British Journal of Mathematical and Statistical Psychology (2002), 55, 305–315

www.bps.org.uk

The comparative sensitivity of ordinal multiple regression and least squares regression to departures from interval scaling

Todd S. Woodward*, Michael A. Hunter and Helena Kadlec
Department of Psychology, University of Victoria, Victoria, British Columbia, Canada V8W 3P5

The comparative sensitivity of ordinal multiple regression (OMR) and least squares regression (LSR) to criterion variable deviations from interval scaling was investigated by way of computer simulation. LSR on raw scores and ranks was compared to OMR on raw scores, ranks and dominances. Simulated data sets varied on predictor variable correlations, amount of prediction error, weight distinctiveness and shape of rating-scale distribution. The results indicated that LSR on raw scores was most affected by discretization in all conditions. In contrast, the performance of LSR approximated that of OMR when the data were first transformed to ranks. The poor performance of LSR on raw scores was most pronounced when the data discretization resulted in a symmetrical distribution. Predictor variable correlations and amount of prediction error did not affect the pattern of results. Weight distinctiveness did not interact with the other factors.

1. Introduction

Cliff and colleagues have argued that because data in psychological research are frequently measured on an ordinal scale, ordinal test theory and statistical techniques designed for use with such data would seem most appropriate (Cliff, 1989, 1991, 1993, 1996a, 1996b; Cliff & Donoghue, 1992). Accordingly, these authors have developed an alternative regression technique deemed appropriate when the criterion variable is measured on an ordinal scale, and when the predictor variables are ordinal or interval. One study has compared the confidence intervals associated with ordinal multiple regression (OMR) to that derived from least squares regression (LSR) in a data simulation study (Long, 1999). It was concluded that the performance of OMR was superior to that of LSR when data conditions were not optimal, including when the data are

*Requests for reprints should be addressed to Todd S. Woodward, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B.C., Canada, V6T 1Z4 (e-mail: twoodward@ cortex.psych.ubc.ca).

non-normally distributed. Furthermore, this author stated that the superiority of OMR was due to its use of only ordinal information, and the resulting lack of sensitivity to monotonic discretizations of the data. Implied in this conclusion is that OMR should be less sensitive than LSR to the loss of true-score information associated with discretization, such as when a Likert-type rating scale is used as a measuring instrument.

The motivation for the current investigation was to compare OMR with LSR on their sensitivity to measurement-induced discretization which resulted in varying degrees of non-normality. Instead of comparing power-related statistics (as did Long, 1999), we tested the effect of these manipulations on straightforward interpretation of weights and fit statistics. Furthermore, we examined the properties of LSR when data were ranked, since this methodology, although based on the least squares algorithm, should be equally unaffected by monotonic discretization, and should approximate the performance of OMR. Stated simply, we considered the comparative sensitivity of OMR and LSR to transformation from interval to ordinal scaling. We tested several different distributional shapes under varying predictor variable correlations, amount of prediction error and weight distinctiveness. The simulation methodology, measures of sensitivity, and the mathematical derivation of OMR are presented below.

2. Measurement–induced interval to ordinal discretizations

For this investigation it is assumed that rating scales attempt to measure true scores of traits or attitudes, and that these true scores are normally distributed, an assumption also made by Jöreskog and Sörbom (1988, p. 1–4):

> For each ordinal variable x, it is assumed that there is a latent continuous variable ξ that is normally distributed . . .

When a trait is quantified by a measuring instrument taking the form of a rating scale, the coarseness of this instrument is assumed to cause the continuous, interval-level true scores to be transformed to discrete, ordinal-level rating-scale scores with measurement error, which may or may not be normally distributed. Defining ordinal variables such as rating scales, Jöreskog and Sörbom (1988, p. 1–4) wrote:

> Assuming that there are k categories on x, we write $x = i$ to mean that x belongs to category i . . . The connection between x and ξ is that $x = i$ is equivalent to $\alpha_{i-1} < \xi \leq \alpha_i$, where $\alpha_0 = -\infty$, $\alpha_1 < \alpha_2 < \ldots < \alpha_{k-1}$, and $\alpha_k = +\infty$ are parameters called threshold values. If there are k categories, there are $k - 1$ unknown thresholds.

We adopted Jöreskog and Sörbom's assumptions and terminology to specify two general types of discretizations which may arise when a latent continuous variable is captured on a discrete rating scale: equal intervals (equally spaced thresholds) $\alpha_i - \alpha_{i-1} = \alpha_j - \alpha_{j-1}$, for all $i, j = 1, 2, \ldots, k - 1$, where $i \neq j$; and unequal thresholds $\alpha_i - \alpha_{i-1} \neq \alpha_j - \alpha_{j-1}$, for any $i, j = 1, 2, \ldots, k - 1$, where $i \neq j$. Thus, rating scale measuring instruments that map normally distributed, interval-level, continuous sets of scores onto discrete ordinal scales with equal intervals would produce an approximately normally distributed discrete variable. However, it is more likely that some type of unequal threshold discretization would take place. We simulated both situations in the present investigation.

3. Ordinal multiple regression and its relation to least squares regression

We now provide a brief introduction to the computational underpinnings of OMR, but encourage the reader to refer directly to Cliff (1994, 1996b). The first step in calculating OMR weights is the discretization of the ordinal-level criterion variable scores, y, to *dominance scores*. A dominance score d_{ih} is calculated for every *pair* of persons (i, h): $d_{ih} = 1$ if $y_i > y_h$, and $d_{ih} = -1$ if $y_i < y_h$. These dominance scores then identify a type of group membership.[1] The first OMR criterion group consists of all pairs of subjects with the dominance score of 1 (i.e., of the $n!/(n-2)!2!$ possible pairs, those whose first member scores higher than the second on the criterion variable). The second group consists of the same pairs in the reverse order, or all pairs with the dominance statistic of -1. This results in a vector of length $n(n-1)$.

The second step is to compute a matrix of raw score differences, rank differences, or dominances from the predictor variables, denoted as matrix $\mathbf{A}$, for the same pairs of persons for whom dominance scores on the criterion variable were computed. The third and final step in computing OMR weights is to solve the normal equations:

$$\mathbf{w}_{\text{omr}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is a vector of the dominance scores derived from the criterion variable, $\mathbf{A}$ contains either the raw score differences, rank differences or dominances, and $\mathbf{A}'$ is the transpose of $\mathbf{A}$.

The fact that OMR weights are computed using the normal equations (i.e., by minimizing the sum of squared errors) suggests a rather direct relationship between OMR and LSR. The nature of this relationship has previously been described by Cliff (1994, equations (17)–(22)), when he notes that if raw score differences are used for matrix $\mathbf{A}$, equation (1) can be rewritten as

$$\mathbf{w}_{\text{omr}} = \frac{2\mathbf{S}^{-1}\mathbf{s}_{ry}}{n}, \tag{2}$$

where $\mathbf{S}$ is the familiar variance–covariance matrix of the raw score predictor variables, and $\mathbf{s}_{ry}$ is the vector of covariances between the raw score predictor variables and the ranked criterion variable. The similarity of (2) to the LSR model,

$$\mathbf{w}_{\text{lsr}} = \mathbf{S}^{-1}\mathbf{s}_{xy} \tag{3}$$

is evident, the only differences being the OMR term $\mathbf{s}_{xy}$ (the vector of covariances between the predictors and the criterion variable raw scores) in place of the LSR term $\mathbf{s}_{ry}$, and a constant.

Thus, OMR and LSR weights can be considered equivalent under certain circumstances. For example, the weights resulting from the use of raw score differences for $\mathbf{A}$ in (1) can be considered equivalent (differing only by a constant) to LSR with a *ranked* criterion variable. Similarly, use of rank differences for $\mathbf{A}$ in (1) can be considered equivalent (differing only by a constant) to LSR with predictor and criterion variables *both ranked*. However, there is no LSR equivalent for the most purely ordinal OMR technique, which is based on Kendall's tau:

$$\mathbf{w}_{\text{omr}} = \mathbf{T}^{-1}\boldsymbol{\tau}_{xy}, \tag{4}$$

---

[1] *$d_{ih} = 0$ if $y_i = y_h$. Ties on the criterion variable can be accounted for by adjustments in the formulae (see Cliff, 1994, p. 138).*

where **T** contains the tau-a correlations[2] among the predictors, and τ_{xy} is the vector of the tau-a correlations between the predictors and the criterion variable. This OMR weight computation is equivalent to using dominances in **A** in (1).

These relationships between LSR and OMR can be understood as occurring along an interval–ordinal continuum. LSR using raw scores assumes interval-level criterion and predictor variables, so can be considered a purely interval technique, occupying the 'interval' end of the continuum. OMR using dominances for **A** in (1) uses only ordinal information, so can be considered a purely ordinal technique occupying the 'ordinal' end of the continuum. OMR on ranks and raw scores and LSR on ranks can be considered intermediate points along this continuum.

4. Computer simulations

Using computer simulation, continuous, normally distributed interval-level criterion variables (ξ_Y) were discretized to 5-point rating scales (**Y**). The following five rating-scale distributions were simulated: low scores dominate (asymmetric); flat distribution (symmetric); extreme scores dominate (symmetric); extreme scores and low values dominate (asymmetric); and a discretized approximation of a normal distribution (symmetric), where the percentages of scores in each of the five rating-scale categories were 9, 24, 34, 24, 9. The frequency distributions resulting from discretization are played graphically in Fig. 1.

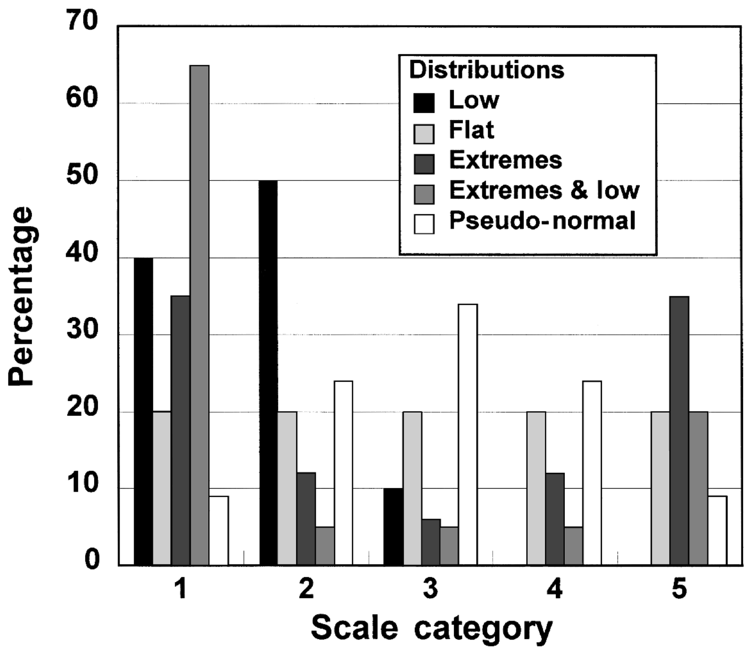

Figure 1. Sample distributions of simulated data following discretization of the criterion variable.

[2] *Tau-a is conceptually equivalent to the proportion of paired rank-order agreement on two variables (Kendall & Gibbons, 1990). Tau-a is widely referred to as Kendall's tau.*

In this preliminary study we elected to focus on the impact of discretization of the criterion variable only. This decision was based on the fact that with three predictor variables, five distributions and possible interactions with predictor-variable-based factors (i.e., intercorrelations and weight distinctiveness), thousands of combinations of predictor/criterion variable condition levels would be possible if the impact of predictor variable discretization were studied simultaneously. In the interest of delivering a clear initial result regarding the comparative sensitivity of OMR and LSR to discretization, investigation of the impact of combinations of predictor variable distributions was left for future investigations.

The criterion variable $\boldsymbol{\xi}_Y$ was generated in three steps:

1. A matrix $\boldsymbol{\Xi}$ consisting of three vectors, $[\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3]$, of normally distributed data with mean 0 and standard deviation 1 were generated to represent three 'true score' predictor variables. These vectors were created with differing levels of mutual intercorrelation (0, 0.25, 0.5 and 0.75), and the sample size was 50.
2. Three sets of weights, $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]$ differing in distinctiveness were used for data generation: an equivalent set [0.5 0.5 0.5], a close set [0.4 0.5 0.6] and a disparate set [0.3 0.5 0.7]. The constant term β_0 for all sets was 0. The matrix of the predictor variables was post-multiplied by the vector of beta weights to create the predicted true-score criterion variable:
$$\hat{\boldsymbol{\xi}}_Y = \boldsymbol{\Xi}\boldsymbol{\beta}. \tag{5}$$
3. Variance not explained by the true-score predictor variable (residual variance) was simulated by adding a normally distributed vector of random numbers, $\boldsymbol{\varepsilon}$, to the predicted true score vector $\hat{\boldsymbol{\xi}}_Y$. The percentage of prediction error was manipulated by specifying the variance of $\boldsymbol{\varepsilon}$, such that the resulting residual variance in the criterion variable took one of three possible values: 25%, 50% or 75%. This sum produced the initially untransformed criterion variables:
$$\boldsymbol{\xi}_y = \hat{\boldsymbol{\xi}}_Y + \boldsymbol{\varepsilon}. \tag{6}$$
The effect of the discretization was assessed by comparing the results of the analyses regressing $\boldsymbol{\xi}_y$ on $\boldsymbol{\Xi}$ to the same analyses using the discretized $\boldsymbol{\xi}_y$. The exact method of comparison will be detailed below.

In summary, four factors were experimentally manipulated: predictor variable intercorrelation, weight set, percentage of prediction error, and criterion variable distribution type. For each combination of factors (180 in total), 100 data sets were simulated ($N = 50$ for each data set[3]). Each data set was subjected to five analyses: ordinary least squares multiple regression on raw scores (LSR_{raw}); ordinary least squares multiple regression on ranks ($\text{LSR}_{\text{ranks}}$), where both the criterion and predictor variables are converted to ranks; OMR with raw score differences as predictors (OMR_{raw}); OMR with rank score differences as predictors ($\text{OMR}_{\text{ranks}}$); and OMR with dominances as predictors (OMR_{dom}).

The model that was fitted for all analyses was

$$\mathbf{y} = b_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3 + \mathbf{e}, \tag{7}$$

where $\mathbf{y}$ was either the normally distributed, continuous variable $\boldsymbol{\xi}_Y$ before discretization, or a discrete, pseudo- or non-normally distributed variable representing a rating

[3] *In pilot tests, it was observed that sample size had a negligible effect on the pattern of results; therefore, this variable was not manipulated in the present investigation.*

scale **Y** after one of the five discretizations. The predictor variables $\mathbf{x}_i$ were always the normally distributed, untransformed variables ($[\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3]$). Note that the estimated weights ($\hat{b}_1, \hat{b}_2, \hat{b}_3$) were expected to be smaller than the weights used for data generation ($\beta_1, \beta_2, \beta_3$), due to the influence of prediction error $\boldsymbol{\varepsilon}$ (added to $\hat{\boldsymbol{\xi}}_Y$, see (6)).

5. Performance indices

A difficulty arises when comparing OMR to LSR, in that the resulting fit statistics and weights are not directly comparable in terms of absolute values or interpretation. Despite this, we developed performance indices which were comparable across regression methods, and were directly interpretable at face value. This was achieved by using proportions instead of absolute values.

5.1. Weights

Traditionally, the rank order of regression weights has been considered a relatively minor issue (Wainer, 1976). However, regarding interpretation of OMR weights, Cliff (1994, p. 131) wrote:

> The interpretation of the relative sizes of weights is purely pragmatic. If one variable has a higher weight than another, this simply means that a difference (or dominance, or rank difference) of a given size on the first variable will predictively override a differences of the same size but opposite direction on the other.

Thus, Cliff emphasizes that the pattern of OMR regression weights must be interpreted in terms of *relative size*. This focus on relative size provides a solution to the problem of how to make the 'apples and oranges' comparison of OMR and LSR weight estimates before and after discretization. What must be preserved is the relative sizes of weights, not their absolute values. Thus, the effect of discretization was assessed by comparing the *rank order* of OMR and LSR weights before and after discretization.

More specifically, for each type of regression analysis, the estimated weights for $\boldsymbol{\xi}_Y$ (untransformed), regressed on the three predictors, were rank-ordered. The same regression analyses were then carried out on the same data, but $\boldsymbol{\xi}_Y$ was discretized to **Y**, representing a rating scale. The weights were again rank-ordered. If the rank order of the weights changed, a 1 was recorded; if the rank order did not change, a 0 was recorded to specify no effect of the discretization. This allowed direct interpretation of the weight change measure as the proportion of samples for which the pattern of *relative size* of the weights was affected by the discretizations. (Recall that both pre-and post-discretization weights were affected by measurement error, such that the set of betas $\boldsymbol{\beta}$ that began as an equivalent set [0.5 0.5 0.5] could be rank-ordered after the addition of error, which changed the relative size in a randomly determined fashion.)

5.2. Fit statistics

Also of interest was the influence of the discretization of the criterion variable on overall model fit statistics. Cliff (1994) introduced O_2 as an objective function to be maximized in ordinal multiple regression. This statistic is directly interpretable as the proportion of pairs for which the predicted dominances were equal to the observed dominances. The formula for computation of O_2 is

$$O_2 = \frac{\sum u_{ih}[\mathrm{sign}(\hat{u}_{ih})]}{n(n-1)}, \qquad (8)$$

Table 1. Proportion of samples for which weight order changed as a function of distribution and analysis type, averaged over all other factors. Regression types are ordinary least squares multiple regression on raw scores (LSR_{raw}), ordinary least squares multiple regression on ranks and ordinal multiple regression with rank score differences as predictors (LSR/OMR_{ranks}), ordinal multiple regression with raw score differences as predictors (OMR_{raw}) and ordinal multiple regression with dominances as predictors (OMR_{dom}). Means and marginal means are presented

	LSR_{raw}	LSR/OMR_{ranks}	OMR_{raw}	OMR_{dom}	Marginal means
Low	0.45	0.40	0.41	0.40	0.41
Extremes and low	0.53	0.47	0.48	0.45	0.48
Extremes	0.39	0.26	0.27	0.29	0.30
Flat	0.31	0.18	0.21	0.20	0.23
Pseudo-normal	0.30	0.26	0.26	0.26	0.27
Marginal means	0.39	0.31	0.33	0.32	

where $\hat{u}$ represents predicted dominances, and the sum is over pairs $((i, h)$, where $i \neq h)$ of scores (Cliff, 1994, p. 130).[4] Correspondingly, the fit statistic we used for LSR was the traditional R^2 statistic.

Performance was evaluated for fit statistics by computing the proportion reduction in the fit statistic resulting from discretization of $\boldsymbol{\xi}_Y$ to $\mathbf{Y}$. In other words, if the measure-of-fit index (O_2 or R^2) computed on pre-discretization data (i.e., on normally distributed, continuous, interval-scale data) was reduced by 20%when the same regression analysis was carried out on the discretized data, a 0.2 was recorded to document this change.

6. Results

All measures of performance were computed on each simulated sample ($N = 50$), and 100 simulated samples were created for each design cell ($4 \times 3 \times 3 \times 5 = 180$). Due to the previously mentioned equivalence of weights resulting from LSR_{ranks} and OMR_{ranks} (see also Cliff, 1994, equation (22)), there are five levels of the type of regression analysis when investigating measures of fit, but only four when investigating the relative importance of weights. The results are presented graphically (significance tests were carried out to guide our selection of reported effects, only those significant at the $p < 0.001$ level are presented).

6.1. Weights

Analysis of the weight performance measures suggested that the only effects warranting investigation were type of regression and similarity of weight sets. Figure 2 provides a visual display of the proportion of samples affected by each discretization type and analysis type, averaged over the remaining between-sample factors (there were no interactions between the factors). Table 1 provides this information in numeric form, with marginal means. The most interesting result was that of analysis type, whereby LSR_{raw} weights (proportion changed = 0.39, averaged over discretization type) changed rank order more frequently than the weights resulting from other analysis types

[4] *It should be noted that (4) maximizes (8) only approximately, and in some situations the single best predictor may maximize (8) (Cliff, 1994).*

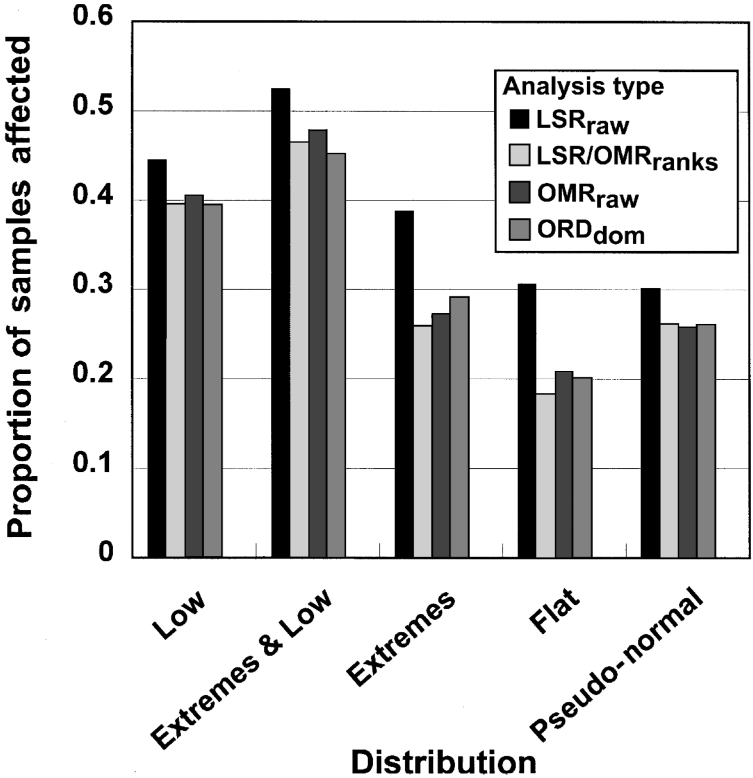

Figure 2. Proportion of samples for which weight order changed as a function of distribution and analysis type, averaged over all other factors. Charted regression types are ordinary least squares multiple regression on raw scores (LSR_{raw}), ordinary least squares multiple regression on ranks and ordinal multiple regression with rank score differences as predictors (LSR/OMR_{ranks}), ordinal multiple regression with raw score differences as predictors (OMR_{raw}) and ordinal multiple regression with dominances as predictors (OMR_{dom}).

(proportion changed = 0.32, averaged over OMR_{raw}, LSR/OMR_{ranks}, OMR_{dom}, and discretization type). In other words, LSR performed as well as OMR when the data were transformed to ranks prior to analysis, but not when raw scores were used.

An effect of weight set similarity was present but uninteresting, because this factor did not interact with other manipulations. As expected, the proportion of samples affected by this factor increased with weight similarity in the following order: disparate (proportion changed = 0.28), close (proportion changed = 0.34), and pre-error equivalent (proportion changed = 0.39) weights.

6.2. Measures of fit

Figure 3 provides a visual display of the proportion of reduction in the fit index (averaged over samples) induced by the discretization, as a function of discretization type and analysis type, averaged over the remaining between-samples factors. Table 2 provides this information in numeric form, with marginal means. With the symmetric 'Flat' and 'Extremes' discretizations, LSR_{raw} (means = 0.15, 0.27) was more affected

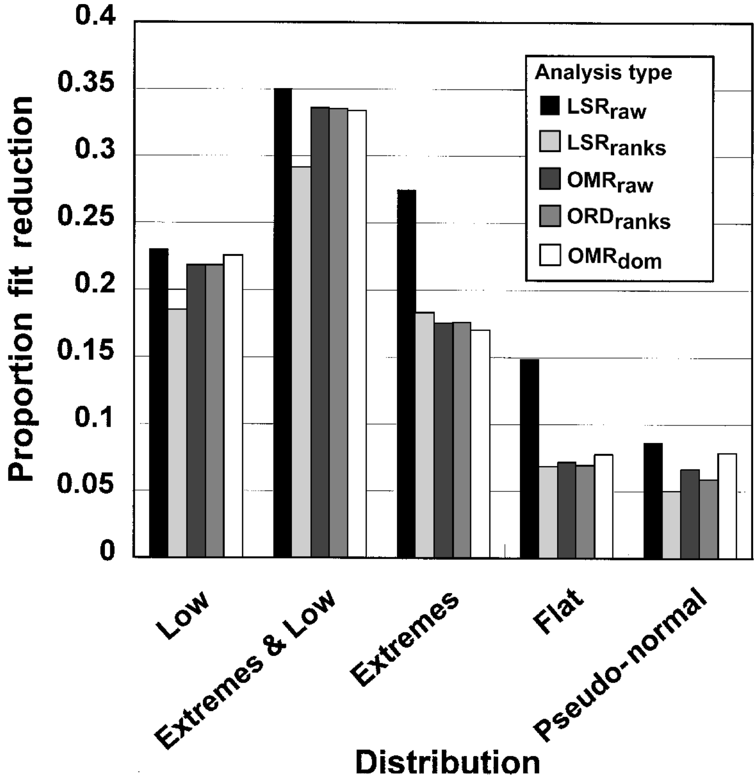

Figure 3. Mean proportion of reduction in fit as a function of distribution and analysis type, averaged over all other factors. Charted regression types are ordinary least squares multiple regression on raw scores (LSR_{raw}), ordinary least squares multiple regression on ranks (LSR_{ranks}), ordinal multiple regression with raw score differences as predictors (OMR_{raw}), ordinal multiple regression with rank score differences as predictors (OMR_{ranks}) and ordinal multiple regression with dominances as predictors (OMR_{dom}).

Table 2. Mean proportion of reduction in fit as a function of distribution and analysis type, averaged over all other factors. Regression types are ordinary least squares multiple regression on raw scores (LSR_{raw}), ordinary least squares multiple regression on ranks (LSR_{ranks}), ordinal multiple regression with raw score differences as predictors (OMR_{raw}), ordinal multiple regression with rank score differences as predictors (OMR_{ranks}), and ordinal multiple regression with dominances as predictors (OMR_{dom}). Means and marginal means are presented

	LSR_{raw}	LSR_{ranks}	OMR_{raw}	OMR_{ranks}	OMR_{dom}	Marginal means
Low	0.23	0.19	0.22	0.22	0.23	0.22
Extremes & low	0.35	0.29	0.34	0.34	0.33	0.33
Extremes	0.27	0.18	0.18	0.18	0.17	0.20
Flat	0.15	0.07	0.07	0.07	0.08	0.09
Pseudo-normal	0.09	0.05	0.07	0.06	0.08	0.07
Marginal means	0.22	0.16	0.17	0.17	0.18	

than the other analysis types (averaged means = 0.07, 0.18, respectively). With the asymmetric 'Low' and 'Extremes & Low' discretizations, LSR_{ranks} (mean = 0.19, 0.29, respectively) was less affected than the other analysis types (mean = 0.23, 0.34, respectively). Thus, on the measure of fit reduction, as was the case for weights analysis, LSR_{raw} performed most poorly.

7. Discussion

This investigation compared the sensitivity of LSR techniques to OMR when regression models were fitted to a criterion variable affected by discretization into various distributions. The most general conclusion was that LSR on raw scores was most affected by discretizations in all conditions. However, the performance of LSR approximated that of OMR when the data were first transformed to ranks. The symmetry of the resulting score distribution had an effect on weight importance, such that regression-method differences in weight importance were less pronounced for asymmetric discretization types. Finally, the symmetry of the resulting score distribution had an effect on fit statistics, such that LSR_{raw} performed poorly relative to the other regression techniques for symmetric discretizations, and LSR_{ranks} performed well relative to the other regression techniques for asymmetric discretizations.

It is important to note that, for all analysis types, in all conditions, the relative size of the weights and fit statistics was affected by discretizations to a rating scale. Regarding the relative size of weights, averaged over prediction error, weight set similarity and predictor variable intercorrelation, the least affected was LSR/OMR_{ranks} for the 'Flat' discretization type (see Table 1 and Fig. 2), for which the order of weight importance changed in 18% of the samples. The most severe effect was obtained for the skewed rating-scale distributions. For example, performing LSR_{raw} with the 'Extremes & Low' discretization type resulted in 53% of the samples being affected. Discretization caused a reduction in fit under all conditions, although this reduction was relatively small for the 'Flat' and 'Normal' discretization types (9% and 7%, averaged over all other manipulations, respectively; see Table 2 and Fig. 3). This reduction was much more severe for other discretization type/analysis type combinations. For example, LSR_{raw} on the 'Extremes & Low' discretization produced a 35% decrease in fit.

In summary, although all regression techniques were affected by the interval-to-ordinal discretization of the criterion variable, LSR_{raw} was more affected than OMR in all conditions. However, if data were transformed to ranks, LSR approximated OMR, and differences in performance were negligible. These results suggest that measurement-induced ordinality will affect LSR appreciably more than it will affect OMR techniques, unless data are transformed to ranks prior to analysis, in which case LSR can approximate its OMR counterpart. As a final point, it should be noted that this comparison of LSR to OMR did not employ data for which both predictor and criterion variables were ordinal; this more complete test, including the effect of combinations of non-normal distributions on predictor and criterion variables, must await further evaluation.

References

Cliff, N. (1989). Ordinal consistency and ordinal true scores. *Psychometrika*, *54*, 75–91.

Cliff, N. (1991). Ordinal methods in the assessment of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change*. Washington, DC: American Psychological Association.

Cliff, N. (1993). Dominance statistics: Ordinal answers to ordinal questions. *Psychological Bulletin*, *114*, 494–509.

Cliff, N. (1994). Predicting ordinal relations. *British Journal of Mathematical and Statistical Psychology*, *47*, 127–150.

Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research*, *31*, 331–350.

Cliff, N. (1996b). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Erlbaum.

Cliff, N., & Donoghue, J. R. (1992). Ordinal test fidelity estimated by an item sampling model. *Psychometrika*, *57*, 217–236.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS user's guide*. Uppsala: Scientific Software.

Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods*, 5th ed. New York: Oxford University Press.

Long, J. (1999). A confidence interval for ordinal multiple regression weights. *Psychological Methods*, *4*, 315–330.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*, 213–217.

Received 11 January 2000; revised version received 31 August 2001