Research Report

Check for updates

A Hybrid Model for Orthogonal Regression

ETS RR–23-04

Michael Kane

December 2023

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

A Hybrid Model for Orthogonal Regression

Michael Kane

ETS, Princeton, NJ

Linear functional relationships are intended to be symmetric and therefore cannot generally be accurately estimated using ordinary least squares regression equations. Orthogonal regression (OR) models allow for errors in both Y and X and therefore can provide symmetric estimates of these relationships. The most well-established OR model, the errors-in-variables (EIV) model, assumes that the observed scatter around the line is due entirely to errors of measurement in Y and X and that the ratio of the error variances is known. If most of the variance around the line is known to be due to the errors of measurement in Y and X, the EIV model can provide an unbiased maximum likelihood estimate for a functional relationship. However, if a substantial part of the variability around the line is due to natural variability, which is not attributable to errors of measurement in Y or X, the ratio of the measurement error variances is not well defined and the EIV model is not directly applicable. The main contribution of this report is the development of a hybrid model that provides plausible estimates for linear functional relationships in cases with substantial natural variability and substantial errors of measurement. An analysis of female and male differential test functioning between an essay test and an objective test used as parts of a licensure examination provides an illustration of the use of the hybrid model.

Keywords Errors-in-variables (EIV) model; orthogonal regression; functional relationships; geometric-mean (GM) model; differential test functioning

doi:10.1002/ets2.12367

A large subset of theoretical relationships in science can be expressed as bivariate linear equations and interpreted as functional relationships between the quantities on the two sides of the equations. These functional relationships are symmetric in that the equations are expected to work equally well in both directions. Sprent (1969) characterized these relationships as "law-like" or "functional" in that "there is a basic mathematical relationship between variables with which their data would accord were it not for the fact that this relationship is obscured to some extent by "random" fluctuations, perhaps associated with both variables" (p. 29). An essential feature of these functional relationships is their symmetry.

Many of the most familiar physical "laws" (e.g., Newton's second law of motion, $F = \text{ma}$, or force equals mass times acceleration) state functional relationships. There are fewer examples of this kind of quantitative relationship in the social sciences, but there are some. For example, Fechner's law states that sensation S increases as the logarithm of the intensity of a stimulus, or $S = k \log R$ (Stevens, 1964). Note that the variables on the two sides of these equations are not the same—a force is not the same thing as an acceleration, and a sensation is not the same thing as the logarithm of a stimulus intensity—but the laws say that the quantities on the two sides of the equations are equivalent in a strong sense (Sprent, 1969).

If the quantities on the two sides of the equation stating a linear functional relationship were measured without error, we would expect the relationship to be represented exactly by the linear equation

$$y = a + bx, \tag{1a}$$

where x and y represent true values of the variables and a and b are parameters that adjust for differences in the scales on which x and y are reported. As an algebraic equation, this relationship could also be written as

$$x = \frac{1}{b}(y - a). \tag{1b}$$

Equations 1a and 1b represent the same functional relationship.

If Equation 1a (or 1b) held exactly, and we plotted the values of y against the values of x, or x against y, for a sample of data points, the points would all fall on a straight line. In practice, of course, because of errors of measurement in the

Corresponding author: M. Kane, E-mail: mkane@ets.org

observed values of the variables Y and X, as well and other more or less random sources of variability, the corresponding observed data points (X, Y) will generally be scattered around the line. The problem addressed in this report is how to estimate the parameters a and b for the underlying true-score relationship from the observed values for Y and X.

Ordinary least squares (OLS) regression is commonly used in predicting unknown values of a dependent variable Y from known values of an independent variable X and therefore is highly directional (i.e., from X to Y). The OLS model is not symmetric and generally yields two distinct lines, one for the regression of Y on X and the other for the regression of X on Y. The slopes of these two lines are biased (in opposite directions) as estimates of the slopes of the underlying functional relationship, and therefore the OLS model is not appropriate for evaluating functional relationships.

In addition to their role in scientific laws and theories and in validating measures of theory-based constructs (Cronbach & Meehl, 1955), binary functional relationships play important roles in addressing some basic issues in educational measurement. A familiar example occurs in equating different forms of a test so that the scores yielded by the different test forms can be considered equivalent. In this case, Y and X would represent observed scores on the forms of the test to be equated, and the relationship of interest would be the linear equating relationship for the two forms (Kolen & Brennan, 2014).

Estimates of functional relationships are also useful in studies of convergent validation and of differential test functioning for distinct groups of test takers. To the extent that the relationship between two tests differs across the groups, one or the other or both tests could be considered biased (Dorans, 2004; Kane & Mroch, 2010). These empirical comparisons do not in themselves indicate which of the measurements is biased, but they can indicate the presence of bias, which needs to be investigated. This use of the methodology developed in this report in evaluating differential test functioning is illustrated in its next-to-last section.

A number of authors (Adcock, 1877; Brace, 1977; Deming, 1943; Fuller, 1987; Isobe et al., 1990; Kane, 2021; Kane & Mroch, 2010, 2020; Kruskal, 1953; Madansky, 1959; Pearson, 1901; Ricker, 1973; Riggs et al., 1978; Sprent, 1990) have discussed statistical models for estimating functional relationships between two variables when both variables contain errors. These models are generally referred to as orthogonal regression (OR) models, because several of the earliest of these models required that the sum of squared orthogonal distances from the points to the line be minimized. The errors-in-variables (EIV) model (Deming, 1943; Fuller, 1987) provides an elegant solution to the problem of estimating functional relationships, but it assumes that all the variability around the line is due to errors of measurement in the two variables. A simpler OR model, the geometric mean (GM) model, focuses on minimizing the squared orthogonal distances of the data points from the line and relies on weaker assumptions.

The requirement that the only sources of variability around the line are the errors of measurement in Y and X constitutes a serious limitation on the usefulness of the EIV model, because most observed relationships also include variability that cannot be explained in terms of measurement errors. This additional variability around the line has been called *natural variability* (Kruskal, 1953; Ricker, 1973) or *equation errors* (Fuller, 1987; Sprent, 1969). I will refer to the additional variability around the line (over and above what can be attributed to errors of measurement in Y and X) as *natural variability*.

Measurement errors are fluctuations in the estimated values of a variable over repeated measurements, using essentially the same measurement procedure, on the same people or objects. Operationally, measurement errors are estimated in studies of the precision, reliability, or generalizability of measurements. The measurement error in X is uniquely associated with X, and the measurement error in Y is uniquely associated with Y.

Natural variability, in contrast, is not uniquely attributable to measurement errors in Y or X; rather, it is attributable to additional sources of random fluctuations, or noise in the system. Sprent (1969) suggested that these fluctuations

> very often represent genuine variability in the experimental material rather than errors of measurement. This is particularly true in most biological or economic studies. In this case the underlying relationship may be looked upon as representing an average relationship for the population under study. Departures from the relationship represent what may be termed individuality. The average relationship is a useful concept in comparing populations. For two or more populations we may well ask whether the average relationships differ. (p. 29)

As Ricker (1973) put it, natural variability is "inherent in the material being measured" (p. 410) and may be due to a variety of environmental or individual effects that are not attributable to errors of measurement in X or in Y. In many studies (especially in the social and biological sciences), the errors of measurement for Y and X can be smaller than the natural variability in the relationship (Ricker, 1973).

The EIV model provides unbiased, maximum likelihood estimates of the parameters a and b in Equation 1 and an explicit model for the structure of the errors in Y and X. However, it assumes that all the variability around the line is due to the errors of measurement in Y and X and therefore that there is no natural variability in the relationship. As indicated earlier, this assumption is untenable in many cases, especially in the social and biological sciences.

The hybrid model developed in this report removes this obstacle to the use of the EIV model by treating the natural variability as an additional source of error in Y and X in a way that preserves the symmetry of the functional relationship. The hybrid model is developed by imposing two natural constraints on the EIV model. The first constraint requires that the error variances in Y and X be adjusted to incorporate the variance attributed to natural variability, and the second constraint requires that the solution generated by the model preserve the symmetry of the functional relationship. The hybrid model reduces to the EIV model when all the variability around the line is due to errors of measurement in Y and X and to the GM model when all the variability is attributable to natural variability. In the middle ground between these extremes, where the data include both substantial errors of measurement and substantial natural variability, the hybrid model provides a plausible, unique estimate of the functional relationship, and it preserves the symmetry of the relationship.

The next section briefly describes some of the properties of the OLS regression model, the GM model, and the classical EIV model. The OLS model is not a plausible model for estimates of functional relationships because of its asymmetry, but it provides a useful contrast and some limiting cases. The following section discusses the main problem in applying the EIV model: the need to assume that all the variability around the line is entirely due to errors of measurement. The third section develops a hybrid model that provides plausible, unique estimates of the functional relationship when the scatter in the data reflects both substantial errors of measurement and substantial natural variability. The fourth section illustrates the use of the hybrid model in analyzing differential test functioning for two measures (a multiple-choice test and an essay test) used in a licensure examination. The final section reviews the main points in this report and presents some general conclusions.

The Ordinary Least Squares, Geometric Mean, and Errors-in-Variables Models

This report is concerned mainly with the EIV model and a new hybrid version of the EIV model, but I begin with brief descriptions of the OLS regression model, the GM model, and the classical EIV model.

The Ordinary Least Squares Model

The standard OLS regression model for predicting a dependent variable Y contingent on an independent variable X is given by

$$Y = \mathrm{a} + bX + \varepsilon, \tag{2}$$

where ε represents a random error that is independent of X and has a mean of 0 and a standard deviation of σ_ε. There is only one error term in the OLS model, and it incorporates all the sources of variability around the line for Y given X; that is, the OLS model assumes that the observed values of Y are the sums of a predicted value associated with the line represented by $Y = a + bX$ and a random component represented by an error term ε.

The OLS model estimates the values of the constants a and b by minimizing the average squared deviation of the observed values of Y from the predicted values given by the regression line; that is, it minimizes the average squared vertical distances of the data points from the line.

The equation for the OLS regression of Y on X has the familiar form

$$\hat{Y} = r_{XY}\frac{s_Y}{s_X}(X - \overline{X}) + \overline{Y}, \tag{3}$$

where X is the independent variable and $\hat{Y}$ is the estimate of the dependent variable, Y, based on X. The estimated values of Y have the same mean but a smaller variance than the observed values of Y, and therefore the estimated values of Y are said to be regressed toward the mean of Y (Campbell & Kenny, 1999; Galton, 1886).

The OLS regression of Y on X provides optimal (i.e., least squares) linear estimates of Y given X, but it is not optimal in this sense for predicting X from Y. It is therefore not appropriate to use Equation 3 in reverse to predict X from Y. To get OLS predictions of X given Y, it would be necessary to estimate the OLS regression of X on Y, which would provide least squares estimates of X given Y. That is, OLS regression is strongly directional and asymmetric. It does not yield a functional relationship that can be used in either direction; rather, it is an algorithm for predicting values of the dependent variable, Y, from observed values of the independent variable, X.

The OLS regression of Y on X can be represented as a "fitted line" in the X-Y plane (Draper, 1991, p. 7):

$$Y = r_{XY}\frac{s_Y}{s_X}(X - \overline{X}) + \overline{Y}. \tag{4}$$

Equation 4 specifies a line in the X-Y plane, and as such, it does not give Y or X a special role as a dependent or independent variable. This small change in notation represents a substantial conceptual shift from prediction to the representation of a functional relationship between Y and X; it constitutes a change in the question being asked and a corresponding change in the interpretation of the "fitted line" (Draper, 1991).

As noted earlier, to obtain an optimal OLS estimate of X given Y, it would be necessary to estimate the line that minimizes the average squared deviations of the observed values of X from the predicted values of X given Y. In this reverse version of the OLS regression model, X is the dependent variable and Y is the independent variable. The OLS regression of X on Y is represented by

$$\widehat{X} = r_{XY}\frac{s_X}{s_Y}(Y - \overline{Y}) + \overline{X}, \tag{5}$$

where $\widehat{X}$ is the predicted value of X given Y. As a "fitted line" in the X-Y plane (Draper, 1991, p. 7), the OLS regression of X on Y can be written as

$$X = r_{XY}\frac{s_X}{s_Y}(Y - \overline{Y}) + \overline{X}. \tag{6}$$

Rewriting this equation to represent Y as a function of X, we have

$$Y = \frac{1}{r_{XY}}\frac{s_Y}{s_X}(X - \overline{X}) + \overline{Y}. \tag{7}$$

The fitted line generated by the OLS regression of Y on X is not the same as the fitted line generated by the OLS regression of X on Y, unless r_{XY} is equal to 1.0; that is, OLS regression is not symmetric and generates biased estimates of the functional relationship between the two variables because of regression toward the mean (Carroll & Ruppert, 1996; Fuller, 1987; Kane & Mroch, 2010, 2020; Linn & Werts, 1971).

The absolute value of the slope of the line given by Equation 4 is smaller than the absolute value of the slope in Equation 7 by a factor of $\left(r_{xy}\right)^2$, because the X-on-Y line is regressed toward the Y axis and the Y-on-X line is regressed toward the X axis (Campbell & Kenny, 1999; Galton, 1886). The two OLS lines intersect at the centroid (the point defined by the means of Y and X) of the bivariate distribution for Y and X but generally have different slopes.

The Geometric Mean Model for Orthogonal Regression

The GM regression equation can be derived using bivariate principal components analysis (Kane & Mroch, 2010, 2020; Riggs et al., 1978), after transforming both variables to z scores:

$$\mathrm{z_Y} = \frac{(\mathrm{Y} - \overline{\mathrm{Y}})}{\mathrm{S_Y}}$$

and

$$z_X = \frac{(X - \overline{X})}{\mathrm{S_X}}.$$

The line for the first principal component is then

$$z_Y = \operatorname{sgn}\left(r_{XY}\right) z_X, \tag{8}$$

where the sgn function simply ensures that the slope is negative when the relationship between Y and X is negative. The second principal component represents the variability orthogonal to the line specified by the first principal component, and it is minimized in the principal components analysis. The GM model focuses on the distribution of the points in the X-Y plane, and it minimizes the mean-squared orthogonal deviations, on the z score scale, of the data points from the line (Kruskal, 1953; Ricker, 1973).

Transforming Equation 8 back to the observed-score scales for Y and X, the GM equation can be written as

$$Y = \text{sgn}\left(r_{XY}\right)\frac{S_Y}{S_X}(X - \overline{X}) + \overline{Y}, \tag{9}$$

or, more simply, as

$$Y = \pm\frac{S_Y}{S_X}(X - \overline{X}) + \overline{Y}, \tag{10}$$

with the sign of the expression on the right-hand side of the equation understood as the sign of r_{XY}. Note that the GM slope depends on the sign, but not the magnitude, of the correlation coefficient.

The GM regression of X on Y is

$$X = \pm\frac{S_X}{S_Y}(Y - \overline{Y}) + \overline{X}. \tag{11}$$

Solving Equation 11 for Y as a function of X, we get Equation 10, and therefore the GM regression model is symmetric in the sense that it yields the same equation in both directions, and it yields a single line in the X-Y plane.

As its name indicates, the absolute value of the slope of the GM estimator is equal to the geometric mean (the square root of the product) of the slopes of the OLS regressions given by Equations 4 and 7 (Brace, 1977; Ricker, 1973; Riggs et al., 1978; Sprent, 1990). Like the OLS lines, the GM line passes through the centroid, and for the z score scales, it bisects the angle between the two OLS lines.

The Errors-in-Variables Model

The EIV model has a long history (Adcock, 1877; Carroll & Ruppert, 1996; Deming, 1943; Fuller, 1987; Madansky, 1959; Pearson, 1901), but it has not seen much use in educational research, in part because of its restrictive assumptions. What Fuller (1987) called the "classical errors-in-variables model" (p. 30) assumes that both the *Y and X* variables contain measurement error:

$$Y = y + u, \tag{12a}$$

$$X = x + v, \tag{12b}$$

where y and x are the true scores for Y and X, respectively, and u and v are random errors of measurement that are uncorrelated with each other and with their respective true scores and that have means of zero. Because the errors are uncorrelated with the true scores, the variances in Y and X can be represented as

$$s_Y^2 = s_y^2 + s_u^2, \tag{13a}$$

$$s_X^2 = s_x^2 + s_v^2. \tag{13b}$$

The EIV model assumes that the true-score relationship

$$y = \beta_0 + \beta_1 x, \tag{14}$$

is linear and that deviations of observed data points from the line are due entirely to the measurement errors u and v. The classical EIV model also assumes that the ratio of the error variances,

$$\delta = \frac{s_u^2}{s_v^2}, \tag{15}$$

is known, at least approximately (Carroll & Ruppert, 1996; Deming, 1943; Fuller, 1987).

Using these assumptions and the method of moments, Fuller (1987) developed a general maximum likelihood solution for EIV regression with the following estimates for the slope and intercept of the line:

$$\widehat{\beta}_1 = \frac{s_Y^2 - \delta s_X^2 + \left[\left(s_Y^2 - \delta s_X^2\right)^2 + 4\delta s_{YX}^2\right]^{\frac{1}{2}}}{2s_{YX}}, \tag{16}$$

$$\widehat{\beta}_0 = \overline{Y} - \beta_1\overline{X}, \tag{17}$$

where s_{YX} is the covariance between Y and X. The likelihood is maximized by taking the positive square root in Equation 16 (Fuller, 1987; Sprent, 1969), thereby ensuring that the slope has the same sign as s_{YX}.

The EIV model includes several regression models as special cases. If one assumes that the error variance for Y is zero, and that the error variance for X is greater than zero, δ would be zero, and taking the positive root in Equation 16, the estimated slope would reduce to

$$\widehat{\beta}_1 = \frac{s_Y^2}{s_{YX}} = \frac{s_Y^2}{r_{YX}\, s_Y\, s_X} = \frac{s_Y}{r_{YX} s_X}. \tag{18}$$

With this expression for β_1 and the expression for β_0 in Equation 14, we have the fitted line for the OLS regression of X on Y in Equation 7. If we reverse the roles of Y *and* X in Equation 16 and assume that the error variance for X is zero and that the error variance in Y is greater than zero, and we take the positive root, we can derive the fitted line for the OLS regression of Y on X in Equation 4. In the intermediate cases where both error variances are greater than zero, the EIV line will lie between the Y-on-X OLS line and the X-on-Y OLS line, and all three lines will pass through the centroid.

We can derive the GM line as a special case of the EIV model by assuming that the ratios of the error variances to observed-score variances (or equivalently, their reliabilities) are equal for Y and X:

$$\frac{s_u^2}{s_Y^2} = \frac{s_v^2}{s_X^2}. \tag{19}$$

Under this assumption,

$$\delta = \frac{s_u^2}{s_v^2} = \frac{s_Y^2}{s_X^2}. \tag{20}$$

Substituting this expression for δ in Equation 16 and simplifying, we get

$$\widehat{\beta}_1 = \frac{S_Y}{S_X}. \tag{21}$$

Using this expression for the slope and Equation 17 to define the intercept, we get the GM line given by Equation 11.

Estimating the Ratio of Error Variances for the Errors-in-Variables Model

In estimating the ratio δ, it might seem simple to estimate the two error variances and then take their ratio, but this is generally not feasible in cases where there is substantial natural variability. To the extent that the variability around the line is due to natural variability, δ is not well defined, because natural variability is not uniquely associated with either Y or X and therefore cannot be unambiguously assigned to the numerator or denominator of δ.

In some special cases, it may be possible to estimate δ directly, without estimating the error variances for Y and X separately. For example, the different forms of a standardized test are designed to have the same length, item types, content categories, and structure, and therefore it would be reasonable to assume that the reliabilities of the different forms would be approximately equal. The cases in which this kind of reasoning can be used to estimate δ are limited, and as shown earlier, if Y and X have the same reliability, the EIV model reduces to the GM model, and we do not need an estimate of δ. In most cases, the use of the classical EIV model will require separate estimates of the error variances in Y and X, which can then be used to estimate δ (Carroll & Ruppert, 1996).

The error variances associated with physical measurements can often be estimated directly by repeating the measurement several times on a sample of objects, organisms, or systems and estimating the variance of the observed values of the measurements for each object or system (Carroll & Ruppert, 1996). In education and the social sciences, where we cannot generally repeat measurements (e.g., using standardized tests or survey instruments) on a person more than once or twice without changing the value of the construct being measured, errors of measurement are usually evaluated less directly, in terms of reliability coefficients or generalizability coefficients (Brennan, 2001a, 2001b; Haertel, 2006). A reliability coefficient (or generalizability coefficient) of a measure for a particular population is defined as the ratio of the true-score variance (or, in G theory, the universe-score variance) to the observed-score variance:

$$r_{XX'} = \frac{s_x^2}{s_X^2} = 1 - \frac{s_u^2}{s_X^2}, \tag{22a}$$

$$r_{YY'} = \frac{s_y^2}{s_Y^2} = 1 - \frac{s_v^2}{s_Y^2}, \tag{22b}$$

where $r_{XX'}$ is the reliability (or generalizability) of X and $r_{YY'}$ is the reliability (or generalizability) of Y. The details of these analyses (using correlations or analysis of variance with multiple sources of error) can get quite complicated, but the basic idea is to treat variability over replications of the measurement procedure as error and each person's average score over the conditions of observation as their true score. The independent replications of the measurement procedure could include multiple sources of error, such as assessment tasks, occasions, contexts, and raters (Brennan, 2001b; Cronbach et al., 1972; Haertel, 2006). Estimates of the error variances for Y or X can then be obtained using the reliability coefficient and the estimated observed-score variances using Equations 22a and 22b.

If the EIV model fits the data, and the measurement errors in Y and X are specified accurately and completely in the reliability coefficients, the following relationship must hold:

$$r_{XY}^2 = r_{XX'} r_{YY'}. \tag{23}$$

In general, the squared correlation between Y and X should be less than or equal to the product of their reliabilities because of natural variability, but if the EIV's assumption is that all the variability around the line is due to the errors of measurement in Y and X, Equation 23 should hold exactly, except for uncertainty in the estimates of the correlation and the reliabilities. The correlation r_{XY} can typically be estimated directly from the data. The reliabilities will usually have to be evaluated in separate studies in which the measurements are repeated on the same objects, persons, or systems, but under different allowable conditions of observation (e.g., different occasions, raters, test forms, or, for coefficient alpha, test items).

The difference between the correlation expected on the basis of Equation 23 and the observed correlation is given by

$$D_{EE} = r_{XX'} r_{YY'} - r_{XY}^2. \tag{24}$$

To the extent that the variability around the line is due to natural variability as well as measurement errors, the squared correlation between the variables will be less than the product of the reliabilities, and D_{EE} will be positive. If D_{EE} is substantial, we have a clear indication that there is substantial natural variability in the data; and therefore that the EIV model does not fit the data or that the reliability of Y or X, or both, has been overestimated; and therefore that the value of δ is ambiguous at best.

If D_{EE} is large after a thorough evaluation of the measurement errors in Y and X, including as many potential sources of error in Y and X as possible, we may need to forgo the use of the classical EIV model or make do with fairly rough estimates of δ (Carroll & Ruppert, 1996; Fuller, 1987; Kane, 2021; Sprent, 1990), with suitable caution in interpreting the results obtained from the EIV regression equation. Note that in applications where the goal is to represent a general trend in a population, the natural variability can account for most of the variability around the line, and measurement errors, as such, may play a minor role (Draper, 1991; Kruskal, 1953; Ricker, 1973).

If Equation 23 does not hold in a case where we might otherwise rely on the EIV model, we have reason to think that something is amiss, and we have several possibilities to consider. First, if a plot of the data suggests that the relationship is not linear, it might be necessary to use a more complex model (Carroll & Ruppert, 1996; Fuller, 1987). Second, there may be additional, unidentified variables that have a substantial impact on the relationship, and if these additional variables can be identified, they could be included in a more complex model.

Third, if the analyses of the measurement errors in Y and X do not include some sources of measurement error that are in play in the data under consideration (Brennan, 2001a, 2001b; Carroll & Ruppert, 1996; Haertel, 2006), the errors of measurement in Y and X will be underestimated, and the reliabilities will be overestimated. An analysis that addresses only one or two of the relevant sources of error is likely to yield an underestimate of the error of measurement and, as a result, an overestimate of the natural variability. In these cases, unidentified and unestimated measurement error will be indistinguishable from other sources of natural variability.

Difficulties in Applying the Errors-in-Variables Model

In some cases, especially in physical sciences, the underlying relationship may be known to be linear (at least over some range of values for the two variables), almost all of the variability around the line is attributable to measurement error, and δ is known, at least approximately. In these cases, we can use the EIV model to estimate the slope of the line representing the relationship. These conditions generally cannot be assumed to hold in education, the social sciences, and many parts of the physical sciences (e.g., in astronomy, geology, or biology) because of the inherent natural variability in the populations being studied (Brace, 1977; Isobe et al., 1990; Ricker, 1973). In these cases, the goal is to estimate a regression line that provides a good estimate of the relationship between the two variables (Draper, 1991; Kermack & Haldane, 1950; Kruskal, 1953; Ricker, 1973; Sprent, 1990).

Ricker (1973) has suggested that in cases where natural variability predominates, "the sources of variability around the line are almost wholly natural, in the sense of not being a part of the measuring process and hence not subject to partition into separate moieties" (p. 414). Ricker went on to say,

> There seems no way, even conceptually, of separating the variable element in the position of a given point into two components, one due to variation in Y at a given X and one due to variation in X at a given Y. Thus the whole idea of two "point variances" and their ratio … becomes meaningless here. (p. 413)

Draper (1991) suggested that, as "practical advice," the maximum likelihood solution in Equation 16 be used if δ "is known (or can reasonably be estimated)" (p. 6) and otherwise suggested that the GM model be used.

Carroll and Ruppert (1996) suggested that the natural variability be assigned to the Y variable in computing δ, but this would destroy the symmetry of the analysis, and therefore this approach would not be appropriate if the goal were to represent a functional relationship. If we assign the variance associated with natural variability to the Y variable, the regression line will be shifted toward the Y-on-X OLS line, and if we assign the variance associated with natural variability to the X variable, the line will be shifted toward the X-on-Y OLS line. As a result, the Y-on-X line will be different from the X-on-Y line. Assigning the natural variability to the Y variable would be appropriate if one's primary goal were to estimate Y contingent on X, but it is not appropriate in estimating a functional relationship.

In the next section, I develop a hybrid model that is appropriate for estimating functional relationships in applications that contain substantial natural variability and substantial errors of measurement. This hybrid model treats the variance due to natural variability as an additional source of random error assigned to Y and X in a way that satisfies the condition in Equation 23 exactly and preserves the symmetry of the functional relationship.

The Hybrid Model: Modeling Measurement Error and Natural Variability

By definition, the sources and composition of the natural variability are unknown, but its overall effect is to add uncertainty to the estimation problem. Natural variability is apparently random variability in the coordinates of the data points that is not associated with the estimated errors of measurement in Y or X.

The natural variability can be represented by two additional error terms, one assigned to Y and one assigned to X:

$$Y = y^* + k + u, \tag{25a}$$

$$X = x^* + l + v, \tag{25b}$$

where y^* and x^* are adjusted true scores for Y and X, respectively; u and v are the original estimated measurement errors in Y and X, respectively; and k and l are additional, postulated error terms in Y and X, respectively, that have means of zero and are assumed to be uncorrelated with each other, with y^* and x^*, and with u and v.

Like the classical EIV model, the hybrid model assumes that the underlying relationship between Y and X is linear,

$$y^* = \beta_0 + \beta_1 x^*,$$

but allows for the possibility that the observed values of the two variables can contain both errors of measurement u and v and additional sources of error k and l due to natural variability. Unlike the classical EIV model, the hybrid model assumes that estimates of the measurement error variances are known, based on prior reliability studies; this assumption is slightly stronger than the corresponding assumption for the EIV model, which is that the ratio of the measurement error variances is known. As noted earlier, to apply the classical EIV model, it is generally necessary to estimate the two measurement error variances separately (see, e.g., Carroll & Ruppert, 1996).

Under these assumptions, the variances in Y and X can be represented as

$$s_Y^2 = s_{y*}^2 + s_k^2 + s_u^2, \tag{26a}$$

$$s_X^2 = s_{x*}^2 + s_l^2 + s_v^2. \tag{26b}$$

Note that the adjusted true-score variances are smaller than the true-score variances derived from the reliability analyses and that the adjusted error variances are generally larger than the measurement error variances from reliability analyses. We can think of the adjustments for natural variability as an additional subtraction from the observed-score variance. The adjusted reliability coefficients are given by

$$r_{YY'}^* = \frac{s_{y*}^2}{s_Y^2} = \frac{s_Y^2 - s_k^2 - s_\mathrm{u}^2}{s_Y^2} = 1 - \frac{s_k^2 + s_u^2}{s_Y^2}, \tag{27a}$$

$$r_{XX'}^* = \frac{s_{x*}^2}{s_X^2} = \frac{s_X^2 - s_l^2 - s_v^2}{s_X^2} = 1 - \frac{s_l^2 + s_v^2}{s_X^2}. \tag{27b}$$

The adjusted error variances for Y and X are larger than they were before the adjustment, and the adjusted true-score variances and reliabilities are smaller than they were before the adjustment.

The empirical estimates of the reliability coefficients $r_{XX'}$ and $r_{YY'}$ (in Equations 22a and 22b) obtained from reliability or generalizability studies will generally be larger than the adjusted reliabilities in Equations 27a and 27b, because the natural variability is ignored by $r_{XX'}$ and $r_{YY'}$. In the hybrid model, the main role of the reliability coefficients in Equations 22a and 22b is to provide estimates of the measurement error variances in Y and X.

The ratio of the overall error variance assigned to Y, including measurement error as such and natural variation, to the overall error variance assigned to X is given by

$$\delta^* = \frac{s_u^2 + s_k^2}{s_v^2 + s_l^2}. \tag{28}$$

Unfortunately, because the sources of the natural variability are not known, neither s_k^2 nor s_l^2 can be estimated directly, but they can be estimated indirectly by adding two very reasonable constraints to the model (see Deming, 1943, esp. pp. 7–8).

The first constraint requires that the relationship in Equation 23 is satisfied when the adjusted reliabilities in Equations 27a and 27b are substituted for the original reliabilities; that is,

$$r_{XX'}^* r_{YY'}^* = r_{XY}^2. \tag{29}$$

This constraint applies to the joint effect of the two natural-error terms, but it does not determine the values of the two extra error variances or the adjusted reliabilities uniquely. For example, we could achieve the constraint imposed by Equation 29 by assigning all the natural variability to the Y variable (and assuming that $s_l^2 = 0$) or by assigning all the natural variability to the X variable (and assuming that $s_k^2 = 0$). However, in either of these cases, the symmetry of the solution would be destroyed, and we would not have a functional relationship.

To get unique values for the two error variances associated with the natural variability and, thereby, the adjusted reliabilities, and to preserve the symmetry of the analysis, we need a second constraint. In estimating the line for a functional relationship, it would seem reasonable to assign the natural variability to Y and X equally or proportionally. Distributing the natural variability equally to the two variables makes the resulting line dependent on the scales of measurement for the Y and X variables. We can achieve both symmetry and scale independence by assuming that the variances associated with the natural variability are proportional to the observed-score variances.

Taking the extra error terms associated with natural variability to be proportional to the variances in Y and X, we have

$$s_k^2 = \Delta s_Y^2, \tag{30a}$$

$$s_l^2 = \Delta s_X^2. \tag{30b}$$

With this constraint, the error ratio for the Y-on-X EIV regression equation is

$$\delta^* = \frac{s_u^2 + \Delta s_Y^2}{s_v^2 + \Delta s_X^2}. \tag{31}$$

As shown in the Appendix, if the value of δ^* for the EIV regression of Y on X is equal to the inverse of the value of δ^* for the EIV regression of X on Y, the EIV regression line will be symmetric, and this is clearly the case for Equation 31.

Substituting Equations 30a and 30b into Equations 27a and 27b, we have expressions for the adjusted reliabilities in Y and X in terms of the original reliabilities, estimated in prior reliability studies:

$$r_{YY'}^* = \frac{s_Y^2 - s_u^2 - \Delta s_Y^2}{s_Y^2} = \frac{s_y^2 - \Delta s_Y^2}{s_Y^2} = r_{YY'} - \Delta, \tag{32a}$$

$$r_{XX'}^* = \frac{s_X^2 - s_v^2 - \Delta s_X^2}{s_X^2} = \frac{s_x^2 - \Delta s_X^2}{s_X^2} = r_{XX'} - \Delta. \tag{32b}$$

To determine an appropriate value for Δ, these adjusted reliabilities can be inserted into Equation 29,

$$\left(r_{XX'} - \Delta\right)\left(r_{YY'} - \Delta\right) = r_{XY}^2,$$

which can also be written as

$$\Delta^2 - \left[r_{XX'} + r_{YY'}\right]\Delta + \left[r_{XX'} r_{YY'} - r_{XY}^2\right] = 0.$$

Using the standard solution for a quadratic equation, the estimate of Δ is given by

$$\Delta = \frac{\left[r_{XX'} + r_{YY'}\right] \pm \sqrt{\left[r_{XX'} + r_{YY'}\right]^2 - 4\left[r_{XX'} r_{YY'} - r_{XY}^2\right]}}{2}. \tag{33}$$

Given this estimate of Δ, the value of δ^* can be estimated using Equation 31, and the result, δ^*, can then be substituted for δ in Equation 16 to provide an estimate of the slope of the hybrid line.

The expression in Equation 33 is somewhat complex, and its implications are not immediately clear. To get a sense of how it functions, we can consider three special cases. First, if there is no natural variability, the expression in the last set of brackets, which is equal to D_{EE}, is zero, and we have two solutions. For the positive square root, we have

$$\Delta = r_{XX'} + r_{YY'},$$

and for the negative square root, we have

$$\Delta = 0.$$

Assuming that there is no natural variability, there is no need for any adjustment, and the second solution makes sense. The first of these two solutions, which could result in a value greater than 1, does not make sense. Therefore $\Delta = 0$; δ^* reduces to its standard form for the EIV model,

$$\delta^* = \frac{s_u^2 + \Delta s_Y^2}{s_v^2 + \Delta s_X^2} = \frac{s_u^2}{s_v^2};$$

and the best-fitting line is the general solution for the EIV model, given by Equations 16 and 17.

Second, if there is no measurement error, and all the variability around the line is attributable to natural variability,

$$r_{XX'} = r_{YY'} = 1.0,$$

and

$$\Delta = \frac{2 \pm \sqrt{2^2 - 4\left[1 - r_{XY}^2\right]}}{2} = 1 - \left|r_{XY}\right|.$$

The proportion Δ has to be between 0 and 1, so we again take the negative sign for the square root. T he adjusted reliability of X for the case with no measurement errors in Y or X can be estimated as

$$r_{XX'}^* = \frac{s_X^2 - \Delta s_X^2}{s_X^2} = \frac{s_X^2 - \left(1 - \left|r_{XY}\right|\right) s_X^2}{s_X^2} = \left|r_{XY}\right|.$$

Similarly, the adjusted reliability for the Y variable is given by

$$r_{YY'}^* = \left|r_{XY}\right|,$$

and the condition in Equation 29 clearly holds. If there is no measurement error in Y or X, δ^* reduces to

$$\delta^* = \frac{\Delta s_Y^2}{\Delta s_X^2} = \frac{s_Y^2}{s_X^2},$$

and the best-fitting line is the GM line, given by Equation 11.

Third, if the reliability of X is equal to the reliability of Y, Equation 33 reduces to the same value for both variables:

$$\Delta = r_{YY'} - \left|r_{XY}\right| = r_{XX'} - \left|r_{XY}\right|.$$

Using Equations 32a and 32b, the adjusted reliabilities for the two variables would be

$$r_{XX'}^* = r_{XX'} - \Delta = \left|r_{XY}\right|,$$

$$r_{YY'}^* = r_{YY'} - \Delta = \left|r_{XY}\right|.$$

So, the constraint in Equation 29 holds.

The equality of the reliabilities for Y and X implies that the measurement error variances for Y and X are proportional to their observed-score variances and that the error terms associated with the natural variability are also proportional to their observed-score variances. Therefore the ratio in Equation 31 reduces to

$$\delta^* = \frac{s_Y^2}{s_X^2},$$

and the hybrid model reduces to the GM model.

The hybrid model imposes two constraints on the EIV model, the first of which is inherent in the definition of the EIV model, the second of which is a basic requirement for a symmetric, functional relationship. It reduces to the classical EIV model, when that model holds, and to the GM model when all the variability around the line is due to natural variability. It provides a reasonable solution to the problem of determining the line representing the functional relationship in cases that involve both substantial error of measurement in Y or X and substantial natural variability, and in these cases, it makes effective use of the information about errors of measurement that is available, while recognizing the impact of natural variability.

An Application of the Hybrid Model

Kane and Mroch (2010) analyzed data from the two components of the New York State Bar Examination, an essay test and an objective test, in terms of convergent validation. In this section, I extend some of those analyses using the hybrid model to examine differential test performance for men and women. In 2005, when the samples used by Kane and Mroch

Table 1 Summary Statistics for Women and Men on the New York State Bar Examination Essay Test and Multistate Bar Examination

	Essay (Y) scores			MBE (X) scores			
	M (Y)	SD (Y)	$r_{YY'}$	M (X)	SD (X)	$r_{XX'}$	r_{XY}
Women	736	68.3	0.8	143	14.5	0.9	0.64
Men	727	70.5	0.8	148	14.2	0.9	0.62

Note. MBE = Multistate Bar Examination.

were collected, the Multistate Bar Examination (MBE) was a daylong, multiple-choice test with 200 items, and the essay test included six constructed-response tasks and also took a day. MBE scores were reported on a 0–200 score scale, and the essay test scores were reported on a 0–1,000 score scale. The sample of first-time test takers in 2005 used in the study included 2,187 women and 2,201 men (Kane et al., 2006).

Although these two tests had different formats and different content specifications, they were both interpreted as measures of a candidate's ability to recognize and deal with realistic legal issues. Therefore they could be considered alternate measures of the same construct and thus suitable candidates for a convergent validity study. For similar reasons, these two measures can be evaluated in terms of differential test functioning between different groups (in this case, men and women).

As indicated in Table 1, the women had a mean score of 736 and a standard deviation of 68.3 on the essay test, and the men had a mean score of 727 and a standard deviation of 70.5 on the essay test. The women had a mean score of 143 and a standard deviation of 14.5 on the MBE, whereas the men had a mean score of 148 and a standard deviation of 14.2 on the MBE. So, the women scored higher on the essay test, and the men scored higher on the MBE. This kind of gender-based disparity typically occurs on the bar examination and in other examination programs (women doing better on essay tests and men doing better on objective tests).

The MBE had a reliability (Cronbach's alpha) of approximately .9, and the essay test had a reliability (also Cronbach's alpha) of approximately .8. These values are probably overestimates of the reliabilities as conceptualized in this report, because alpha defines measurement error in terms of person–item interactions and ignores other potential sources of measurement error (e.g., day-to-day variability in performance, context effects). Because the different essay questions were scored by different raters, the estimated alpha for the essay test includes variability due to rater effects. The reliability estimates of .9 and .8 for the MBE and the essay test, respectively, are used in the following calculations for both men and women because separate measures for women and men were not available.

The correlation between essay scores and MBE scores was .64 for women and .62 for men. Therefore, for women,

$$D_{EE} = r_{XX'} r_{YY'} - r_{XY}^2 = (0.9)(0.8) - (0.64)^2 = 0.72 - 0.41 = 0.31,$$

and for men,

$$D_{EE} = r_{XX'} r_{YY'} - r_{XY}^2 = (0.9)(0.8) - (0.62)^2 = 0.72 - 0.38 = 0.34.$$

These two values of D_{EE} are large enough to suggest that the relationship between the MBE scores and the essay scores contains substantial natural variability and therefore that these data would provide a useful example for how the hybrid model works. The product of the reliabilities is almost twice the squared correlation in each of the two cases.

Using Equation 33 and the summary statistics given earlier, Δ equals 0.21 for women and 0.23 for men. So, the hybrid model is assigning over 20% of the observed-score variances in Y and X to the error variances associated with natural variability, s_k^2 and s_l^2, in computing δ^*:

$$\delta^* = \frac{\left(1 - r_{YY}\right) s_Y^2 + \Delta s_Y^2}{\left(1 - r_{XX}\right) s_X^2 + \Delta s_X^2}.$$

For women, this would yield

$$\delta_W^* = \frac{(0.2 + 0.21) s_Y^2}{(0.1 + 0.21) s_X^2} = 1.32 \frac{(68.3)^2}{(14.4)^2} = 29.29,$$

and for men,

$$\delta_M^* = \frac{(0.2 + 0.23) s_Y^2}{(0.1 + 0.23) s_X^2} = 1.30 \frac{(70.5)^2}{(14.2)^2} = 32.04.$$

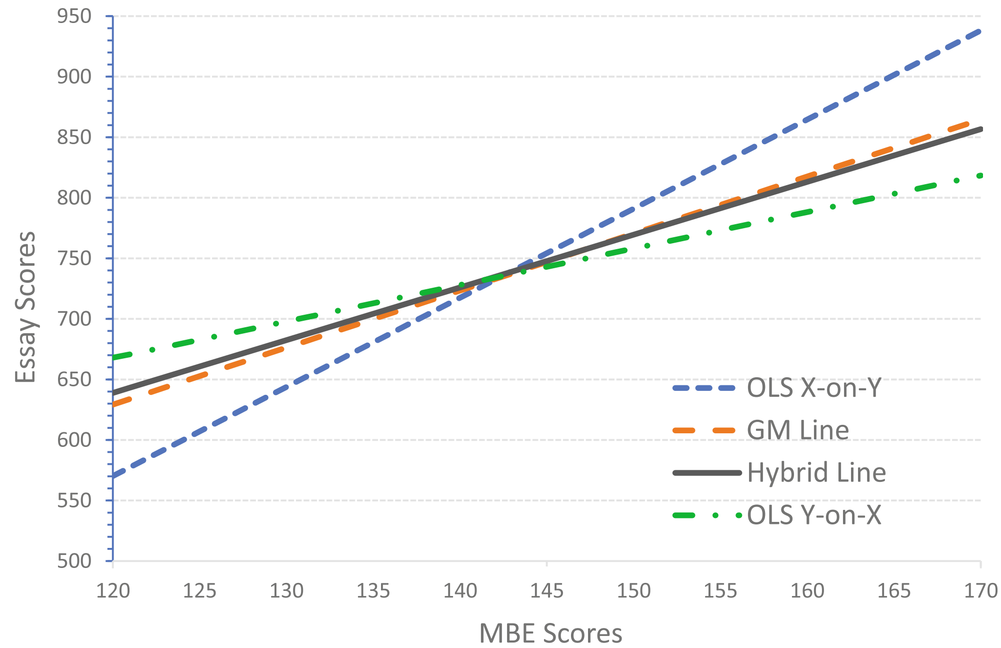

Figure 1 Regression lines for women.

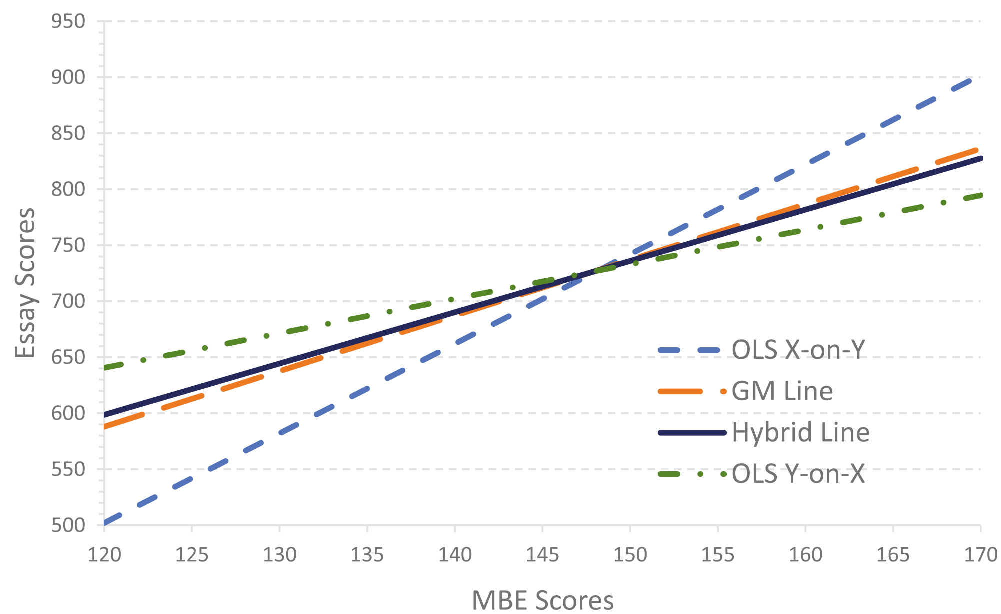

Figure 2 Regression lines for men.

Inserting these values for δ into Equation 16, we get slopes for the hybrid lines for the women and the men. For the women, the hybrid slope is 4.36. For the men, the hybrid slope is 4.59. Using Equation 17, the corresponding intercept is 112.5 for the women and 47.7 for the men. In their original analyses, Kane and Mroch (2010) used the GM model in analyzing these data, and in this section, their original results are compared to results using the hybrid model.

Figure 1 includes the OLS regression of essay scores on the MBE scores and the OLS regression of MBE scores on essay scores, as well as the GM and hybrid regression lines, for the sample of women. T he GM and hybrid lines are very close to each other and lie between the two OLS lines, and all four lines pass through the centroid (143, 736) for the women.

Figure 2 provides comparable results for the men. Again, the GM and hybrid lines are very close to each other and lie between the two OLS lines. As in Figure 1, all four of the lines pass through the centroid (148, 727) for the men, which is to the right and below the centroid for the women.

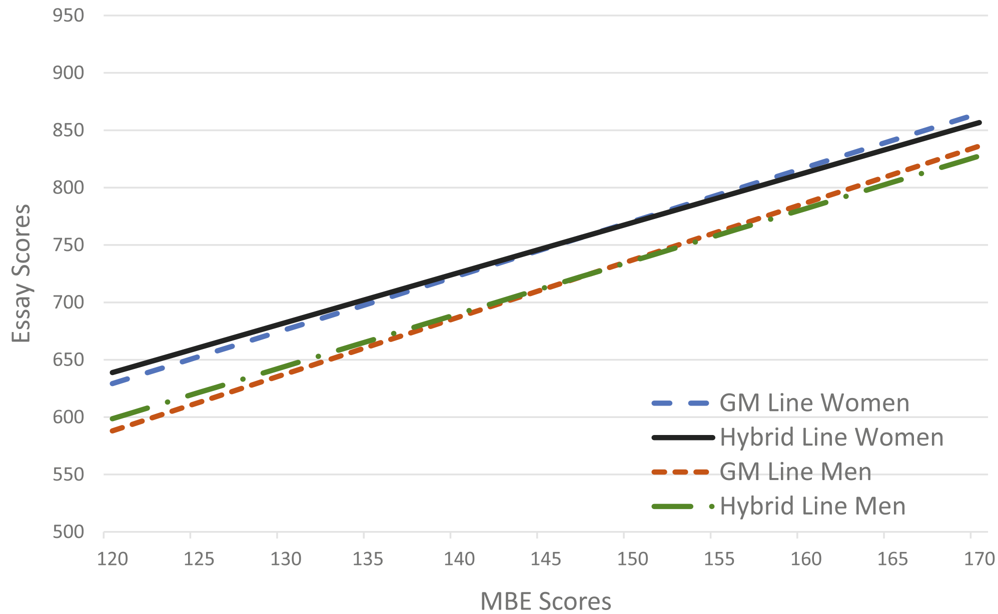

Figure 3 Geometric mean and hybrid lines for women and men.

Figure 3 presents the GM lines and hybrid lines for women and men. The two lines for the women are above the two lines for the men. The two lines for the women are close to each other; they have similar slopes (4.71 for the GM line and 4.36 for the hybrid line). Similarly, the two lines for the men are close to each other; they have similar slopes (4.96 for the GM line and 4.59 for the hybrid line).

These results are generally consistent with the results in the original analysis (Kane & Mroch, 2010), indicating that the functional relationship between the essay scores and the MBE scores is different for men and women, with the women doing better on the essay test and the men doing better on the MBE. For both the GM model and the hybrid model, the line representing the functional relationship for women is above the line for the men, and for both models, the two lines are more or less parallel. The two tests exhibit differential functioning for women relative to men, but it is not clear from this analysis which of the two tests to prefer as a measure of each candidate's ability to handle legal issues, and in practice, a weighted average of the two scores (plus one other score) was used (Kane et al., 2006).

In this example, the hybrid line is close to the GM line for both groups, but this will not always be the case. As noted earlier, if all the variability around the line is due to errors of measurement in Y and X, the hybrid model reduces to the classical EIV model. If most of the variability around the line is accounted for by the estimated errors of measurement, the hybrid line will be close to the EIV line that we would get by ignoring the natural variability. Under these circumstances, the hybrid line could lie anywhere between the OLS regression of Y on X and the OLS regression of X on Y, depending on the reliabilities of the two variables. If the measure of X were much more reliable than the measure of Y, the hybrid line would be close to the OLS regression of Y on X, and if the measure of Y were much more reliable than the measure of X, the hybrid line would be close to the OLS regression of X on Y. The GM line is always in the middle; on the z score scale, the GM line bisects the angle between the two OLS regression lines.

If most of the variability around the line were due to natural variability or the reliabilities of the two variables were more or less equal, the hybrid line would be close to GM line. For the example in this section, both these conditions hold to a large extent (i.e., much of the variability around the line was natural variability, and the two reliabilities were not very different), and therefore the hybrid line and the GM line are close to each other.

Conclusions

The EIV model provides an unbiased maximum likelihood estimate of the true-score relationship between two variables, assuming that the underlying relationship is linear, that all the variation around the regression line is due to errors of measurement, and that the ratio of the two error variances is known. In education and the social sciences, and in many parts of the physical sciences, these assumptions are highly unrealistic because much of the variability around the line tends to be due to natural variability (or equation error), which is not uniquely attributable to Y or X.

The assumption that the errors of measurement in Y and X account for all the variability around the line implies that the product of the reliabilities of Y and X should be equal to the squared correlation between the two variables, and therefore this relationship must hold, at least approximately, if the EIV model is to be appropriate in a particular case. In practice, this condition often does not hold, but it does provide us with a way to estimate the impact of the natural variability.

The hybrid model developed in this report provides a framework for addressing the impact of natural variability on estimates of a linear functional relationship by representing the natural variability as additional sources of random error in Y and X and thereby adjusting the reliabilities of Y and X. The hybrid model assumes that the relationship between the true scores on Y and X is linear; that each of the observed scores includes two sources of error, one due to errors of measurement and one due to natural variability; and that the measurement error variances in Y and X have been estimated and can be taken to be known. In addition, the hybrid model adds two reasonable constraints to the EIV model. The first constraint requires that the condition in Equation 29 hold for the adjusted reliabilities. To achieve this goal, the natural variability is distributed over the errors of measurement in Y and X, thereby decreasing the reliabilities in Y and X such that the constraint in Equation 29 is satisfied.

The second constraint requires that the regression equation for the functional relationship be symmetric. As shown in the Appendix, this second constraint is satisfied if the additional error variances assigned to Y and X are proportional to their observed-score variances. Imposing these two constraints, an adjusted value for the ratio of error variances for the EIV model is uniquely determined, and therefore the hybrid regression equation can be uniquely determined.

The hybrid model has several desirable properties. It maintains the symmetry inherent in functional relationships. It reduces to the classical EIV model if the variability around the line is due entirely to measurement errors in Y and X. It reduces to the GM model if all the variability around the line is due to natural variability, and it reduces to one of the OLS models if all the variability around the line can plausibly be attributed to one variable or the other.

The hybrid model can be applied in cases with both substantial errors of measurement in Y or X and substantial natural variability. In cases where most of the variability around the line is due to errors of measurement, the hybrid model provides an approximation to the classical EIV model with a reasonable adjustment for the natural variability. In cases where most of the variability in the data is due to natural variability, the hybrid model provides results that are similar to those of the GM model but also takes account of the relative sizes of the errors of measurement in Y and X.

References

Adcock, R. (1877). Note on the method of least squares. *The Analyst*, *4*(6), 183–184. https://doi.org/10.2307/2635777

Brace, R. A. (1977). Fitting straight lines to experimental data. *American Journal of Physiology*, *233*(3), R94–R99. https://doi.org/10.1152/ajpregu.1977.233.3.R94

Brennan, R. (2001a). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, *36*(4), 295–317. https://doi.org/10.1111/j.1745-3984.2001.tb01129.x

Brennan, R. L. (2001b). *Generalizability theory*. Springer. https://doi.org/10.1007/978-1-4757-3456-0

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. Guilford Press.

Carroll, R., & Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *American Statistician*, *50*(1), 1–6. https://doi.org/10.1080/00031305.1996.10473533

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Deming, W. E. (1943). *Statistical adjustment of data*. Dover.

Dorans, N. J. (2004). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement*, *41*(1), 43–68. https://doi.org/10.1111/j.1745-3984.2004.tb01158.x

Draper, N. (1991, April 28–30). *Straight line regression when both variables are subject to error*. Conference on Applied Statistics in Agriculture, Kansas State University, Manhattan, Kansas, USA. https://doi.org/10.4148/2475-7772.1414

Fuller, W. A. (1987). *Measurement error models*. Wiley. https://doi.org/10.1002/9780470316665

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, *15*(1886), 246–263. https://doi.org/10.2307/2841583

Haertel, E. H. (2006). Reliability. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). American Council on Education/Praeger.

Isobe, T., Feigelson, E., Akritas, M., & Babu, G. (1990). Linear regression in astronomy, I. *Astrophysical Journal*, *364*, 104–113. https://doi.org/10.1086/169390

Kane, M. (2021). *Symmetric least-squares estimates of functional relationships* (Research Report No. RR-21-21). ETS. https://doi.org/10.1002/ets2.12331

Kane, M., & Mroch, A. (2010). Modeling group differences in OLS and orthogonal regression: Implications for differential validity studies. *Applied Measurement in Education*, *23*(3), 215–241. https://doi.org/10.1080/08957347.2010.485990

Kane, M., & Mroch, A. (2020). *Orthogonal regression, the Cleary criterion, and Lord's paradox: Asking the right questions* (Research Report No. RR-20-14). ETS. https://doi.org/10.1002/ets2.12298

Kane, M., Mroch, A., Ripkey, D., & Case, S. (2006). *Impact of the increase in the passing score on the New York Bar Examination*. http://nybarexam.org/press/ncberep.pdf

Kermack, K., & Haldane, J. (1950). Organic correlation and allometry. *Biometrika*, *37*(1–2), 30–41. https://doi.org/10.1093/biomet/37.1-2.30

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Kruskal, W. (1953). On the uniqueness of the line of organic correlation. *Biometrics*, *9*(1), 47–58. https://doi.org/10.2307/3001632

Linn, R., & Werts, C. (1971). Considerations for studies of test bias. *Journal of Educational Measurement*, *8*(1), 1–4. https://doi.org/10.1111/j.1745-3984.1971.tb00898.x

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, *54*(285), 173–205. https://doi.org/10.1080/01621459.1959.10501505

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *6*(2), 559–572. https://doi.org/10.1080/14786440109462720

Ricker, W. (1973). Linear regressions in fishery research. *Journal of Fishery Research Board Canada*, *30*(3), 409–434. https://doi.org/10.1139/f73-072

Riggs, D., Guarnieri, J., & Addelman, S. (1978). Fitting straight lines when both variables are subject to error. *Life Sciences*, *22*(13–15), 1305–1360. https://doi.org/10.1016/0024-3205(78)90098-X

Sprent, P. (1969). *Models in regression and related topics*. Methuen.

Sprent, P. (1990). Some history of functional and structural relationships. In P. J. Brown & W. A. Fuller (Eds.), *Statistical analysis of measurement error models and applications: Proceedings of the AMS-IMS-SIAM joint summer research conference held June 10–16, 1989, Providence, RI* (pp. 3–15). AMS/IMS/SIAM.

Stevens, S. (1964). Concerning the psychophysical power law. *Quarterly Journal of Experimental Psychology*, *16*(4), 383–385. https://doi.org/10.1080/17470216408416398

Appendix

Symmetry and Scale Invariance

In evaluating OR models, it is desirable that we get the same line whether we label one variable as "X" in the analysis and the other one as "Y" or we switch the labels. As noted earlier, this property of *symmetry* is lacking in OLS regression, in which the Y-on-X line is generally quite different from the X-on-Y line.

It is not difficult to show that the GM and EIV models are symmetric. For any regression, the solution can be represented as

$$Y = \beta_1^{Y|X} X + \beta_0^{Y|X}. \tag{A1a}$$

If we reverse the roles of Y and X, we have

$$X = \beta_1^{X|Y} Y + \beta_0^{X|Y}, \tag{A1b}$$

and solving for Y, we get

$$Y = \frac{1}{\beta_1^{X|Y}} X + \frac{\beta_0^{X|Y}}{\beta_1^{X|Y}}.$$

If these two equations are to represent the same line, it is necessary that

$$1/\beta_1^{X|Y} = \beta_1^{Y|X},$$

or

$$\beta_1^{Y|X}\beta_1^{X|Y} = 1. \tag{A2}$$

As noted earlier, both the EIV line and the GM line pass through the centroid, and for any line passing through the centroid, the Y intercept is given by

$$\beta_0 = \overline{Y} - \beta_1\overline{X}. \tag{A3}$$

So, two lines with the same slope that both pass through the centroid will have the same Y intercept, and therefore it is necessary only to show that the Y-on-X regression has the same slope as the X-on-Y regression to demonstrate that the results are symmetric.

The GM model obviously satisfies the requirement in Equation A2,

$$\beta_1^{Y|X}\beta_1^{X|Y} = \frac{s_Y}{s_X}\frac{s_X}{s_Y} = 1, \tag{A4}$$

and therefore the GM model is symmetric.

Demonstrating symmetry for the general EIV model is a bit more complicated. The standard result for the EIV model has the general form

$$\beta_1^{Y|X} = \frac{s_Y^2 - \delta s_X^2 + \left[\left(s_Y^2 - \delta s_X^2\right)^2 + 4\delta s_{YX}^2\right]^{\frac{1}{2}}}{2s_{YX}}. \tag{A5}$$

But we can simplify this proof a bit by defining a quantity K^2 as (Riggs et al., 1978)

$$K^2 = \delta\frac{s_X^2}{s_Y^2} = \frac{s_u^2}{s_v^2}\frac{s_X^2}{s_Y^2} = \frac{s_u^2}{s_Y^2}\bigg/\frac{s_v^2}{s_X^2}. \tag{A6}$$

K^2 equals the ratio of the proportion of the variance in Y that is due to errors of measurement to the proportion of the variance in X that is due to errors of measurement.

Dividing the numerator and denominator of Equation A6 by s_Y^2 and replacing $\delta\frac{s_X^2}{s_Y^2}$ by K^2, the slope of the EIV line for Y given X can be written as

$$\beta_1^{Y|X} = \frac{1 - K^2 + \left[\left(1 - K^2\right)^2 + 4\delta s_{YX}^2/s_Y^4\right]^{\frac{1}{2}}}{2s_{XY}/s_Y^2}, \tag{A7}$$

and making use of the relationship between covariance and correlation,

$$s_{XY} = r_{XY}s_Xs_Y,$$

we have

$$\beta_1^{Y|X} = \frac{1 - K^2 + \left[\left(1 - K^2\right)^2 + 4\delta r_{YX}^2 s_X^2/s_Y^2\right]^{\frac{1}{2}}}{2r_{XY}s_Xs_Y/s_Y^2},$$

or

$$\beta_1^{Y|X} = \left[\frac{1 - K^2 + \left[\left(1 - K^2\right)^2 + 4K^2r_{YX}^2\right]^{\frac{1}{2}}}{2r_{XY}}\right]\frac{s_Y}{s_X}. \tag{A8}$$

Note that the expression within the square brackets in this equation depends on two quantities, r_{XY} and K^2, and that if K^2 equals 1 (i.e., Y and X have the same reliability), it reduces to 1, and the EIV slope reduces to the GM slope.

If we reverse the direction of the EIV regression to get an equation for X as a function of Y, the roles of Y and X are reversed, and we have the slope for the EIV regression of X given Y:

$$\beta_1^{X|Y} = \frac{s_X^2 - \delta' s_Y^2 + \left[\left(s_X^2 - \delta' s_Y^2\right)^2 + 4\delta' s_{YX}^2\right]^{\frac{1}{2}}}{2s_{YX}}, \tag{A9}$$

where δ' is given by

$$\delta' = \frac{s_v^2}{s_u^2} = \frac{1}{\delta}. \tag{A10}$$

So, the value of δ for the EIV regression of X on Y is the inverse of the value of δ for the EIV regression of Y on X. As shown shortly, this is a sufficient condition for the symmetry of the resulting regression equation.

The analog of K^2 for the regression of X on Y can then be written as

$$\delta' \frac{s_Y^2}{s_X^2} = \frac{s_u^2}{s_e^2}\frac{s_Y^2}{s_X^2} = \frac{1}{K^2}. \tag{A11}$$

Multiplying the numerator and denominator of Equation A9 by $1/s_X^2$ and using Equation A11, we can write the slope of the EIV regression of X on Y as

$$\beta_1^{X|Y} = \frac{1 - \frac{1}{K^2} + \left[\left(1 - 1/K^2\right)^2 + \left(4/K^2\right) r_{YX}^2\right]^{\frac{1}{2}}}{2r_{XY}} \frac{s_X}{s_Y}, \tag{A12}$$

$$\beta_1^{X|Y} = \frac{\frac{K^2-1}{K^2} + \left(\frac{1}{K^2}\right)\left[\left(K^2 - 1\right)^2 + 4K^2 r_{YX}^2\right]^{\frac{1}{2}}}{2r_{XY}} \frac{s_X}{s_Y},$$

$$\beta_1^{X|Y} = \frac{\left(K^2 - 1\right) + \left[\left(K^2 - 1\right)^2 + 4K^2 r_{YX}^2\right]^{\frac{1}{2}}}{2r_{XY}K^2} \frac{s_X}{s_Y}. \tag{A13}$$

Notice that the expression within the square root in Equation A13 is equal to the corresponding expression in Equation A8. We can represent the square root by

$$\Gamma = \left[\left(K^2 - 1\right)^2 + 4K^2 r_{YX}^2\right]^{\frac{1}{2}}. \tag{A14}$$

Then,

$$\beta_1^{X|Y} = \frac{-\left(1 - K^2\right) + \Gamma}{2r_{XY}K^2} \frac{s_X}{s_Y}, \tag{A15}$$

and

$$\beta_1^{Y|X} = \frac{\left(1 - K^2\right) + \Gamma}{2r_{XY}} \frac{s_Y}{s_X}. \tag{A16}$$

Multiplying the expressions in Equations 15 and 16, we have

$$\beta_1^{Y|X}\beta_1^{X|Y} = \left(\frac{\left(1 - K^2\right) + \Gamma}{2r_{XY}} \frac{s_Y}{s_X}\right)\left(\frac{-\left(1 - K^2\right) + \Gamma}{2r_{XY}K^2} \frac{s_X}{s_Y}\right), \tag{A17}$$

or

$$\beta_1^{Y|X}\beta_1^{X|Y} = \left(\frac{-\left(1 - K^2\right)^2 + \Gamma^2}{4K^2 r_{YX}^2}\right).$$

Using Equation A14, we get

$$\beta_1^{Y|X}\beta_1^{X|Y} = \left(\frac{-\left(1 - K^2\right)^2 + \left(K^2 - 1\right)^2 + 4K^2 r_{YX}^2}{4K^2 r_{YX}^2}\right) = 1. \tag{A18}$$

Therefore, if the condition in Equation A10 holds, the line representing the EIV regression of Y on X has the same slope and intercept as the line representing the EIV regression of X on Y, and the EIV model is symmetric. That is, Equation A10 is a sufficient requirement for the symmetry of the EIV regression model.

Scale Invariance

An expectation of any regression analysis is that the resulting line should not depend on the units (e.g., inches, meters, centimeters) for Y or X, as stated by Riggs et al. (1978),

> a fundamental requirement for a generally useful method of curve-fitting is that it must be invariant under all linear transformations. … It is a poor method indeed whose results depend upon the particular units which we happened to choose for measuring the variables. (p. 1313)

To say that a regression analysis relating Y and X is scale invariant is to say that if we apply linear transformations to X and/or Y and rerun the regression analysis, we would get the same result as we would get by applying the same transformations to the slope of the original regression equation.

It is easy to see that the GM model is scale invariant. Its slope is s_Y/s_X, which clearly has the required units. The scale invariance of the general EIV model is almost as obvious. Note that the quantity K is unitless and that the correlation r_{XY} is unitless, so the expression in square brackets in Equation A8 is unitless, and the units of the slope in Equation A8 are determined by the second term, s_Y/s_X, which has the units required for scale invariance.

Suggested citation:

Kane, M. (2023). *A hybrid model for orthogonal regression* (Research Report No. RR-23-04). ETS. https://doi.org/10.1002/ets2.12367

Action Editor: Michael Walker

Reviewers: Daniel F. McCaffrey and John Donoghue