0.1004	-0.0335	0.0076	0.0060	3
IPS	0.25	0.0322	0.0322	0.0000	0.0000	1
IPS	0.50	-0.1191	-0.1191	0.0000	0.0000	1
IPS	0.75	-0.0135	-0.0135	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.1520	0.0507	0.0244	0.0006	3
IPS	0.25	0.0754	0.0754	0.0000	0.0000	1
IPS	0.50	0.0499	0.0499	0.0000	0.0000	1
IPS	0.75	0.0266	0.0266	0.0000	0.0000	1
IDIST	3.00	0.0098	0.0001	0.0666	0.0044	81
NSUBJ	50.00	-0.8267	-0.0306	0.0656	0.0043	27
ER2T	0.1000	-0.6476	-0.0720	0.0654	0.0043	9
IBETA	1.00	-0.2297	-0.0766	0.0747	0.0056	3
IPS	0.25	-0.1538	-0.1538	0.0000	0.0000	1
IPS	0.50	-0.0712	-0.0712	0.0000	0.0000	1
IPS	0.75	-0.0046	-0.0046	0.0000	0.0000	1
IBETA	2.00	-0.2459	-0.0820	0.0697	0.0049	3
IPS	0.25	-0.1194	-0.1194	0.0000	0.0000	1
IPS	0.50	-0.1249	-0.1249	0.0000	0.0000	1
IPS	0.75	-0.0016	-0.0016	0.0000	0.0000	1
IBETA	3.00	-0.1720	-0.0573	0.0787	0.0062	3
IPS	0.25	-0.1471	-0.1471	0.0000	0.0000	1
IPS	0.50	-0.0245	-0.0245	0.0000	0.0000	1
IPS	0.75	-0.0005	-0.0005	0.0000	0.0000	1
ER2T	0.3000	-0.3307	-0.0367	0.0590	0.0035	9
IBETA	1.00	-0.0877	-0.0292	0.0671	0.0045	3
IPS	0.25	-0.1067	-0.1067	0.0000	0.0000	1
IPS	0.50	-0.0102	-0.0102	0.0000	0.0000	1
IPS	0.75	-0.0088	-0.0088	0.0000	0.0000	1
IBETA	2.00	-0.1325	-0.0442	0.0431	0.0019	3
IPS	0.25	-0.0816	-0.0816	0.0000	0.0000	1
IPS	0.50	-0.0539	-0.0539	0.0000	0.0000	1
IPS	0.75	0.0030	0.0030	0.0000	0.0000	1

a	b	c	d	e	f	g
IBETA	3.00	-0.1105	-0.0368	0.0861	0.0074	3
IPS	0.25	-0.1342	-0.1342	0.0000	0.0000	1
IPS	0.50	-0.0051	-0.0051	0.0000	0.0000	1
IPS	0.75	0.0289	0.0289	0.0000	0.0000	1
ER2T	0.5000	0.1516	0.0168	0.0413	0.0017	9
IBETA	1.00	0.0967	0.0322	0.0319	0.0010	3
IPS	0.25	0.0063	0.0063	0.0000	0.0000	1
IPS	0.50	0.0678	0.0678	0.0000	0.0000	1
IPS	0.75	0.0226	0.0226	0.0000	0.0000	1
IBETA	2.00	0.0607	0.0202	0.0227	0.0005	3
IPS	0.25	0.0019	0.0019	0.0000	0.0000	1
IPS	0.50	0.0456	0.0456	0.0000	0.0000	1
IPS	0.75	0.0132	0.0132	0.0000	0.0000	1
IBETA	3.00	-0.0058	-0.0019	0.0662	0.0044	3
IPS	0.25	-0.0783	-0.0783	0.0000	0.0000	1
IPS	0.50	0.0388	0.0388	0.0000	0.0000	1
IPS	0.75	0.0337	0.0337	0.0000	0.0000	1
NSUBJ	100.00	0.1554	0.0058	0.0600	0.0036	27
ER2T	0.1000	-0.4204	-0.0467	0.0564	0.0032	9
IBETA	1.00	-0.1457	-0.0486	0.0465	0.0022	3
IPS	0.25	-0.0907	-0.0907	0.0000	0.0000	1
IPS	0.50	-0.0563	-0.0563	0.0000	0.0000	1
IPS	0.75	0.0013	0.0013	0.0000	0.0000	1
IBETA	2.00	-0.2091	-0.0697	0.0859	0.0074	3
IPS	0.25	-0.0419	-0.0419	0.0000	0.0000	1
IPS	0.50	-0.1661	-0.1661	0.0000	0.0000	1
IPS	0.75	-0.0011	-0.0011	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.0656	-0.0219	0.0378	0.0014	3
IPS	0.25	-0.0655	-0.0655	0.0000	0.0000	1
IPS	0.50	-0.0023	-0.0023	0.0000	0.0000	1
IPS	0.75	0.0022	0.0022	0.0000	0.0000	1
ER2T	0.3000	0.0197	0.0022	0.0266	0.0007	9
IBETA	1.00	0.0328	0.0109	0.0125	0.0002	3
IPS	0.25	-0.0034	-0.0034	0.0000	0.0000	1
IPS	0.50	0.0171	0.0171	0.0000	0.0000	1
IPS	0.75	0.0191	0.0191	0.0000	0.0000	1
IBETA	2.00	-0.0266	-0.0089	0.0146	0.0002	3
IPS	0.25	-0.0107	-0.0107	0.0000	0.0000	1
IPS	0.50	-0.0224	-0.0224	0.0000	0.0000	1
IPS	0.75	0.0065	0.0065	0.0000	0.0000	1
IBETA	3.00	0.0135	0.0045	0.0465	0.0022	3
IPS	0.25	-0.0492	-0.0492	0.0000	0.0000	1
IPS	0.50	0.0315	0.0315	0.0000	0.0000	1
IPS	0.75	0.0313	0.0313	0.0000	0.0000	1
ER2T	0.5000	0.5561	0.0618	0.0340	0.0012	9
IBETA	1.00	0.2424	0.0808	0.0429	0.0018	3
IPS	0.25	0.0703	0.0703	0.0000	0.0000	1
IPS	0.50	0.1280	0.1280	0.0000	0.0000	1
IPS	0.75	0.0442	0.0442	0.0000	0.0000	1
IBETA	2.00	0.1800	0.0600	0.0410	0.0017	3
IPS	0.25	0.0679	0.0679	0.0000	0.0000	1
IPS	0.50	0.0965	0.0965	0.0000	0.0000	1
IPS	0.75	0.0157	0.0157	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.1336	0.0445	0.0105	0.0001	3
IPS	0.25	0.0327	0.0327	0.0000	0.0000	1
IPS	0.50	0.0528	0.0528	0.0000	0.0000	1
IPS	0.75	0.0481	0.0481	0.0000	0.0000	1
NSUBJ	250.00	0.6811	0.0252	0.0640	0.0041	27
ER2T	0.1000	-0.2761	-0.0307	0.0581	0.0034	9
IBETA	1.00	-0.0626	-0.0209	0.0282	0.0008	3
IPS	0.25	-0.0131	-0.0131	0.0000	0.0000	1
IPS	0.50	-0.0521	-0.0521	0.0000	0.0000	1
IPS	0.75	0.0026	0.0026	0.0000	0.0000	1
IBETA	2.00	-0.2023	-0.0674	0.0965	0.0093	3
IPS	0.25	-0.0254	-0.0254	0.0000	0.0000	1
IPS	0.50	-0.1779	-0.1779	0.0000	0.0000	1
IPS	0.75	0.0010	0.0010	0.0000	0.0000	1
IBETA	3.00	-0.0112	-0.0037	0.0114	0.0001	3
IPS	0.25	-0.0168	-0.0168	0.0000	0.0000	1
IPS	0.50	0.0025	0.0025	0.0000	0.0000	1
IPS	0.75	0.0032	0.0032	0.0000	0.0000	1
ER2T	0.3000	0.1884	0.0209	0.0272	0.0007	9
IBETA	1.00	0.1075	0.0358	0.0121	0.0001	3
IPS	0.25	0.0455	0.0455	0.0000	0.0000	1
IPS	0.50	0.0397	0.0397	0.0000	0.0000	1
IPS	0.75	0.0223	0.0223	0.0000	0.0000	1
IBETA	2.00	-0.0133	-0.0044	0.0367	0.0013	3
IPS	0.25	0.0257	0.0257	0.0000	0.0000	1
IPS	0.50	-0.0453	-0.0453	0.0000	0.0000	1
IPS	0.75	0.0064	0.0064	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.0942	0.0314	0.0034	0.0000	3
IPS	0.25	0.0313	0.0313	0.0000	0.0000	1
IPS	0.50	0.0348	0.0348	0.0000	0.0000	1
IPS	0.75	0.0280	0.0280	0.0000	0.0000	1
ER2T	0.5000	0.7689	0.0854	0.0400	0.0016	9
IBETA	1.00	0.3022	0.1007	0.0528	0.0028	3
IPS	0.25	0.1415	0.1415	0.0000	0.0000	1
IPS	0.50	0.1196	0.1196	0.0000	0.0000	1
IPS	0.75	0.0412	0.0412	0.0000	0.0000	1
IBETA	2.00	0.1986	0.0662	0.0460	0.0021	3
IPS	0.25	0.1079	0.1079	0.0000	0.0000	1
IPS	0.50	0.0740	0.0740	0.0000	0.0000	1
IPS	0.75	0.0168	0.0168	0.0000	0.0000	1
IBETA	3.00	0.2680	0.0893	0.0238	0.0006	3
IPS	0.25	0.1165	0.1165	0.0000	0.0000	1
IPS	0.50	0.0724	0.0724	0.0000	0.0000	1
IPS	0.75	0.0791	0.0791	0.0000	0.0000	1
IDIST	4.00	0.7289	0.0090	0.0604	0.0036	81
NSUBJ	50.00	-0.5782	-0.0214	0.0669	0.0045	27
ER2T	0.1000	-0.7030	-0.0781	0.0607	0.0037	9
IBETA	1.00	-0.2283	-0.0761	0.0733	0.0054	3
IPS	0.25	-0.1548	-0.1548	0.0000	0.0000	1
IPS	0.50	-0.0637	-0.0637	0.0000	0.0000	1
IPS	0.75	-0.0098	-0.0098	0.0000	0.0000	1
IBETA	2.00	-0.2482	-0.0827	0.0754	0.0057	3
IPS	0.25	-0.1546	-0.1546	0.0000	0.0000	1
IPS	0.50	-0.0893	-0.0893	0.0000	0.0000	1
IPS	0.75	-0.0043	-0.0043	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.2265	-0.0755	0.0602	0.0036	3
IPS	0.25	-0.1307	-0.1307	0.0000	0.0000	1
IPS	0.50	-0.0846	-0.0846	0.0000	0.0000	1
IPS	0.75	-0.0113	-0.0113	0.0000	0.0000	1
ER2T	0.3000	-0.2337	-0.0260	0.0442	0.0020	9
IBETA	1.00	-0.0663	-0.0221	0.0539	0.0029	3
IPS	0.25	-0.0821	-0.0821	0.0000	0.0000	1
IPS	0.50	-0.0064	-0.0064	0.0000	0.0000	1
IPS	0.75	0.0222	0.0222	0.0000	0.0000	1
IBETA	2.00	-0.1009	-0.0336	0.0466	0.0022	3
IPS	0.25	-0.0850	-0.0850	0.0000	0.0000	1
IPS	0.50	-0.0218	-0.0218	0.0000	0.0000	1
IPS	0.75	0.0059	0.0059	0.0000	0.0000	1
IBETA	3.00	-0.0665	-0.0222	0.0510	0.0026	3
IPS	0.25	-0.0732	-0.0732	0.0000	0.0000	1
IPS	0.50	-0.0222	-0.0222	0.0000	0.0000	1
IPS	0.75	0.0288	0.0288	0.0000	0.0000	1
ER2T	0.5000	0.3586	0.0398	0.0322	0.0010	9
IBETA	1.00	0.1579	0.0526	0.0356	0.0013	3
IPS	0.25	0.0115	0.0115	0.0000	0.0000	1
IPS	0.50	0.0722	0.0722	0.0000	0.0000	1
IPS	0.75	0.0741	0.0741	0.0000	0.0000	1
IBETA	2.00	0.1172	0.0391	0.0375	0.0014	3
IPS	0.25	0.0157	0.0157	0.0000	0.0000	1
IPS	0.50	0.0824	0.0824	0.0000	0.0000	1
IPS	0.75	0.0192	0.0192	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.0835	0.0278	0.0320	0.0010	3
IPS	0.25	0.0006	0.0006	0.0000	0.0000	1
IPS	0.50	0.0199	0.0199	0.0000	0.0000	1
IPS	0.75	0.0630	0.0630	0.0000	0.0000	1
NSUBJ	100.00	0.4417	0.0164	0.0505	0.0026	27
ER2T	0.1000	-0.2688	-0.0299	0.0327	0.0011	9
IBETA	1.00	-0.0884	-0.0295	0.0345	0.0012	3
IPS	0.25	-0.0667	-0.0667	0.0000	0.0000	1
IPS	0.50	-0.0231	-0.0231	0.0000	0.0000	1
IPS	0.75	0.0014	0.0014	0.0000	0.0000	1
IBETA	2.00	-0.1403	-0.0468	0.0412	0.0017	3
IPS	0.25	-0.0568	-0.0568	0.0000	0.0000	1
IPS	0.50	-0.0821	-0.0821	0.0000	0.0000	1
IPS	0.75	-0.0015	-0.0015	0.0000	0.0000	1
IBETA	3.00	-0.0401	-0.0134	0.0232	0.0005	3
IPS	0.25	-0.0400	-0.0400	0.0000	0.0000	1
IPS	0.50	-0.0026	-0.0026	0.0000	0.0000	1
IPS	0.75	0.0025	0.0025	0.0000	0.0000	1
ER2T	0.3000	0.0890	0.0099	0.0273	0.0007	9
IBETA	1.00	0.0481	0.0160	0.0274	0.0008	3
IPS	0.25	-0.0147	-0.0147	0.0000	0.0000	1
IPS	0.50	0.0379	0.0379	0.0000	0.0000	1
IPS	0.75	0.0250	0.0250	0.0000	0.0000	1
IBETA	2.00	0.0193	0.0064	0.0109	0.0001	3
IPS	0.25	-0.0038	-0.0038	0.0000	0.0000	1
IPS	0.50	0.0179	0.0179	0.0000	0.0000	1
IPS	0.75	0.0051	0.0051	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.0216	0.0072	0.0450	0.0020	3
IPS	0.25	-0.0440	-0.0440	0.0000	0.0000	1
IPS	0.50	0.0255	0.0255	0.0000	0.0000	1
IPS	0.75	0.0401	0.0401	0.0000	0.0000	1
ER2T	0.5000	0.6214	0.0690	0.0301	0.0009	9
IBETA	1.00	0.2659	0.0886	0.0223	0.0005	3
IPS	0.25	0.0776	0.0776	0.0000	0.0000	1
IPS	0.50	0.1143	0.1143	0.0000	0.0000	1
IPS	0.75	0.0740	0.0740	0.0000	0.0000	1
IBETA	2.00	0.1912	0.0637	0.0430	0.0019	3
IPS	0.25	0.0736	0.0736	0.0000	0.0000	1
IPS	0.50	0.1010	0.1010	0.0000	0.0000	1
IPS	0.75	0.0166	0.0166	0.0000	0.0000	1
IBETA	3.00	0.1643	0.0548	0.0187	0.0003	3
IPS	0.25	0.0336	0.0336	0.0000	0.0000	1
IPS	0.50	0.0619	0.0619	0.0000	0.0000	1
IPS	0.75	0.0688	0.0688	0.0000	0.0000	1
NSUBJ	250.00	0.8654	0.0321	0.0513	0.0026	27
ER2T	0.1000	-0.1500	-0.0167	0.0306	0.0009	9
IBETA	1.00	-0.0192	-0.0064	0.0076	0.0001	3
IPS	0.25	-0.0088	-0.0088	0.0000	0.0000	1
IPS	0.50	-0.0125	-0.0125	0.0000	0.0000	1
IPS	0.75	0.0021	0.0021	0.0000	0.0000	1
IBETA	2.00	-0.1012	-0.0337	0.0492	0.0024	3
IPS	0.25	-0.0117	-0.0117	0.0000	0.0000	1
IPS	0.50	-0.0901	-0.0901	0.0000	0.0000	1
IPS	0.75	0.0006	0.0006	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	-0.0296	-0.0099	0.0245	0.0006	3
IPS	0.25	-0.0381	-0.0381	0.0000	0.0000	1
IPS	0.50	0.0042	0.0042	0.0000	0.0000	1
IPS	0.75	0.0043	0.0043	0.0000	0.0000	1
ER2T	0.3000	0.2485	0.0276	0.0108	0.0001	9
IBETA	1.00	0.0925	0.0308	0.0137	0.0002	3
IPS	0.25	0.0394	0.0394	0.0000	0.0000	1
IPS	0.50	0.0380	0.0380	0.0000	0.0000	1
IPS	0.75	0.0150	0.0150	0.0000	0.0000	1
IBETA	2.00	0.0584	0.0195	0.0099	0.0001	3
IPS	0.25	0.0294	0.0294	0.0000	0.0000	1
IPS	0.50	0.0194	0.0194	0.0000	0.0000	1
IPS	0.75	0.0096	0.0096	0.0000	0.0000	1
IBETA	3.00	0.0977	0.0326	0.0054	0.0000	3
IPS	0.25	0.0386	0.0386	0.0000	0.0000	1
IPS	0.50	0.0283	0.0283	0.0000	0.0000	1
IPS	0.75	0.0307	0.0307	0.0000	0.0000	1
ER2T	0.5000	0.7669	0.0852	0.0406	0.0016	9
IBETA	1.00	0.2952	0.0984	0.0442	0.0020	3
IPS	0.25	0.1336	0.1336	0.0000	0.0000	1
IPS	0.50	0.1128	0.1128	0.0000	0.0000	1
IPS	0.75	0.0488	0.0488	0.0000	0.0000	1
IBETA	2.00	0.2011	0.0670	0.0487	0.0024	3
IPS	0.25	0.0833	0.0833	0.0000	0.0000	1
IPS	0.50	0.1056	0.1056	0.0000	0.0000	1
IPS	0.75	0.0123	0.0123	0.0000	0.0000	1

[cont'd]

a	b	c	d	e	f	g
IBETA	3.00	0.2705	0.0902	0.0382	0.0015	3
IPS	0.25	0.1335	0.1335	0.0000	0.0000	1
IPS	0.50	0.0613	0.0613	0.0000	0.0000	1
IPS	0.75	0.0757	0.0757	0.0000	0.0000	1

Total cases = 324

Appendix B

COMPARISONS OF SELECTED AND CORRECTED COEFFICIENTS IN TERMS OF BIAS($r^2_{xy_s}$), BIAS($r^2_{xy_c}$), AND VAR($r^2_{xy_s}$) - VAR($r^2_{xy_c}$) THAT CORRESPOND WITH THE FUNCTIONS INVESTIGATED IN TABLES 8-12

Table 14

Selected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Sample Size, and the Proportion Selected[a]

Distribution of x scores	Proportion selected								
	.25			.50			.75		
	Number of subjects			Number of subjects			Number of subjects		
	50	100	250	50	100	250	50	100	250
Normal	-.1669	-.1721	-.1890	-.0148	-.0272	-.0518	-.0886	-.0913	-.1006
Uniform	-.1937	-.2163	-.2451	.1010	.1357	.1100	-.0932	-.0795	-.0863
Right-skewed	-.1804	-.2162	-.2360	-.1332	-.1685	-.1759	-.1080	-.1288	-.1390
Left-skewed	-.1893	-.2137	-.2224	-.1905	-.2135	-.2171	-.1481	-.1489	-.1364

[a]Values presented are averaged over 9 different conditions.

Table 15

Corrected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Sample Size, and the Proportion Selected[a]

Distribution of x scores	Proportion selected								
	.25			.50			.75		
	Number of subjects			Number of subjects			Number of subjects		
	50	100	250	50	100	250	50	100	250
Normal	-.0383	.0048	.0012	.1821	.1723	.1628	.0035	-.0007	.0013
Uniform	.0147	.0276	.0194	.3601	.3951	.3758	.0017	.0053	.0015
Right-skewed	-.0621	-.0233	-.0141	.0226	.0168	.0264	-.0104	-.0112	-.0107
Left-skewed	-.0029	-.0169	-.0080	-.0248	-.0537	-.0670	-.0074	-.0084	-.0079

[a]Values presented are averaged over 9 different conditions.

Table 16

A Comparison of Selected and Corrected Coefficients in Terms of $VAR(r^2_{xy_S}) - VAR(r^2_{xy_C})$ as a Function of the Distribution of x Scores, the Sample Size, and the Proportion Selected[a]

Distribution of x scores	Proportion selected								
	.25			.50			.75		
	Number of subjects			Number of subjects			Number of subjects		
	50	100	250	50	100	250	50	100	250
Normal	-.0991	-.0310	-.0041	-.0053	-.0006	-.0001	-.0013	-.0002	-.0000
Uniform	-.1082	-.0950	-.0366	-.0087	-.0008	-.0001	-.0013	-.0001	-.0000
Right-skewed	-.1257	-.0643	-.0149	-.0148	-.0037	-.0004	-.0021	-.0003	-.0000
Left-skewed	-.1143	-.0611	-.0119	-.0393	-.0073	-.0007	-.0058	-.0006	-.0000

[a]Values presented are averaged over 9 different conditions.

Table 17

Selected Coefficients in Terms of Bias as a Function of the Distribution of x Scores and the Form of the Regression[a]

Distribution of x scores	Form of the regression		
	Linear	Convex	Concave
Normal	-.1002	-.0355	-.1650
Uniform	-.0610	.0401	-.1682
Right-skewed	-.1688	-.1046	-.2219
Left-skewed	-.1901	-.1386	-.2312

[a]Values presented are averaged over 27 different conditions.

Table 18

Corrected Coefficients in Terms of Bias as a Function of the Distribution of x Scores and the Form of the Regression[a]

Distribution of x scores	Form of the regression		
	Linear	Convex	Concave
Normal	.0585	.1715	-.0670
Uniform	.1523	.2782	-.0302
Right-skewed	.0044	.1028	-.1291
Left-skewed	-.0212	.0634	-.1079

[a]Values presented are averaged over 27 different conditions.

Table 19

A Comparison of Selected and Corrected Coefficients in Terms of $VAR(r^2_{xy_s}) - VAR(r^2_{xy_c})$ as a Function of the Distribution of x Scores and the Form of the Regression[a]

Distribution of x scores	Form of the regression		
	Linear	Convex	Concave
Normal	-.0176	-.0152	-.0144
Uniform	-.0255	-.0214	-.0367
Right-skewed	-.0264	-.0172	-.0318
Left-skewed	-.0270	-.0202	-.0332

[a]Values presented are averaged over 27 different conditions.

Table 20

Selected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Form of the Regression, and the Strength of the Relationship [$E(r^2_{xy_t})$][a]

Distribution of x scores	Strength of the relationship								
	.1			.3			.5		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	.0265	.0686	-.0134	-.1115	-.0395	-.1586	-.2157	-.1357	-.3231
Uniform	.0734	.1791	-.0046	-.0538	.0319	-.1755	-.2027	-.0908	-.3225
Right-skewed	-.0119	.0342	-.0518	-.1705	-.1070	-.2350	-.3161	-.2411	-.3787
Left-skewed	-.0345	-.0084	-.0599	-.1895	-.1361	-.2418	-.3464	-.2713	-.3921

[a]Values presented are averaged over 9 different conditions.

Table 21

Corrected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Form of the Regression, and the Strength of the Relationship [$E(r^2_{xy_t})$]

Distribution of x scores	Strength of the relationship								
	.1			.3			.5		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	.1478	.2156	.0680	.0485	.1736	-.0766	-.0208	.1252	-.1924
Uniform	.2536	.3035	.1115	.1281	.2863	-.0299	.0753	.2447	-.1720
Right-skewed	.0983	.1748	.0089	-.0008	.0968	-.1386	-.0844	.0368	-.2577
Left-skewed	.0673	.1351	.0368	-.0227	.0647	-.1261	-.1081	-.0096	-.2345

[a]Values presented are averaged over 9 different conditions.

Table 22

A Comparison of Selected and Corrected Coefficients in Terms of $VAR(r^2_{xy_S}) - VAR(r^2_{xy_C})$ as a Function of the Distribution of x Scores, the Form of the Regression, and the Strength of the Relationship $[E(r^2_{xy_t})]$

Distribution of x scores	Strength of the relationship								
	.1			.3			.5		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	-.0265	-.0209	-.0239	-.0169	-.0129	-.0108	-.0094	-.0120	-.0085
Uniform	-.0378	-.0296	-.0375	-.0284	-.0242	-.0354	-.0104	-.0104	-.0372
Right-skewed	-.0314	-.0205	-.0299	-.0237	-.0184	-.0361	-.0243	-.0127	-.0293
Left-skewed	-.0311	-.0259	-.0293	-.0274	-.0170	-.0354	-.0223	-.0177	-.0348

[a]Values presented are averaged over 9 different conditions.

Table 23

Selected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Form of the Regression, and the Number of Subjects[a]

Distribution of x scores	Number of subjects								
	.1			.3			,5		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	-.0835	-.0455	-.1413	-.1102	-.0025	-.1580	-.1069	-.0387	-.1958
Uniform	-.0625	.0451	-.1685	-.0533	.0424	-.1493	-.0672	.0326	-.1868
Right-skewed	-.1444	-.0759	-.2013	-.1768	-.1135	-.2231	-.1852	-.1245	-.2412
Left-skewed	-.1807	-.1298	-.2173	-.2003	-.1421	-.2337	-.1893	-.1439	-.2427

[a]Values presented are averaged over 9 different conditions.

Table 24

Corrected Coefficients in Terms of Bias as a Function of the Distribution of x Scores, the Form of the Regression, and the Number of Subjects[a]

Distribution of x scores	Number of subjects								
	50			100			250		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	.0467	.1587	-.0582	.0628	.1716	-.0580	.0660	.1841	-.0849
Uniform	.1563	.2672	-.0470	.1479	.2818	-.0018	.1528	.2855	-.0417
Right-skewed	-.0112	.0884	-.1272	.0130	.0944	-.1251	.0112	.1255	-.1351
Left-skewed	-.0179	.0548	-.0719	-.0196	.0620	-.1213	-.0260	.0735	-.1305

[a]Values presented are averaged over 9 different conditions.

Table 25

A Comparison of Selected and Corrected Coefficients in Terms of $VAR(r^2_{xy_s}) - VAR(r^2_{xy_c})$ as a Function of the Distribution of x Scores, the Form of the Regression, and the Number of Subjects[a]

Distribution of x scores	Number of subjects								
	50			100			250		
	Form of the regression			Form of the regression			Form of the regression		
	Linear	Convex	Concave	Linear	Convex	Concave	Linear	Convex	Concave
Normal	-.0365	-.0355	-.0337	-.0151	-.0091	-.0076	-.0013	-.0012	-.0017
Uniform	-.0317	-.0308	-.0556	-.0321	-.0251	-.0388	-.0129	-.0082	-.0157
Right-skewed	-.0482	-.0349	-.0594	-.0273	-.0126	-.0284	-.0038	-.0040	-.0075
Left-skewed	-.0536	-.0452	-.0607	-.0250	-.0135	-.0305	-.0023	-.0019	-.0084

[a]Values presented are averaged over 9 different conditions.

Table 26

Selected Coefficients in Terms of Bias as a Function of the Strength of the Relationship [$E(r^2_{xy_t})$] and the Proportion Selected[a]

$E(r^2_{xy_t})$	Proportion selected		
	.25	.50	.75
.1	-.0427	.1300	-.0399
.3	-.2068	-.0686	-.1218
.5	-.3608	-.2728	-.1755

[a]Values presented are averaged over 36 different conditions.

Table 27

Corrected Coefficients in Terms of Bias as a Function of the Strength of the Relationship [$E(r^2_{xy_t})$] and the Proportion Selected[a]

$E(r^2_{xy_t})$	Proportion selected		
	.25	.50	.75
.1	.0644	.3310	.0099
.3	-.0207	.1296	-.0080
.5	-.0682	-.0685	-.0127

[a]Values presented are averaged over 36 different conditions.

Table 28

A Comparison of Selected and Corrected Coefficients in Terms of $VAR(r^2_{xy_S}) - VAR(r^2_{xy_C})$ as a Function of the Strength of the Relationship [$E(r^2_{xy_t})$] and the Proportion Selected[a]

$E(r^2_{xy_t})$	Proportion selected		
	.25	.50	.75
.1	-.0073	-.0070	-.0017
.3	-.0638	-.0070	-.0009
.5	-.0504	-.0064	-.0004

[a]Values presented are averaged over 36 different conditions.

Appendix C

A PRESENTATION OF THE RESULTS OF A FIVE-FACTOR FACTORIAL ANOVA†

---

†IDIST = the distribution of x scores
1 = normal distribution
2 = uniform distribution
3 = right-skewed distribution
4 = left-skewed distribution
NSUBJ = the number of subjects
ER2T = the expected value of the squared correlation in the total group ($E(r^2_{xy_t})$
IBETA = the regression curve
1 = linear regression
2 = convex regression
3 = concave regression
IPS = the proportion selected

Source of variation	SS	df	MS
Main effects	1.132	11	0.103
IDIST	0.187	3	0.062
NSUBJ	0.100	2	0.050
ER2T	0.606	2	0.303
IBETA	0.083	2	0.042
IPS	0.156	2	0.078
2-way interactions	0.618	48	0.013
IDIST x NSUBJ	0.014	6	0.002
IDIST x ER2T	0.003	6	0.001
IDIST x IBETA	0.036	6	0.006
IDIST x IPS	0.168	6	0.028
NSUBJ x ER2T	0.001	4	0.000
NSUBJ x IBETA	0.008	4	0.002
NSUBJ x IPS	0.117	4	0.029
ER2T x IBETA	0.018	4	0.004
ER2T x IPS	0.161	4	0.040
IBETA x IPS	0.091	4	0.023
3-way interactions	0.187	104	0.002
IDIST x NSUBJ x ER2T	0.006	12	0.000
IDIST x NSUBJ x IBETA	0.002	12	0.000
IDIST x NSUBJ x IPS	0.009	12	0.001
IDIST x ER2T x IBETA	0.023	12	0.002
IDIST x ER2T x IPS	0.013	12	0.008
IDIST x IBETA x IPS	0.097	12	0.000
NSUBJ x ER2T x IBETA	0.002	8	0.000
NSUBJ x ER2T x IPS	0.003	8	0.000
NSUBJ x IBETA x IPS	0.005	8	0.001
ER2T x IBETA x IPS	0.026	8	0.003
4-way interactions	0.062	112	0.001
IDIST x NSUBJ x ER2T x IBETA	0.008	24	0.000
IDIST x NSUBJ x ER2T x IPS	0.005	24	0.000
IDIST x NSUBJ x IBETA x IPS	0.009	24	0.000
IDIST x ER2T x IBETA x IPS	0.030	24	0.001
NSUBJ x ER2T x IBETA x IPS	0.011	16	0.001
5-way interactions	0.016	48	0.000
IDIST x NSUBJ x ER2T x IBETA x IPS	0.016	48	0.000
Explained	2.015	323	0.006
Residual	0.000	0	0.000
Total	2.015	323	0.006

REFERENCES

Birnbaum, Z.W., Paulson, E., & Andrews, F.C. (1950). On the effect of selection performed on some coordinates of the multi-dimensional population. Psychometrika, 15, 191-204.

Bobko, P., & Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. Applied Psychological Measurement, 4, 385-398.

Brewer, J.K., & Hills, J.R. (1969). Univariate selection: The effects of size of correlation, degree of skew, and degree of restriction. Psychometrika, 34, 347-361.

Bulgren, W.G. (1971). On representations of the doubly noncentral F distribution. Journal of the American Statistical Association, 66, 184-186.

Cohen, A.C. (1955). Restriction and selection in samples from bivariate normal distributions. Journal of the American Statistical Association, 50, 884-893.

Forsyth, R.A. (1971). An empirical note on correlation coefficients corrected for restriction on range. Educational and Psychological Measurement, 31, 115-123.

Graybill, F.A. (1961). An introduction to linear statistical models, Volume I. New York: McGraw-Hill.

Greener, J.M., & Osburn, H.G. (1979). An empirical study of the accuracy of corrections for restriction in range due to explicit selection. Applied Psychological Measurement, 3, 31-41.

Greener, J.M., & Osburn, H.G. (1980). Accuracy of corrections for restriction in range due to explicit selection in heteroscedastic and nonlinear distributions. Educational and Psychological Measurement, 40, 337-345.

Gross, A.L. (1982). Relaxing the assumptions underlying corrections for restriction of range. Educational and Psychological Measurement, 42, 795-801.

Gross, A.L., & Fleischman, L. (1983). Restriction of range corrections when both distributions and selection assumptions are violated. Applied Psychological Measurement, 2, 227-237.

Gross, A.L., & Kagen, E. (1983). Not correcting for restriction of range can be advantageous. Educational and Psychological Measurement, 43, 389-396.

Gross, A.L., & Perry, P. (1983). Validating a selection test, a predictive probability approach. Psychometrika, 48, 115-127.

Gullickson, A., & Hopkins, K. (1976). Interval estimation of correlation coefficients corrected for restriction of range. Educational and Psychological Measurement, 36, 9-25.

IBM Technical Publications Department. (1968). IBM scientific package, 4th ed. White Plans, NY: IBM.

Johnson, N.I., & Kotz, S. (1970). Distributions in statistics: Continuous univariate distributions - 2. Boston, MA: Houghton Mifflin Company.

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. Royal Society of Edinburgh Proceedings, Section A, 62, 28-30.

Levin, J. (1972). The occurrence of an increase in correlation by restriction of range. Psychometrika, 37, 93-97.

Linn, R.L. (1968). Range restriction problems in the use of self-selected groups for test validation. Psychological Bulletin, 69, 69-73.

Linn, R.L. (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. Journal of Educational Measurement, 20, 1-15.

Linn, R.L., & Dunbar, S.B. (1982). Predictive validity of admissions measures: Correlations for selection on several variables. Journal of College Student Personnel, 23, 222-226.

Linn, R.L., Harnisch, D.L., & Dunber, S.B. (1981). Corrections for range restriction: An empirical investigation of conditions resulting in conservative correction. Journal of Applied Psychology, 66, 655-663.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Moran, P.A.P. (1970). On asymptotically optimal tests of composite hypotheses. Biometrika, 57, 47-55.

Novick, M.R., & Jackson, P. (1974). Statistical methods for educational and psychological research. New York: McGraw-Hill.

Novick, M.R., & Thayer, D.T. (1969). An investigation of the accuracy of the Pearson selection formulas (ETS-RM-69-22). Princeton, NJ: Educational Testing Service.

Nunnally, J.C. (1978). Psychometric theory, 2nd ed. New York: McGraw-Hill.

Pearson, K. (1903). On the influence of natural selection on the variability and correlation of organs. Philosophical Transactions Royal Society of London, Series A, 200, 1-66.

Roe, R.A. (1979). The correction for restriction of range and the difference between intended and actual selection. Educational and Psychological Measurement, 39, 551-559.

Searle, S.R. (1971). Linear models. New York: John Wiley.

Watterson, G.A. (1959). Linear estimation in censored samples from multivariate normal populations. Annals of Mathematical Statistics, 30, 814-824.