ON THE USE OF ORDINAL AND DUMMY EXPLANATORY VARIABLES: AN ERRORS-IN-VARIABLES APPROACH*

RICHARD T. CARSON

RESOURCES FOR THE FUTURE
UNIVERSITY OF CALIFORNIA, BERKELEY

ABSTRACT

A commonly followed procedure in econometric research has been to dummy out a qualitative ordinal explanatory variable before entering it into a standard regression equation. This paper demonstrates using the framework of Aigner (1974) that this practice creates an errors-in-variables problem similar to entering the ordinal variable directly into a regression equation when the underlying "true" unobserved variable has an interval or ratio measurement scale. On the basis of a number of Monte Carlo experiments, we find that there is no a priori reason for favoring the dummy variable representation over the ordinal variable representation. Optimal transformations for ordinal proxy variables are discussed and suggestions are made on methods to assess and mitigate the errors-in-variables problem.

For presentation at The Econometric Society Winter Meeting, San Francisco (December, 1983).

1. Introduction

The dummy explanatory variable section of any standard econometrics text portrays one of two situations: either the need to dummy out a nominal level variable (e.g., war/no war, male/female) or the need to dummy out an ordinal (ranked categorical) level variable (e.g., education, income). This paper addresses the second situation and the implications of alternative methods of estimating regression equations when working with ordinal explanatory variables.

The question of whether to use an ordinal explanatory variable directly or to use its dummy variable representation[1] when estimating a regression equation has long been a topic of controversy in the other social sciences [e.g., Labovitz (1970); Wilson (1971); Bonrnstedt and Carter (1971); Kim (1975, 1978)]. The basic position of those advocating the use of ordinal variables in regression equations is that regression techniques are robust to most allowable (order preserving) transformations and thus the greater power of parametric statistics justified the use of ordinal variables. The opponents demonstrated that radically different results could be obtained using different allowable but extreme transformations. Lewis-Beck (1980) in a popular statistics series summarized current thinking when he compares two OLS equations (one in which an ordinal variable is entered directly and one in which its dummy variable representation is used) and concludes, "In this particular case, regression analysis with the ordinal variable arrives at the same conclusion as the <u>more proper</u> (emphasis added) analysis with dummy variables."

There is a fundamental problem with this whole debate. The presence of either an ordinal variable or its dummy variable representation is almost always[2] indicative of an errors-in-variables problem since the "true" underlying variable is usually measured on an interval (or ratio) scale. The errors-in-variables nature of this problem is clearly seen once it is realized that the "true" interval-level explanatory-variable in the model can be represented as a function of either the ordinal variable and an error term, or the dummy variable(s) and an error term. Both the ordinal and dummy variable representations may be thought of as proxies measured with error, for the "true" interval level variable in the model being estimated.

Econometric texts, in general, either offer no guidance to the researcher on what to do with ordinal explanatory variables, or tell the researcher to dummy the ordinal variable(s) out with little or no recognition of the errors-in-variables nature of the situation. The increasing use of micro surveys in economic research with their abundant ordinal categorical data make the implications of how one handles these ordinal proxies highly relevant to applied econometric research.

Fortunately, Aigner (1974) has provided a framework for comparing various methods for estimating regression equations when an ordinal proxy for a "true" unobserved-interval variable is available. Assuming no other information, such as variance ratios, covariances, or multiple equations, is available,[3] the options a researcher has, when confronted with the above situation, can be divided into three categories:

(1) estimate the model without a proxy for the unavailable interval explanatory variable,

(2) dummy out the ordinal proxy variable before estimating the model, or

(3) estimate the model using the ordinal variable directly.

This paper will show that the omission of any explanatory variable, using a dummy variable representation of the ordinal proxy variable or using the ordinal variable directly when the "true" model contains an unavailable interval level variable, results in similar (but potentially different magnitude) errors-in-variables problems. For the simple case where the "true" model consists of a constant term and two variables (one of which is interval level and unavailable), we will derive the mean square error for the estimated parameter of the variable observed without error and the $\bar{R}^2$ for the equation as a whole, for each of the three possible methods of estimating the equation. Finally, Monte Carlo results are presented for variants of the three cases with particular emphasis given to the effect of different numbers of ranked categories in the available ordinal variable. Violations of the assumptions made and optimal transformations are also considered.

2. Preliminaries

Consider the following "true" model:

$$Y_j = \beta_0 + \beta_1 X_j + \beta_2 Z_j + u_j; \tag{1}$$

where Y, X, and Z are variables measured on an interval or ratio scale with finite and observable first and second movements; u is an unobservable $\sim(0,\sigma_{uu})$ disturbance, distributed identically and independently of X and Z; and there are j=1, ..., n observations. Dropping the j subscript to avoid notational confusion and taking deviations from the means, equation (1) becomes

$$y = \beta_1 x + \beta_2 z + u \tag{2}$$

where small letters represent deviations from the means.[4]

The covariance matrix of x and z is given by

$$= E\left\{\begin{pmatrix} x \\ z \end{pmatrix} (x \ \ z)\right\} = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & \sigma_{zz} \end{pmatrix}. \tag{3}$$

Z is unavailable, but available as a proxy is an ordinal representation of Z, which we will call Z*. The researcher must choose: (1) whether to estimate the model with or without the proxy variable; and (2) if so, whether to use the ordinal variable Z* directly or to "dummy out" Z* using one or more dummy variables in the equation to be estimated.

3. Omitting a Proxy for Z

The first case to be considered is leaving Z* or its dummy representation out altogether. The model to be estimated is then

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X \tag{4}$$

when eq. (1) is the true model. We give the asymptotic $MSE(\hat{\beta}_1)_{OLS}$ developed by Aigner (1974) and $\bar{R}^2$ for eq. (4) below[5]

$$MSE(\beta_1) = \left(\frac{\sigma_{uu}}{n\sigma_{xx}} + \beta_2^2\left(\frac{\sigma_{xz}}{\sigma_{xx}}\right)^2\right) \tag{5}$$

and

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n} \hat{\beta}_2^2(\sigma_{zz}-\sigma_{xx})/n-2}{\sum_{j=1}^{n} y_j^2/n-1} \tag{6}$$

We will now assume that the decision to use the proxy variable in some form has been made based on the need for consistent prediction of Y, the desire for some notion of the significance and/or effect of Z in eq. (1), interests in minimizing bias $(\hat{\beta}_1)$, or interests in minimizing $MSE(\hat{\beta}_1)$. Use of a proxy variable, in the absence of any a priori information, to achieve the first two goals is fairly clear; use of a proxy variable to fulfill the last two goals is more open to question and circumstance.[6] McCallum (1972) and Wickens (1972) have demonstrated, using a number of restrictions on covariances between variables in the estimated equation, that the bias of $\hat{\beta}_1$ when the proxy variable is used is always less than or equal to the case where the proxy variable is not used. Since Aigner (1974) has shown that, except in a limited range of circumstances[7], that $MSE(\hat{\beta}_1)$ was smaller using the proxy variable than without it, we will assume that the researcher will want to consider one of the two proxy variable methods.

4. Direct Use of an Ordinal Proxy Variable

The relationship between Z and the ordinal variable proxy, Z*, is dependent upon the particular ranking or ordering system chosen for Z*. We will assume that this relationship is at least monotonic and further limit the class of possible transformation functions to those which are order preserving. It is important to note that the class of allowable transformations between Z and Z* is quite broad and includes linear, polynomial, logarithmic, exponential and others. The linear transformation is the easiest to work with, and specific results developed will apply only to that transformation. However, the general development and conclusions can be adapted to any admissible transformation. One of the more troublesome transformations, the exponential, is sketched out in the footnote below.[8]

Taking the linear case, we can represent the ordinal proxy, Z*, as

$$Z^* = \alpha_0 + \alpha_1 Z + e \ . \tag{7}$$

Z can be defined in terms of Z*

$$Z = \frac{-\alpha_0}{\alpha_1} + \frac{Z^*}{\alpha_1} - \frac{e}{\alpha_1} \ . \tag{8}$$

Equation (8) is unidentified, since α_1 is unknown; however, since the scale of Z is unobserved we can assume Z is scaled so that $\alpha_1 = 1$ without affecting the results to follow.[9] It is clear from these two equations that some ordinal representations of the same ranked categories will be better than others, while the dummy variable representation will be invariant to the particular ordinal representation.

The errors in variables problem becomes readily apparent when we substitute the right hand side of eq. (8) for Z in eq. (1),

$$Y = \beta_0 + \beta_1 X + \beta_2(-\alpha_0 + Z^* - e) + u \ . \tag{9}$$

$-\beta_2\alpha_0$ is a constant and the error terms $-\beta_2 e$ and u can be combined so

$$Y = (\beta_0 - \beta_2\alpha_0) + \beta_1 X + \beta_2 Z^* + (-\beta_2 e + u) \ , \tag{10}$$

which can be further simplified in notation by letting $\beta_0^* = \beta_0 - \beta_2\alpha_0$ and letting $v = -\beta_2 e + u$:

$$Y = \beta_0^* + \beta_1 X + \beta_2 Z^* + v \ . \tag{11}$$

The usual errors-in-variables result that $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are biased and inconsistent can be shown to hold.[10]

Taking deviations from the means eq. (11) can be written as

$$y = \beta_1 X + \beta_2 z^* + v \ . \tag{12}$$

Making explicit assumptions that $\sigma_{xe} = 0$, $\sigma_{eu} = 0$, and $\sigma_{ze} = 0$. The covariance matrix Σ_o of x and z* can be represented as

$$\Sigma_o = \begin{pmatrix} \sigma_{xx} & \sigma_{xz^*} \\ \sigma_{z^*x} & \sigma_{z^*z^*} \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & (\sigma_{zz}+\sigma_{ee}) \end{pmatrix}. \tag{13}$$

The $MSE(\hat{\beta}_1)_{OLS}$ derived by Aigner (1974) is

$$MSE(\hat{\beta}_1) = \frac{\sigma_{uu}}{n}\left(\frac{\sigma_{zz}+\sigma_{ee}}{\phi}\right) + \frac{\beta_2^2}{n}\left\{\frac{\sigma_{ee}+(\sigma_{zz}\sigma_{xx}-\sigma_{zx}^2)}{\phi}\right\}.$$
$$\cdot \left\{\frac{\sigma_{zz}+\sigma_{ee}}{\phi}\right\} + \beta_2^2\left(\frac{\sigma_{xz}\sigma_{ee}}{\phi}\right)^2 \tag{14}$$

where $\phi = \sigma_{xx}(\sigma_{ee}+\sigma_{zz}(1-\sigma_{xz}^2/\sigma_{xx}\sigma_{zz}))$.

We can simplify this expression in an instructive manner using Aigner's notation by defining $\lambda_o = \sigma_{ee}/(\sigma_{zz}+\sigma_{ee})$ which is the proportion of the variance in z* accounted for by measurement error (e); ρ_{xz}^2, as the squared simple (population) correlation coefficient between x and z; and $\Delta_o = \sigma_{ee}\sigma_{xx} - \lambda_o\sigma_{xz}^2$:

$$MSE(\beta_1) = \frac{\sigma_{uu}}{n\Delta_o} + \left\{\frac{\lambda_o\beta_2^2\sigma_{ee}^2\sigma_{xz}^2}{\Delta_o^2}\right\} \cdot \left\{\lambda_o + \frac{1-\rho_{xz}^2}{n\rho_{xz}^2}\right\} \tag{15}$$

With the choosing of any sample n, $\bar{Y}$, σ_{yy}, $\bar{X}$, σ_{xx}, r_{xz} the sample estimate of ρ_{xz}, $\hat{u}$, and $\hat{\sigma}_{uu}$ are determined; thus the one manipulatable feature of eq. (15) is the choice of the proxy variable Z*, which will affect $MSE(\hat{\beta}_1)$ through λ_o (or equivalently

σ_{ee}). To minimize MSE($\hat{\beta}_1$), Z* should be chosen to minimize σ_{ee}, which can be minimized by minimizing $\sum_{j=1}^{n} v_i$, where $v = -{}_2e + u$. Minimization of $\sum_{j=1}^{n} v_i$ is reflected in Theil's $\bar{R}^2$ statistic

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n} (-\beta_2 e+u)_j^2/n-3}{\sum_{j=1}^{n} y_j/n-1} , \tag{16}$$

which becomes

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n} (\beta_2^2\sigma_{ee}+\sigma_{uu})_j/n-3}{\sum_{j=1}^{n} y_j^2/n-1} . \tag{17}$$

It is obvious from eq. (17) that $\bar{R}^2$ will increase as σ_{ee} decreases, with $\bar{R}^2$ reaching its highest possible maximum (for the model in eq. (1)) when $\sigma_{ee} = 0$, at which point VAR(Z*) = VAR(Z).

5. The Dummy Variable(s) Proxy Case

Z can be represented by i - 1 (or fewer) dummy variables, D_i, a constant and an error term where i is the number of ranked categories present in ordinal variable Z*

$$(C_0+C_2D_2+C_3D_3+...+C_iD_i) = Z + \varepsilon , \tag{18}$$

or alternatively

$$Z = (C_0+C_2D_2+C_3D_3+...+C_iD_i) - \varepsilon . \tag{19}$$

We can simplify notation by letting $\tilde{D} = (C_2D_2+C_3D_3+\ldots+C_iD_i)$.

Substituting the right hand side of eq. (19) for Z in eq. (1) we have

$$Y = \beta_0 + \beta_1 X + \beta_2(C_0+\tilde{D}-\varepsilon) + u \tag{20}$$

combining terms

$$Y = (\beta_0+\beta_2 C_0) + \beta_1 X + \beta_2\tilde{D} + (-\beta_2\varepsilon+u) \tag{21}$$

which can be simplified in notation to

$$Y = \beta_0^d + \beta_1 X + \beta_2\tilde{D} + w \tag{22}$$

where $\beta_0^d = \beta_0 + \beta_2 C_0$ and $w = -\beta_2\varepsilon + u$.

The usual errors in variables result that all of the parameter estimates are biased and inconsistent can be shown to hold except for the special case described in footnote 10.

Taking deviations from the means eq. (22) can be written as

$$y = \beta_1 x + \beta_2\tilde{d} + w \tag{23}$$

where $\tilde{d} = \tilde{D} - E(\tilde{D})$.

Now making assumption about the covariances parallel to those made for the ordinal variable case in section 3 we can represent the covariance matrix Σ_d of x and $\tilde{d}$ as

$$\Sigma_d = \begin{pmatrix} \sigma_{xx} & \sigma_{xd} \\ \sigma_{dx} & \sigma_{dd} \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & (\sigma_{zz}+\sigma_{\varepsilon\varepsilon}) \end{pmatrix} \tag{24}$$

The $MSE(\hat{\beta}_1)_{OLS}$ from eq. (23) is

$$MSE(\hat{\beta}_1) = \frac{\sigma_{uu}}{n}\left(\frac{\sigma_{zz}+\sigma_{\varepsilon\varepsilon}}{\phi}\right) + \frac{\beta_2^2}{n}\left\{\frac{\sigma_{\varepsilon\varepsilon}(\sigma_{zz}\,\sigma_{xx}-\sigma_{zx}^2)}{\phi}\right\}$$

$$\cdot\left(\frac{\sigma_{zz}+\sigma_{\varepsilon\varepsilon}}{\phi}\right) + \beta_2^2\left(\frac{\sigma_{xz}+\sigma_{\varepsilon\varepsilon}}{\phi}\right)^2 \tag{25}$$

where $\phi = \sigma_{xx}(\sigma_{\varepsilon\varepsilon}+\sigma_{zz}(1-\sigma_{xz}^2/\sigma_{xx}\sigma_{zz})$.

Defining λ_d as $\sigma_{\varepsilon\varepsilon}/(\sigma_{zz}+\sigma_{\varepsilon\varepsilon})$ and λ_d as $\sigma_{\varepsilon\varepsilon}\sigma_{xx} - \lambda_d\sigma_{xz}$, eq. (25) can be rewritten as

$$MSE(\hat{\beta}_1) = \frac{\sigma_{uu}}{n\Delta_d} + \left\{\frac{\lambda_d\beta_2^2\sigma_{\varepsilon\varepsilon}^2\sigma_{xz}^2}{\Delta_d^2}\right\}\cdot\left\{\lambda_d + \left(\frac{1-\rho_{xz}^2}{n\rho_{xz}^2}\right)\right\} \tag{26}$$

As in the case of the direct use of the ordinal variable proxy, Z*, the parameters of eq. (1) are fixed, with the choice of any sample, except for λ_d (or equivalently $\sigma_{\varepsilon\varepsilon}$), which is determined by which proxy is dummied out, and how it is dummied out.

The $\bar{R}^2$ for eq. (23) is

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n}(-\beta_2\varepsilon+u)_j^2/n-2-(i-1)}{\sum_{j=1}^{n} y_j^2/n-1}, \tag{27}$$

or

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n} (\beta_2^2\sigma_{\varepsilon\varepsilon}+\sigma_{uu})_j/n-2-(i-1)}{\sum_{j=1}^{n} y_j^2/n-1} \quad . \tag{28}$$

$\bar{R}^2$ is maximized by minimizing $\sigma_{\varepsilon\varepsilon}$, and will reach the maximum possible for the model in eq. (1) when $\sigma_{\varepsilon\varepsilon} = 0$.

6. Comparison of the Three Cases

A comparison of directly entering the ordinal proxy variable and using a dummy representation of it when the true model is eq. (1) is now straight forward. Using eq. (9), the ordinal variable case, and eq. (20), the dummy variable case, it is possible to substitute the right hand side of eq. (9) for Y in eq. (20)

$$\beta_0 + \beta_1 X + \beta_2(-\alpha_0+Z-e) + u$$

$$= \beta_0 + \beta_1 X + \beta_2(C_0+\tilde{D}-\varepsilon) + u \tag{29}$$

Subtracting the common elements (β_0, β_1, X, and u) from both sides we get

$$\beta_2(-\alpha_0+Z^*-e) = \beta_2(C_0+\tilde{D}-\varepsilon) \tag{30}$$

By dividing both sides by β_2 and substituting for Z* in terms of Z and $(C_0+\tilde{D})$ in terms of Z we get

$$-\alpha_0 + \alpha_0 + Z + e - e = Z + \varepsilon - \varepsilon \qquad (31)$$

Since $E(e) = 0$ and $E(\varepsilon) = 0$, taking $E[Z^* - (C+\tilde{D})]$ yields

$$-\alpha_0 = E[Z^* - (C+\tilde{D})] \ . \qquad (32)$$

Thus, the effective measurement error, in the ordinal and dummy variable cases, differs by the term $-\alpha_0$ when the relationship between Z and Z* is linear. α_0 may take on both positive and negative values.

From the development of the three cases (omission of a proxy for Z, direct use of Z*, and the dummy variable representation of Z*) and given the assumptions made, we conclude

- A similar error is made when either an ordinal variable or a dummy variable representation is used as a proxy for a true interval variable. The same type of error is also made if a proxy variable is not used. The magnitude of the error is dependent on ρ_{xz} and on σ_{ee}, $\sigma_{\varepsilon\varepsilon}$, and σ_{zz}.

- Further, any method of dummying out a "true" interval explanatory variable is equivalent to dummying out any ordinal variable with those ranks.

- There is no way to know a priori what the relative magnitudes of σ_{ee}, $\sigma_{\varepsilon\varepsilon}$, and σ_{zz} are with out additional knowledge. However, σ_{zz} is by definition always greater than σ_{ee} or $\sigma_{\varepsilon\varepsilon}$.

- Hence, there is no a priori reason for preferring to dummy out the ordinal variable before using it as a proxy.

- $\bar{R}^2$ may be used to choose between different proxy variables and/or different representations of a proxy variable. Maximizing $\bar{R}^2$ will minimize VAR $(Y-\hat{Y})$ and also have the effect of minimizing $MSE(\hat{\beta}_1)$.

7. Consequences of Relaxing Assumptions

Let us relax some of the assumptions made:

(1) Suppose now that σ_{eu} is not equal to 0, $\bar{R}^2$ becomes

$$\bar{R}^2 = 1 - \frac{\sum_{j=1}^{n} (\beta_2^2\sigma_{ee}+2\sigma_{eu}+\sigma_{uu})_j/n-3}{\sum_{j=1}^{n} y_j^2/n-1} \tag{33}$$

The direction of a change in $\bar{R}^2$ with a change in σ_{ee} is no longer determinant being now also dependent on the sign and magnitude of $2\sigma_{eu}$.

(2) Suppose now that σ_{ze} is not equal 0, the variance-covariance matrix in eq. (13) becomes

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & (\sigma_{zz}+2\sigma_{ze}+\sigma_{ee}) \end{pmatrix} \tag{34}$$

with the result that $MSE(\hat{\beta}_1)$ is now also dependent on the sign and magnitude of $2\sigma_{ze}$.

(3) Suppose now that $Z^* = \alpha_0 + \alpha_1 X + \alpha_2 Z + e$ instead of $Z^* = \alpha_0 + \alpha_1 Z + e$. This assumption introduces a circular errors in variables argument, the effect of which in the first round is that σ_{xe} is no longer equal zero. The appropriate off diagonal variance-covariance element is now $(\sigma_{xz}+\sigma_{xe})$ instead of σ_{xe} with the size and magnitude of the off diagonal element now dependent on the size and magnitude of both covariants.

(4) Suppose now that the model contains non-linear terms, say $Y = \beta_0 + \beta_1 X + \beta_2(Z^2) + u$, or $Y = \beta_0 + \beta_1 X + \beta_2(\text{Log } Z) + u$. In

this case the best proxy, Z*, a transformation of Z*, or Z*'s dummy variable representation is the one that minimizes σ_{ee}, $\sigma_{\pi\pi}$, or $\sigma_{\varepsilon\varepsilon}$ respectively in the following equations for the log example

$$Z^* = \alpha_0 + \alpha_1(\text{Log } Z) + e \tag{35}$$

$$(\gamma_0 + \gamma_1 Z^*) = \alpha_0 + \alpha_1(\text{Log } Z) + \pi \tag{36}$$

$$(C_0 + \tilde{D}) = \text{Log } Z + \varepsilon \tag{37}$$

(5) In view of the possible deviations from the models developed above, an F test for the significance of the difference between the sums of squared error of two competing specifications should be conducted before a particular specification is rejected on the basis of a strict $\bar{R}^2$ comparison.

8. Monte Carlo Experiments: Design and Purpose

The primary purpose of the Monte Carlo experiments is to examine the difference between using an ordinal variable proxy, Z* for Z in eq. (1), and using a dummy variable representation of Z*. We have not attempted to look at the full spectrum of possible Z*'s (and relationships between Y, X, and Z), but have instead concentrated on examining the effects of the number of ranked categories in Z*, good and bad approximations[11] to the metric of Z by Z*, and cases of low and moderate correlation between X and Z. Some attention is paid to the situation in which the transformation between the true underlying continuous variable and its ordinal proxy is not an order preserving monotonic transformation.

Monte Carlo experiments (134) were conducted which can be divided into three groups:

(A) 42 experiments which examine the effects of altering the number of ranked categories (2-10) of Z* (and equivalently its dummy variable representations) on two different mapping schemes between Z and Z* when the correlation between X and Z is small (r_{XZ} = .088) and not statistically significant,

(B) 42 experiments parallel to (A) but with the correlation coefficient between X and Z of moderate size (r_{XZ} = .366),

(C) 50 experiments which examine in a much less detailed fashion the effects of a higher correlation (r_{XZ} = .729) between X and Z, sample size (25, 100, and 250), a logarithmic functional form for the underlying continuous variable, the addition of another continuous variable (X,Q,Z), and a nonmonotonic transformation.

In all of the experiments of groups A and B as well as the first thirteen of group C, the true model was constructed to be $Y = \beta_0 + \beta_1 X + \beta_2 Z + u$ where $\beta_0 = 100$, $\beta_1 = 2$, and $\beta_2 = 3$. In all of the other experiments $\beta_0 = 100$, and $\beta_1 = 2$. The true model, however, varies as Z enters the true model in a log form, in conjunction with an additional continuous variable X_2 and Z^2. Mean, bias, mean square error, mean absolute deviation (MASD), and variance[12] were measured for the parameter β_1. R^2 and $\bar{R}^2$ were calculated for the regression as a whole. There were 100 replications of each experiment.

The following conventions have been adopted for identifying experiments:

(1) OLS (k, l), ordinary least squares (OLS) regression on the variables k, l in the parentheses,

(2) io, OLS on (X, Z^*) where Z^* has i ranked categories (i = 2, 3, 4, 5, 6, 7, 8, 9, 10) which correspond to dividing Z into i equal distance categories (e.g. for 4o, $Z^* = 1$ if $Z < 25$; $Z^* = 2$ if $25 \leq Z < 50$; $Z^* = 3$ if $50 \leq Z < 75$; $Z^* = 4$ if $25 \leq Z$);

(3) id, OLS on $(X, D_2, \ldots, D_i)$ where there are i - 1 dummy variables corresponding to the ranked categories of the ordinal variable used in io (e.g. for 4d, $D_2 = 1$ if $Z^* = 2$ else $D_2 = 0$; $D_3 = 1$ if $Z^* = 3$ else $D_3 = 0$; $D_4 = 1$ if $Z^* = 4$ else $D_4 = 0$),

(4) iow, OLS on $(X, Z^*0$ where Z^* has i ranked categories with one ranked category representing the lower 70% of the range of Z and the other i - 1 ranked categories, dividing the upper 30% of the range of Z into equal distance categories (e.g. for 4ow, $Z^* = 1$ if $Z < 70$; $Z^* = 2$ if $70 \leq Z < 80$; $Z^* = 3$ if $80 \leq Z < 90$; $Z^* = 4$ if $90 \leq Z$),

(5) idw, OLS on $(X, D_3, \ldots, D_i)$ where there are i - 1 dummy variables corresponding to the i ranked categories of the ordinal variable used in iow (e.g. for 4dw, $D_1 = 1$ if $Z^* = 2$ else $D_2 = 0$; $D_3 = 1$ if $Z^* = 3$ else $D_3 = 0$; $D_4 = 1$ if $Z^* = 4$ else $D_4 = 0$).

The particular features and results of each of these experiments are described below.

Table I shows the group A experiments. The following features are common to each of these 42 experiments: $n = 100$, $r_{XZ} = .088$, X is normally distributed, Z is uniformly distributed between 0 and 100, and u is $N(0, \sigma_{uu})$.

Table II shows the group B experiments. The features are the same as the group A experiments except that $r_{XZ} = .366$

Table III shows 6 experiments similar to the group A and group B experiments except that $r_{XZ} = .729$ and only 4 ranked categories are used for those experiments using Z*.

Table IV shows 18 experiments where the sample size varies (25, 100, 250). r_{XZ} is approximately .5 in these experiments and results are given only for Z* with 4 ranked categories.

Table V shows 9 experiments where the true model was $Y = 100 + 2X + 40 (\text{Log } Z) + u$; $r_{XLogZ} = .353$. The experiments OLS (X only), OLS (X, Z) estimate the wrong functional form. 4o and 4ow use Z* directly (not Log Z*); 4lo and 4low use OLS to estimate $Y = \beta_0 + \beta_1 X + \beta_2 (\text{Log } (\alpha_0 + \alpha_1 Z^*)) + u$, where $Z = \alpha_0 + \alpha_1 Z^* + e$ and where Z* is 4o and 4ow, respectively.

Table VI shows the results of 6 experiments where the true model was $Y = 100 + 2X + 2Q + 3Z + u$; $r_{XQ} = .477$, $r_{XZ} = .313$, and $r_{QZ} = .270$. X and Q are continuous variables.

Table VII shows the results of 9 experiments where the true model is $Y = \beta_0 + \beta_1 X + \beta_2 Z^2 + u$ where $\beta_1 = 2$, $\beta_2 = 1$, $r_{XZ} = .336$ and $r_{XZ^2} = .245$. Z* is an ordinal representation of Z, however Z takes on negative values so that Z* (in the 4o experiment) is not an ordinal representation of Z^2 while Z* (4ow) is. In the 4os and 4ows experiments, it is assumed that the researcher knows the mean value of the Z* possessed and has a rough idea of the mean value of Z, and scales Z* to have the same mean as Z. In the 4os experiment Z* (4os) = 20Z*(4o), and in the 4os experiment Z* (4ows) = 20Z*(4ow).

Table VIII shows the result of 9 experiments similar (different error terms) to those in Table VII except that Z* had ten ranked categories instead of four. In the 10os experiment Z*(10os) = 10 Z*(10$_o$) and Z*(10ows) = 20Z*(10ow). In these experiments, r_{XZ} = .455 and $r_{XZ}2$ = .365.

Table 1

EXP	MEAN $\hat{\beta}_1$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
OLS (X only)	2.121	.121	.155	.318	.142	.1543	.1462
OLS (X, Z)	1.977	-.022	.140	.298	.141	.5113	.5013
2o	1.831	-.168	.166	.308	.140	.4302	.4184
2d	1.831	-.168	.166	.308	.140	.4302	.4184
2ow	2.114	.114	.154	.316	.142	.3923	.3798
2dw	2.114	.114	.154	.316	.142	.3923	.3798
3o	2.053	.053	.143	.298	.142	.4720	.4616
3d	2.047	.047	.166	.318	.166	.4769	.4661
3ow	1.994	-.005	.141	.290	.143	.3955	.3830
3dw	2.053	.053	.146	.304	.145	.4101	.3916
4o	2.014	.014	.140	.292	.142	.4873	.4767
4d	1.981	-.018	.142	.292	.143	.4990	.4779
4ow	1.936	-.063	.147	.292	.145	.3951	.3826
4dw	2.030	.030	.151	.312	.152	.4158	.3912
5o	1.868	-.131	.158	.303	.142	.4940	.4835
5d	1.864	-.135	.161	.301	.144	.5101	.4841
5ow	1.954	-.045	.144	.290	.144	.3890	.3764
5dw	2.045	.045	.151	.310	.153	.4228	.3921
6o	1.954	-.045	.142	.289	.141	.5003	.4900
6d	1.944	-.055	.170	.313	.169	.5227	.5081
6ow	1.897	-.102	.153	.297	.144	.3824	.3697
6dw	2.025	.025	.152	.307	.153	.4283	.3914
7o	2.046	.046	.142	.297	.141	.5039	.4936
7d	2.044	.044	.154	.311	.154	.5255	.4894
7ow	1.926	-.073	.148	.292	.145	.3794	.3666
7dw	2.045	.045	.158	.320	.157	.4288	.3786
8o	1.926	-.073	.146	.291	.142	.5055	.4953
8d	1.908	-.091	.161	.308	.154	.5364	.4957
8ow	1.900	-.099	.153	.297	.145	.3791	.3663
8dw	2.036	.036	.155	.313	.155	.4384	.3890
9o	1.915	-.084	.147	.292	.141	.5067	.4965
9d	1.894	-.105	.160	.306	.151	.5437	.4981
9ow	1.912	-.087	.151	.295	.145	.3817	.3690
9dw	2.035	.035	.161	.318	.162	.4383	.3821
10o	1.982	-.017	.140	.289	.142	.5077	.4965
10d	1.999	-.001	.172	.336	.173	.5492	.4986
10ow	1.912	-.087	.151	.295	.145	.3802	.3675
10dw	2.030	.030	.157	.317	.158	.4561	.3950

*$\beta_1 = 2$, $r_{XZ} = .088$, sample size = 100

Table 2

EXP	MEAN $\hat{\beta}_1$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
ols (X only)	2.695	.695	.510	.695	.026	.6390	.6353
ols (X, Z)	1.988	-.011	.035	.144	.035	.7655	.7607
2o	2.117	.117	.046	.176	.033	.7249	.7192
2d	2.117	.117	.046	.176	.033	.7249	.7192
2ow	2.300	.300	.124	.314	.034	.7148	.7089
2dw	2.300	.300	.124	.314	.034	.7148	.7089
3o	2.117	.117	.045	.179	.032	.7457	.7405
3d	2.119	.119	.052	.188	.037	.7486	.7407
3ow	2.605	.260	.103	.281	.036	.7205	.7147
3dw	2.644	.264	.105	.284	.036	.7200	.7112
4o	2.064	.064	.039	.160	.035	.7526	.7475
4d	2.049	.049	.038	.155	.036	.7581	.7479
4ow	2.248	.248	.099	.274	.038	.7101	.7042
4dw	2.255	.255	.103	.280	.038	.7222	.7105
5o	1.974	-.025	.038	.151	.038	.7554	.7503
5d	1.971	-.028	.038	.148	.037	.7631	.7505
5ow	2.260	.260	.104	.282	.036	.7095	.7036
5dw	2.258	.258	.104	.281	.037	.7258	.7112
6o	2.003	.003	.035	.147	.035	.7586	.7536
6d	2.000	.000	.042	.160	.043	.7694	.7545
6ow	2.253	.253	.100	.275	.036	.7058	.6997
6dw	2.259	.259	.104	.285	.037	.7283	.7107
7o	2.040	.040	.035	.149	.034	.7604	.7554
7d	2.041	.041	.039	.157	.037	.7735	.7562
7ow	2.256	.256	.103	.279	.037	.7076	.7015
7dw	2.263	.263	.108	.287	.039	.7310	.7105
8o	1.977	-.022	.036	.146	.036	.7611	.7562
8d	1.959	-.040	.040	.154	.038	.7791	.7570
8ow	2.260	.260	.104	.282	.037	.6973	.6911
8dw	2.623	.262	.106	.286	.038	.7256	.7015
9o	1.969	-.030	.036	.148	.036	.7617	.7568
9d	1.959	-.040	.040	.154	.038	.7791	.7570
9ow	2.261	.261	.105	.283	.037	.6917	.6854
9dw	2.266	.266	.110	.293	.040	.7279	.7007
10o	2.001	.001	.035	.146	.035	.7549	.7498
10d	2.011	.011	.043	.169	.043	.7823	.7578
10ow	2.264	.264	.106	.286	.037	.7057	.6996
10dw	2.260	.260	.107	.288	.039	.7390	.7097

*$\beta_1 = 2$, $r_{XZ} = .366$, sample size = 100

Table 3

EXP	MEAN $\hat{\beta}_1$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
ols (X only)	3.760	1.760	3.215	1.760	.120	.4865	.4813
ols (X, Z)	1.986	-.014	.306	.438	.309	.5736	.5648
4o	2.294	.294	.388	.499	.304	.5736	.5648
4d	2.300	.300	.396	.502	.309	.5822	.5646
4ow	2.939	.939	1.071	.949	.192	.5232	.5134
4dw	2.910	.910	1.028	.922	.202	.5399	.5205

*β_1 = 2, r_{XZ} = .729, sample size = 100

Table 4

EXP	MEAN $\hat{\beta}_1$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
ols (X only)	2.835	.835	.908	.843	.214	.5292	.5244
ols (X, Z)	1.960	-.040	.327	.453	.328	.6697	.6629
4o	2.053	.053	.313	.434	.313	.6631	.6562
4d	2.109	.109	.343	.448	.334	.6918	.6788
4ow	2.266	.266	.392	.501	.324	.5925	.5841
4dw	2.200	.200	.448	.529	.412	.6344	.6190

* $\beta_1 = 2$, $r_{XZ} = .531$, sample size = 25

ols (X only)	2.695	.695	.524	.695	.042	.5613	.5568
ols (X, Z)	1.986	-.014	.055	.181	.055	.6749	.6682
4o	2.044	.044	.052	.179	.050	.6614	.6544
4d	2.066	.066	.068	.207	.064	.6708	.6569
4ow	2.246	.246	.120	.296	.060	.6276	.6199
4dw	2.253	.253	.124	.301	.061	.6406	.6255

* $\beta_1 = 2$, $r_{XZ} = .501$, sample size = 100

ols (X only)	2.747	.747	.576	.747	.017	.5782	.5739
ols (X, Z)	2.008	.008	.023	.119	.023	.6847	.6782
4o	2.087	.087	.030	.139	.022	.6717	.6649
4d	2.082	.082	.029	.139	.023	.6792	.6657
4ow	2.297	.297	.113	.303	.025	.6335	.6259
4dw	2.284	.284	.105	.291	.025	.6457	.6308

* $\beta_1 = 2$, $r_{XZ} = .532$, sample size = 250

Table 5

EXP	MEAN $\hat{\beta}_1$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
ols (X only)	2.258	.258	.099	.272	.033	.5787	.5744
ols (X, Z)	1.974	-.026	.050	.177	.049	.6118	.6038
ols (X, ln Z)	2.005	.005	.041	.160	.042	.6290	.6213
4o	2.021	.021	.050	.177	.050	.6106	.6026
4d	2.049	.049	.053	.183	.051	.6219	.6059
4ow	2.111	.111	.057	.191	.045	.5877	.5792
4dw	2.122	.122	.058	.198	.044	.5977	.5808
41o	2.044	.044	.046	.168	.045	.6140	.6060
41ow	2.108	.108	.057	.190	.045	.5891	.5806

*True model $Y = \beta_0 + \beta_1 X + \beta_2(\text{Log } Z) + u$, $\beta_1 = 2$, $r_{X\text{Log } Z} = .353$, sample size = 100

TABLE 6

EXP		MEAN $\hat{\beta}_i$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
OLS (X, Q only)	β_1	4.306	2.306	6.163	2.306	.853	.5386	.5291
	β_2	2.499	.499	.301	.499	.055		
OLS (X, Q, Z)	β_1	2.583	.583	.915	.583	.581	.7297	.7208
	β_2	2.019	.019	.053	.189	.053		
4o	β_1	2.575	.575	.860	.680	.535	.7097	.7008
	β_2	2.176	.176	.081	.236	.050		
4d	β_1	2.567	.567	.852	.676	.535	.7164	.7014
	β_2	2.255	.255	.110	.294	.046		
4ow	β_1	3.144	1.144	1.902	1.154	.600	.6690	.6587
	β_2	2.479	.479	.284	.480	.055		
4dw	β_1	3.097	1.097	1.599	1.097	.159	.6801	.6631
	β_2	2.454	.454	.267	.454	.061		

*True Model $Y = \beta_0 + \beta_1 X + \beta_2 Q + \beta_3 Z$, $\beta_1 = \beta_2 = 2$, $r_{XQ} = .477$, $r_{XZ} = .313$
$r_{QZ} = .270$, sample size = 100

Table 7

EXP	MEAN $\hat{\beta}_i$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
OLS (X only)	39.901	37.901	1441.834	37.901	5.438	.0599	.0503
OLS (X, Z)	-5.908	-7.908	70.271	7.908	7.805	.6777	.6710
OLS (X, Z^2)	.433	-1.566	8.861	2.363	6.474	.9781	.9786
4o	2.108	.108	6.350	1.946	6.339	.6136	.6056
4d	11.163	9.163	90.221	9.169	6.325	.7190	.7071
4ow	12.259	10.259	111.280	10.259	6.086	.8251	.8216
4dw	9.181	7.181	58.608	7.181	7.118	.8456	.8391
4os	4.367	2.818	11.760	2.367	6.219	.6831	.6765
4ows	13.857	11.857	146.625	11.857	6.107	.8357	.8324

*True Model $Y = \beta_0 + \beta_1 X + \beta_2 Z^2 + u$, $\beta_1 = 2$, $r_{XZ}2 = .245$, sample size = 100.

Table 8

EXP	MEAN $\hat{\beta}_i$	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
OLS (X only)	87.375	85.375	7289.796	85.375	.990	.1385	.1297
OLS (X, Z)	.1014	-1.899	6.433	1.932	2.858	.6972	.6909
OLS (X, Z^2)	2.619	.619	1.932	1.17	1.564	.9891	.9889
10o	12.668	10.668	116.530	10.668	2.761	.5123	.5021
10d	-3.938	-5.938	38.116	5.938	2.882	.7751	.7495
10ow	-3.755	-5.755	34.950	5.755	1.879	.7547	.7496
10dw	2.928	.9278	1.942	1.131	1.092	.8599	.8440
10os	-3.833	-5.833	36.377	5.833	2.378	.6374	.6298
10ows	2.353	.235	2.049	1.420	2.226	.7995	.7954

*True Model $Y = \beta_0 + \beta_1 X + \beta_2 Z^2 + u$, $\beta_1 = 2$, $r_{XZ}2 = .365$, sample size = 100.

9. Discussion of Monte Carlo Results

Looking at Table 1, there is no real pattern to the size of the MSE($\hat{\beta}_1$) or to the bias, MASD, or variance of $\hat{\beta}_1$. All three cases (omission of a proxy, an ordinal proxy and a dummy proxy) have larger MSE's than the OLS estimate of the true model [MSE(β_1) = .14]. In the experiment where no proxy for Z was used, MSE(β_1) equaled .16. For the o series (i.e., 2o, 3o, 4o, 5o, 6o, 7o, 8o, 9o, 10o), MSE(β_1) ranged from .14 to .17 with no apparent pattern as the number of ranked categories in Z* increased. For the ow series, MSE(β_1) ranged from .14 to .15. For both the o and ow series bias, MASD, and variance of $\hat{\beta}_1$ followed patterns similar to MSE($\hat{\beta}_1$). For the d series, MSE(β_1) ranged from .14 to .17 with no apparent pattern as the number of dummy variables increased. For the dw series, MSE($\hat{\beta}_1$) ranged from .15 to .16. Fluctuations in bias, MSAD, and variance of $\hat{\beta}_1$ tended to follow MSE($\hat{\beta}_1$).

A comparison of the o and d series shows that the MSE($\hat{\beta}_1$) for the o series is less than or equal to the d series (for experiments with an equal number of ranked categories). Comparing the ow and dw series, with the exception of 6ow (MSE = .153) and 6dw (MSE = .152), the MSE($\hat{\beta}_1$) is smaller for the ow series. Overall the o series produced the smaller MSE($\hat{\beta}_1$) in six of the nine quadruples (e.g., 4o, 4d, 4ow, 4dw) of ranked categories, and the ow series had the smallest MSE three times.

$\bar{R}^2$ for the case where a proxy for Z was not used was only 29% of the OLS estimate of the true model. For the o series, $\bar{R}^2$ increased from .4184 (83% of that of that obtained by the OLS regression of the true model) to .4965 (99%) as the number of ranked categories increased. For the d series, $\bar{R}^2$ increased from .4184 (83%) to .4986 (99%) as the number of dummy variables increased. For the ow series, $\bar{R}^2$ showed no pattern as the number of ranked categories increased and varied from .3663 (73%) for 8ow to .3830 (76%) for 3ow. The dw series $\bar{R}^2$'s also showed no pattern as the number of dummy variables increased varying from .3786 (75%) for 7dw to .3950 (78%) for 10dw. The o and d series resulted in the highest $\bar{R}^2$ for each quadruple with the d series slightly higher. The ow and dw series were uniformly lower than the o and d series with the dw series resulting in higher $\bar{R}^2$'s than the ow series.

From these results, we conclude that if r_{XZ} is small and a researcher's sole interests are in minimizing MSE($\hat{\beta}_1$) that it does not matter how he or she estimates the model (leaving out a proxy, using the ordinal variable proxy directly, or using a dummy variable proxy). Nor does it matter if Z* represents a good approximation to the metric of Z. If $\bar{R}^2$ is a consideration then the model should be estimated with a proxy, and the better the proxy the closer the $\bar{R}^2$ will be to the $\bar{R}^2$ of the OLS estimate of the true model for both the ordinal and dummy variable cases. An alternative prediction criteria to $\bar{R}^2$ such Amemiya's prediction criteria (Judge et al., 1980), which imposes a greater penalty on the number of variables, will tend to favor use of the ordinal variable proxy.

With a moderate correlation between X and Z (Table II), the o and d series have both smaller MSE($\hat{\beta}_1$) and higher $\bar{R}^2$ than comparable experiments of the ow and dw series. Omission of a proxy for Z produces inferior results on both a MSE [MSE(β_1) = .26] and $\bar{R}^2$ basis compared to any experiment using a proxy. The o series had the lowest MSE($\hat{\beta}_1$) except for 4o where 4d's MSE($\hat{\beta}_1$) was slightly lower. The o series had MSE(β_1) ranging from slightly below .02 to slightly above .02. There was a tendency for the MSE($\hat{\beta}_1$) to decline as the number of categories increased. The d series smallest MSE was for 5d and largest MSE occurred at 3d. The ow series MSE($\hat{\beta}_1$) showed no pattern as the number of ranked categories increased ranging from a low of MSE(β_1) = .05 for 4ow to a high of .06 for 2ow. The dw series also showed no pattern with regard to MSE as the number of categories increased, but was always slightly higher or equal to the comparable ow experiment.

$\bar{R}^2$ for the d series was always the highest ranging from .7192 (95% of attainable) for 2d to .7578 (99%) at 10d. The o series $\bar{R}^2$'s ranged from .7192 (95%) for 2o to .7498 (99%) for 10o. The ow series $\bar{R}^2$'s showed no pattern as the number of ranked categories increased ranging from .6854 (90%) for 9ow to .7147 (94%) for 3ow while the dw series had $\bar{R}^2$'s ranging from .7007 (92%) for 9dw to .7112 (94%) for 3dw. For the case where no proxy for Z was used the $\bar{R}^2$ was .6353 (84%).

Our earlier development indicated that $\bar{R}^2$ could be used to select the model with the smallest $MSE(\hat{\beta}_1)$. The results of Table 2 offer some support for that conclusion when the difference between the MSE of two competing models is large, the choice of the model with the highest $\bar{R}^2$ will be the one with the lower $MSE(\hat{\beta}_1)$. $\bar{R}^2$ is, however, insensitive to relatively small differences in $MSE(\beta_1)$. In particular on the $\bar{R}^2$ criteria, the dummy variable case will usually have a slightly higher $\bar{R}^2$ than the equivalent ordinal variable case while having a slightly larger $MSE(\hat{\beta}_1)$.

Table 3 displaying the results from experiments where r_{XZ} = .729 shows the increased importance of using a proxy variable (and a good proxy variable) when the correlation between X and Z is large. For the case where a proxy for Z was left out $MSE(\beta_1)$ is 1.61. For the o case, $MSE(\beta_1)$ = .19 and for the d case, $MSE(\beta_1)$ = .20 where $MSE(\beta_1)$ = .15 for the OLS estimate of the true model. For the ow and dw cases $MSE(\beta_1)$ was .54 and .51 respectively. The $\bar{R}^2$ measures for the 6 experiments in Table 3 correctly rank order the $MSE(\hat{\beta}_1)$ measures.

Table 4 shows the effect of different sample sizes. $MSE(\hat{\beta}_1)$ generally decreases as sample size increases due to the decrease in $VAR(\hat{\beta}_1)$, but will not converge to zero even as the sample size goes to infinity because the bias does not disappear. These limited results suggest that reductions in MSE ($\hat{\beta}_1$) with increased sample size are not large for samples of over 100 cases.

A logarithmic functional form ($Y = \beta_0 + \beta_1 X + \beta_2(\text{Log } Z) + u$), was examined in the experiments shown in Table 5. Here the lowest MSE (other than the true model) was achieved by the 4lo experiment where the ordinal proxy was Log ($\gamma_0 + \gamma_1 Z^*$), where γ_0 and γ_1 were determined by using OLS to regression Z* on Z. The next lowest MSE ($\hat{\beta}_1$) was for the 4o experiment which was an estimation of the wrong functional form. It is interesting to note here that while the range of Z was not large (0-100), 4lo and 4o both had smaller MSE than did 4d and that 4low and 4ow had smaller MSE($\hat{\beta}_1$) than did 4dw. While the $\bar{R}^2$ here would have allowed us to select the model with the lowest MSE($\hat{\beta}_1$), it would not have allowed us to rank order several of the other experiments where the differences in MSE were smaller.

The experiments in Table 6 simply introduce another continuous independent variable, Q. The results are similar to those in Table 2 where the level of correlation between Z and X was similar. Noteworthy perhaps is that 4o and 4d produce a $\hat{\beta}_1$ with a smaller MSE than the ols regressions on the true model although their estimate of $\hat{\beta}_2$ has a higher MSE. There is no real basis to choose between the 4o and 4d regressions since each performs better than the other on one of the two coefficients. Between the 4ow and 4dw experiments, 4dw performs marginally better. Leaving out a proxy for Z produces noticeably worse results. $\bar{R}^2$ is a fairly accurate guide to the results.

The experiments in Table 7 represent an intentional effort to show what might happen when the order preserving transformation rule was violated. Here the transformation between $Z \rightarrow Z^*$ is order preserving but it is Z^2 which is the variable in the true model. Keeping in mind how Z^* is created in the 4o case (4 equal sized categories), we note that most of the values of Z represented by $Z^* = 1$ are negative while those of $Z^* = 2, 3, 4$ are positive. $Z^*(4o)$ is not, however, an order preserving transformation of Z^2 since many of the values of Z^2 represented by $Z^* = 1$ are larger than those represented by $Z^* = 2$. For $Z^*(4ow)$ however this is not the case since $Z^* = 1$ contains the first 70% of the cumulative distribution of Z and the largest absolute value of Z contained in $Z^* = 1$ is smaller than the smallest absolute value in $Z^* = 2$. The experiments 4os and 4ows imitate the common practice of scaling an ordinal variable to have the same mean as the unobserved Z, when the mean value of Z is known from outside information.

Examining the results in Table 7, we are struck by a number of peculiar results, 4o produces the best estimate of β_1 although it has a lower R^2 than any case except the omitted variables case. The misspecified case (Z instead of Z^2) produces estimates that are wrong by a wide margin and worse than several of the cases which use an ordinal or dummy proxy. The 4o case as already noted is uniformly preferred to 4d while 4dw is preferred to 4ow. It is interesting to note that the 4os and 4ows transformations result in clearly inferior estimates of β_1. This result will have implications for our later discussion.

In Table 8, we display the results of experiments with the same basic parameters but different seed numbers for X, Z and the u's, except that we have created 10 ranked categories in Z*(10o) and Z*(10ow) and nine appropriate dummy variables (dropping the last category) for 10d and 10dw. Now, 10o is a much more distorted representation of the values of Z^2 while 10ow is still an order preserving representation of Z^2. The 10os and 10ows experiments represent transformations which scale Z*(10o) and Z*(10ow) to the approximate mean value of Z (keep in mind that the true variable is Z^2).

The results in Table 8 suggest that the good performance of 4o in Table 7 was to some degree a fluke of the particular scaling and categorization scheme used. In Table 8, on a MSE base the experiments can be ranked: ols(X, Z^2), 10dw, 10ows, ols(X,Z), 10ow, 10os, 10d, 10o, ols(X).

10. Optimal Transformation of Ordinal Variables

In the previous section, we gave the results of several experiments [Tables 5(4lo, 4low); 7(4os, 4ows) and 8(10os, 10ows)] where the ordinal proxy had been rescaled from the simple equal distance, 1, 2, 3... numerical assignment used in most of the experiments. In this section we take up the issue of how the optimal transformation of an ordinal proxy might be found. It is first necessary, however, to make clear the distinction between how well the ranked categories of the ordinal proxy divided the true unobserved variable's distribution up into equal distance intervals and the particular scheme used to assign numeric values to each of the ordinal categories. The researcher typically has no control over the categories in the available ordinal proxy but complete control over the value given to each of those categories.

The experiments reported have used several of the more commonly used methods of assigning numbers to ordinal categories. The io experiments have used a scheme equivalent to assigning the mean of the underlying values of Z represented by each category of Z*. The iow experiments conform to the practice of assigning equal distances between categories when those distances are unknown. The 4lo and 4low experiments in Table 5 assumed that the least squares transformation between Z and Z* (4o; 4ow) was known. The os and ows experiments in Tables 7 and 8 assumed that only the mean value of Z was known. There are obviously an infinite number of transformations which might be made.

It is clear that the optimal transformation of the ordinal proxy Z* must be defined in terms of the researcher's objectives. Those objectives can be seen to fall into three general categories:

(1) estimation of the parameter β_1, associated with the observed continuous variable, X, of interest,

(2) estimation of the parameter, β_2, associated with the unobserved Z or obtaining some idea of the significance and importance of $\beta_2 Z$ with respect to the dependent variable Y, or

(3) prediction of Y from the observed continuous variables X_i and the available proxy Z*.

If the researcher is willing to accept the assumptions of sections 3-6, then finding the transformation of Z* which maximizes $\bar{R}^2$ will be optimal for objectives (1) and (3) under a mean square error loss function. A number of suggestions for doing this have been made in the literature, Bonacich and Kirby (1976), de Leeuw et al. (1976) and Young et al. (1926). The technique proposed by Young et al. is the most general and allows almost any monotonic transformation of Z*.

In general, however, these assumptions often will not hold and we have demonstrated in Tables 7 and 8 the large distortions possible if the form of Z, say $\phi(Z)$, which is part of the true model, is not monotonic in the available Z*. Brieman and Friedman (1982) have recently proposed a nonparametric iterative method based on alternating conditonal expectations which minimizes

$$\frac{E\{[\theta(Y) - \sum_{j=1}^{p}\phi(X_j)]^2\}}{VAR\ [\theta(Y)]}, \qquad (38)$$

where there are p independent variables and $\theta(.)$ and $\phi_j(.)$ are transformations to be estimated. No restrictions are placed on the transformations $\theta(.)$ and $\phi_j(.)$ and the algorithm has been shown to converge under farily weak conditions. While the use of such a tool maximizes the predictive properties of $\hat{Y}$, the properties of the $\beta_1(X)$ obtained are unclear. Without additional information or the assumption of various restrictions, singular pursuit of the best estimator of β_1 appears to be impossible.

The second goal, estimation of $\beta_2 Z$, is frought with difficulties. The standard practice has been for researchers to assess the direction and significance of β_2 from the significance and sign of the ordinal or dummy variable proxy(s). If the ordinal proxy is similar to the io proxy used in Tables 1, 2 and 3 and the true model is similar to the one in those tables, this practice has some merit as a rule of thumb, although strict hypothesis testing is invalid. We only need to turn to any of the ow experiments to begin finding t-statistics which bear no resemblance to those obtained by estimating the true model.

Dummy variable proxies do not force the researcher to choose a scaling scheme for the ordinal proxy and instead allow the data to choose the transformation (not necessarily order-preserving) which maximizes R^2. This property lies behind much of the popularity of using dummy variable proxies instead of an ordinal variable proxy. When the assumptions of sections 3-6 are fulfilled and the number of categories to be dummied out small, the practice of using dummy proxies comes close to achieving the same result as the optimal transformation in terms of minimizing MSE ($\hat{\beta}_1$) and the mean square error of the regression.

The transformation implied by the dummy variable proxies is, however, always readily available to the researcher as an ordinal scaling.[13] This scaling may be obtained by estimating

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{C}_2 D_2 + \hat{C}_3 D_3 + \ldots + \hat{C}_i D_i + w \qquad (39)$$

and forming the vector,

$$Z^{*D} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hat{C}_2 \\ \vdots \\ \hat{C}_2 \\ \vdots \\ \hat{C}_i \end{bmatrix} \qquad (40)$$

where zero replaces the category represented by the omitted dummy variable and the C_i's obtained from equation (39) replace their respective categories. Z^{*D} can then be used in estimating,

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 Z^{*D} + w \qquad (41)$$

$\hat{\beta}_2$ will always equal one and the $\hat{\beta}_1$ obtained from equations (39) and (415) will be equal. This result always holds since $Z^{*D} = \hat{C}_2 D_2 + \ldots + \hat{C}_i D_2$. An example of using Z^{*D} is given in Table 9, experiment 4owsd. Notice that the 4dw and 4owsd experiment are comparable except for slight differences in the VAR($\hat{\beta}_1$) and a slightly higher $\bar{R}^2$ in the 4owds experiment.

A reasonable question to ask is whether this increase in $\bar{R}^2$ is worth the expense involved in estimating the regression equation twice. To answer the question, we must point out that $\bar{R}^2$, while correcting for the degrees of freedom problem that R^2 suffers from, does not account for the random error introduced into the coefficients of the regression equation by adding additional variables to that equation. Breiman and Freidman (1983) have recently unified a large literature on the optimal number of regressors to retain and have shown that minimizing the unconditional mean square error prediction criteria can be fulfilled by minimizing,

Table 9

EXP	MEAN β_1	BIAS	MSE	MASD	VAR	R^2	$\bar{R}^2$
OLS (X only)	44.424	42.424	1799.929	42.424	.182	.3295	.3226
OLS (X, Z)	4.157	2.157	4.857	2.157	.206	.4144	.4023
OLS (X, Z^2)	2.429	.429	.497	.617	.316	.9971	.9470
4o	25.614	23.615	557.707	23.615	.064	.3588	.3456
4d	14.371	12.371	153.136	12.371	.088	.6614	.6472
4ow	-9.062	-11.062	122.573	11.062	.202	.7804	.7759
4dw	-6.222	-8.222	67.784	8.222	.179	.8043	.7961
4osd	14.371	12.371	153.264	12.371	.218	.6614	.6544
4owd	-6.222	-8.222	67.838	8.222	.234	.8043	.8003

*True Model $Y = \beta_0 + \beta_1 X + \beta_2 Z^2 + u$, $\beta_1 = 2$, $r_{XZ}2 = .553$, sample size = 100.

$$U_{np} = (\sigma^2 + \sigma^2_p)\left(1 + \frac{p}{n-1-p}\right) \quad (42)$$

with respect to the number of included variables, p, where n equals the number of observations, σ^2 equals the variance of the true error term and,

$$\sigma^2_p = \text{VAR}\left(\sum_{j=p+1}^{\infty} \beta_j X_j \mid X_1, \ldots, X_p\right) \quad (43)$$

Equation (43) depends on unobserved parameters and minimization of

$$\hat{U}_{np} = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\left(1 + \frac{p}{n-1-p}\right), \quad (44)$$

which depends only on p and the data and serves as a good approximation.

$\hat{U}_{np}$ has been calculated for the experiments in Table 10. The experiment 4owsd uses the transformation indicated by the dummy variable regression (4wd). $\hat{U}_{np}$ clearly favors the ordinal variable proxy over the dummy variable proxies and because the unconditional mean square error increases much more rapidly in p than $\bar{R}^2$ decreases in p. Clearly, since the dummy variable proxies do not convey any more information than the ordinal proxy using the dummy variable scaling, use of the dummy variable proxies in the final regression equation only introduces noise into the estimates. This effect will be noticeable in all but very large data sets.

11. Concluding Remarks

The debate over how to handle ordinal data is old and goes back at least to 1900 when Pearson and Yule expressed their opinions in the Philosophical Transactions of the Royal Society. Pearson believed that ordinal variables were imperfect measurements of continuous variables while Yule believed that treating them as possessing only nominal level information was most appropriate.[14] We have obviously chosen Pearson's view of the world.

Table 10

EXP	Mean β_1	MSE	R^2	$\bar{R}^2$	RSSE	RMSE	$\hat{U}np$
OLS (X only)	3.451	2.145	.4665	.4611	26538693	270803	276387
OLS (X, Z)	1.763	.107	.7022	.6960	14842308	153014	157796
4ow	2.610	.425	.6103	.6022	19437995	200392	206787
4dw	2.590	.398	.6217	.6057	18871765	198650	209216
4owd	2.589	.405	.6170	.6091	19105406	196963	203118

*True Model $Y = \beta_0 + \beta_1 X + \beta_2 Z + u$, $\beta_T = 2$, $r_{XZ} = .455$, sample size = 100, RSEE is residual sum of squared error, RMSE is residual mean squared error.

We feel that the use of dummy variable proxies encourages the interpretation of a slope parameter as intercept shifters and conceals the errors-in-variables nature of the problem. Since we demonstrated in the last section that the dummy variable solution can always be incorporated into an ordinal scaling scheme (whose use results in an improvement in the mean square error prediction rate for the regression over direct use of the dummy variable proxies), we can see no justification for ever using them in preference to an ordinal proxy.

There is obviously much work to be done. Barrow (1976) and Frost (1979) have shown that the choice of techniques for dealing with the errors-in-variables problem becomes more difficult when more than one variable is measured with error while Kinal and Lahiri (1983) have applied Aigner's (1974) framework to the case of stochastic regressors measured with error. Extention of this work to the special attributes of ordinal independent variables would undoubtedly be fruitful as these are the conditions under which most applied work is done. Another important area is the case where the unobserved variable, Z, for which the ordinal proxy, Z*, is available, is the variable of prime interest. Hypothesis testing in the errors-in-variables framework is still woefully inadequate. Finally, we should note that the Monte Carlo results on ordinal proxies presented here is only a beginning of what needs to be done.

While wishing for better (i.e., interval level) data is a pipe dream in many cases, as survey researchers have long known, survey questions could often be designed to provide "better" ordinal variables if guidelines were known. We have shown that simply increasing the number of response categories is not necessarily the answer. Researchers will have to be clearer about their models and in particular about the functional forms they should take if guidelines are to develop with an eye toward better multivariate work in a regression framework.

Researchers using ordinal proxies are not totally without guidance from their data and outside information. The correlation between X and Z was shown to be a key parameter in assessing the severity of the errors-in-variables problem. The correlation coefficient between X and Z* is easily calculated. Since the correlation coefficient between X and Z* is as a rule smaller than that between X and Z, a high r_{XZ*} is a sure sign of problems particularly if Z* is not a good representation of Z. The quality of the ordinal proxy Z* can often be determined by reference to outside information and simple examination of the frequency of each category of Z*. Estimation of the regression equation using different ordinal scalings and the dummy variable transformation often reveals much about the sensitivity of the β_1 parameters to the form of the proxy and of possible nonlinearities with respect to the unobserved variable Z.

This "auxillary" analysis should be reported to the reader who should also be cautioned against too literal of an interpretation of parameter estimates and t-statistics. It is all to easy to generate cross distortions, particularly when the variable of interest is only marginally significant in the true model and highly correlated with the unobserved variable for which the ordinal proxy is being used.

FOOTNOTES

*The author wishes to thank Michael Hazilla, P.A.V.B. Swamy, Raymond Kopp, Robert Mitchell, William Vaughan, and Charles Paulsen for their helpful comments. Any remaining errors are the responsibility of the author.

[1]The following example will help clarify what is meant by ordinal and dummy variable representations. Assign a person the value of 1 if his/her income is less than $10,000, 2 if income is between $10,000 and $25,000, and 3 if over $25,000. One ordinal representation is the 1, 2, 3. Other ordinal representations for the same data include any order preserving transformation of 1, 2, 3 (e.g., 12, 81, 302). The dummy variable representation is $D_1 = 1$ if income is less than $10,000, else $D_1 = 0$; $D_2 = 1$ if income is between $10,000 and $25,000, else $D_2 = 0$; $D_3 = 1$ if income is greater than $25,000, else $D_3 = 0$. Note that while there are an infinite number of ordinal representation of the same ranked categorical data there is only one dummy variable representation (the only choice being which of the dummies to drop). For more on nominal, ordinal, interval, and ratio scales see Roberts and Schulze (1973), Krantz et al. (1971), or any measurement theory text.

[2]It is possible to think of examples where the "true" underlying variable has an ordinal measurement scale rather than interval (e.g., union seniority: rank rather than years). In these cases an errors-in-variables situation is created if then interval level variable is used in the estimated model.

[3]For more information on errors-in-variables estimators and how to use outside information in particular see Judge et al. (1980) and Fuller (1980).

[4]It should be noted that some of the distributional properties of u change with the taking of deviations from the means. These changes do not affect any of the results and the u notation is maintained for convenience.

[5]All of the results to follow should be considered to be asymptotic. Aigner (1974) making a number of normality assumptions derives exact distributions. A function of the R^2 measure is asymptotically proportional to the F statistic. See Dhrymes (1978) for use of the R^2 measure in the errors-in-variables case.

[6]See Aigner (1974), Aigner and Goldberger (1977), Barrow (1976), Maddala (1977), Dhrymes (1978), and Frost (1979) for a discussion on different qualifications for using proxy variables.

[7] The larger the correlation between X and Z, the smaller the ratio of the variance of the measurement error to the variance of Z, and the larger the number of observations, the more likely use of a proxy will result in a smaller $MSE(\hat{\beta}_1)$. Aigner (1974) gives these trade offs in detail. Survey data is usually blessed with such large n that it will always be desirable to use a proxy if available.

[8] A general representation of an exponential transformation between Z* and Z can be represented by $Z^* = Ae^{Z}e^{\gamma}$ where A is a constant and γ is the error component. Z* in terms of Z is Log Z* - Log A - Log γ. Eq. (1) becomes $Y = \beta_0 + \beta_1 X + \beta_2(\text{Log } Z^* - \text{Log } A - \text{Log } \gamma) + u$ which can be rewritten as $Y = \beta_0 - \beta_2 \text{Log } A + \beta_1 X + \beta_2 \text{Log } Z^* - \beta_2 \text{Log } \gamma + u$. OLS minimizes $\sum_{j=1}^{n}(-\beta_2 \text{Log } \gamma + u)$, and the coefficient on Z* will be M where $\beta_2 \text{Log } Z^* = MZ^*$ and $M = (\beta_2 \text{Log } Z^*)/Z^*$.

[9] If α_1 is known Z can be scaled so α_1 is one. Further, since both α_1 and e will vary directly with the scaling of Z*, the parameter α_1 can not be manipulated for any gain by the researcher.

[10] Except when $r_{XZ} = 0$ which is equivalent to identification of the equation through a restriction on the covariance between X and Z.

[11] To simulate a bad metric we constructed Z* to provide little information about the much of Z's range and much information about the upper 30% of Z's range. Labovitz (1970) and O'Brien (1979) suggest that this dichotomizing transformation of an interval variable will produce the most distortion in $\hat{\beta}_1$.

[12] See Smith (1973) for details of construction.

[13] I am indebted to Leo Breiman for this observation.

[14] We have not considered the line of development which descends from Yule's contingency table approach. These methods of dealing with ordinal variables in a contingency table/log-linear framework have recently been surveyed by Agresti (1983). They are difficult to use in combination with continuous independent variables or large numbers of variables, and interpretation of the results is often non-intuitive. The errors-in-variables/missing information problem with the data is often unrecognized. A more promising line of attack using the ANOVA approach is Chamberlain (1980) who uses the concept of latent variables.

REFERENCES

Agresti, A. (1983), "A Survey of Strategies for Modeling Cross-Classifications Having Ordinal Variables," Journal of the American Statistical Association, 78, 184-198.

Aigner, D. J. (1974), "MSE Dominance of Least Squares with Errors of Observation," Journal of Econometrics, 2, 365-372.

Aigner, D. J. and A. S. Goldberger, eds. (1977), Latent Structures in Socioeconomic Models, Amsterdam: North-Holland.

Barrow, B. S. (1976), "The Use of Proxy Variables When One or Two Independent Variables Are Measured with Error, American Statistician,"30, 119-121.

Bonacich, P. and D. Kirby (1976), "Establishing a Metric from Assumption About Linearity," Journal of the American Statistical Association,"71, 56-61.

Bonrnstedt, G. W. and T. M. Carter (1971), "Robustness in Regression Analysis, 118-146, in Sociological Methodology, H. L. Costner, San Francisco: Jossey Bass.

Brieman, L. and J. Friedman (1982), "Estimating Optimal Transformations for Multiple Regression and Correlation," Technical Report No. 9, Department of Statistics, University of California.

________, and D. Freedman (1983), "How many Variables Should Be Entered in a Regression Equation," Journal of the American Statistical Association, 78, 131-136.

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," Review of Economic Studies, 47, 225-238.

de Leeuw, J., F. Young, and Y. Takane (1976), "Additive Structure in Qualitative Data: An Alternating Least Squares Method with Optimal Scaling Features," Psychometrika, 471-503.

Dhrymes, P. J. (1978), Introductory Econometrics, New York: Springer-Verlag.

Frost, P. A. (1979), "Proxy Variables and Specification Bias," Review of Economics and Statistics, 61, 323-325.

Fuller, W. A. (1980), "Properties of Some Estimators for the Errors-in Variables Model," The Annals of Statistics, 8, 407-422.

Judge, G. G., W. E. Griffiths, R. C. Hill and T. Lee (1980), The Theory and Practice of Econometrics, New York: John Wiley.

Kim, J. (1975), "Multivariate Analysis of Ordinal Variables," American Journal of Sociology, 81, 261-298.

Kim, J. (1978), "Multivariate Analysis of Ordinal Variables Revisited," American Journal of Sociology, 84, 448-456.

Kinal, T. and K. Lahiri (1983), "Specification Error with Stocastic Regressors," Econometrica, 51, 1209-1219.

Krantz, D. H., R. D. Luce, P. Suppes and A. Tversky (1971), Foundations of Measurement, vol. I, New York: Academic Press.

Labovitz, S. (1970), "The Assignment of Numbers to Rank Order Categories," American Sociological Review, 35, 515-524.

Lewis-Beck, M. S., 1980, Applied Regression: An Introduction, Beverly Hills, Ca.: Sage.

Maddala, G. S. (1977), Econometrics, New York: McGraw-Hill.

McCallum, B. T. (1972), "Relative Asymptotic Bias from Errors of Omission and Measurement," Econometrica, 40, 757-758.

O'Brien, R. M. (1979), "The Use of Pearson's r with Ordinal Data," American Sociological Review, 44, 851-857.

Pearson, K. (1900), "Mathematical Contributions to the Theory of Evolution in the Inheritance of Characteristics Not Capable of Exact Quantitative Measurement," Philosophical Transactions of the Royal Society, (Series A) 195, 79-150.

Roberts, B. and D. L. Schulze (1973), Modern Mathematics and Economic Analysis, New York: Norton.

Smith, V. K. (1973), Monte Carlo Methods: Their Role for Econometrics, Lexington, Massachusetts: Lexington Books.

Wickens, M. R. (1972), "A Note on the Use of Proxy Variables," Econometrica, 40, 759-761.

Wilson, T. P. (1971), "Critique of Ordinal Variables," Social Forces, 49, 432-444.

Young, F. J. de Leeuw, and Y. Takane (1976), "Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features," Psychometrika, 505-529.

Yule, G. U. (1900), "On the Association of Attributes in Statistics," Philosophical Transactions of the Royal Society, (Series A) 194, 257-319.