Miscellanea

A new residual for ordinal outcomes

By CHUN LI and BRYAN E. SHEPHERD

Department of Biostatistics, Vanderbilt University, Nashville, Tennessee 37232, U.S.A.

chun.li@vanderbilt.edu   bryan.shepherd@vanderbilt.edu

Summary

We propose a new residual for regression models of ordinal outcomes, defined as $E\{\text{sign}(y, Y)\}$, where y is the observed outcome and Y is a random variable from the fitted distribution. This new residual is a single value per subject irrespective of the number of categories of the ordinal outcome, contains directional information between the observed value and the fitted distribution, and does not require the assignment of arbitrary numbers to categories. We study its properties, describe its connections with other residuals, ranks and ridits, and demonstrate its use in model diagnostics.

Some key words: Model diagnostics; Ordinal outcome; Ordinal regression; Residual.

1. Introduction

Residuals are an important component of regression analysis. The most basic residual, from linear regression, has many nice features. Specifically, the residual

(a) results in only one value per subject;
(b) reflects the overall direction of the observed value compared with the fitted value;
(c) is monotonic with respect to the observed value for those with the same covariates;
(d) has a range of possible values that is symmetric about zero; and
(e) has expectation zero.

This list is by no means comprehensive, and some of these features are more important than others. But these combined features make the linear regression residual popular and useful for diagnostics and tests of conditional independence.

Residuals exist for ordinal outcomes, notably Pearson, cumulative Pearson and deviance residuals (McCullagh & Nelder, 1989). Although these residuals have some use for checking model assumptions and fit, they do not satisfy (a)–(e). Liu et al. (2009) proposed some new residuals; their residual based on the sum of cumulative residuals satisfies (a)–(e), but it implicitly assigns equal-distance scores to the categories; see the Supplementary Material. For ordinal outcomes, in addition to (a)–(e), the ideal residual

(f) should preserve order without assigning arbitrary scores to the categories.

A residual satisfying (a)–(f) could permit the application of diagnostic tools developed for linear regression to ordinal models, and could provide a framework for testing conditional independence with ordinal variables. Here we study a new residual for ordinal data. It was introduced as a tool for constructing tests of association between ordinal variables (Li & Shepherd, 2010). However, its properties and other uses have not been studied.

2. New residual for ordinal outcomes

2·1. *Definition*

Consider a set of s ordered categories, $S = \{1, \ldots, s\}$, with order $1 < \cdots < s$. For a category y in S and a distribution F over S, we define a residual

$$r(y, F) = E\{\text{sign}(y, Y)\} = \text{pr}(y > Y) - \text{pr}(y < Y),$$

where Y is a random variable with distribution F, and $\text{sign}(a, b)$ is $-1, 0$ and 1 for $a < b, a = b$ and $a > b$, respectively.

In ordinal regression models, we fit an ordinal outcome variable Y on covariates Z. Some common models include continuation ratio, proportional odds and other cumulative link models (Agresti, 2002). For subject i, let Y_i be the outcome and let $F_{Z_i;\theta}$ be the distribution of Y_i given covariates Z_i under a model with parameters θ. We define

$$R_i = r(Y_i, F_{Z_i;\theta}). \tag{1}$$

Given data (y_i, z_i) and a fitted model with parameter estimates $\hat{\theta}$, the residual for subject i is $\hat{r}_i = r(y_i, F_{z_i;\hat{\theta}})$. Notice that $\hat{r}_i$ is not a realization of R_i, but of the random variable $\hat{R}_i = r(Y_i, F_{Z_i;\hat{\theta}})$. If $\hat{\theta} \to \theta$ in probability, then $F_{Z_i;\hat{\theta}} \to F_{Z_i;\theta}$ and $\hat{R}_i \to R_i$ in distribution. Therefore, moment properties of R_i are applicable to $\hat{R}_i$ asymptotically.

Our residual has all the desirable features listed in § 1; (a), (b) and (f) are obvious and (c)–(e) will result from Properties 1 and 8 below. Proofs of all the properties are given in the Appendix.

2·2. *Properties of $r(y, F)$, R_i and $\hat{r}_i$*

First consider a distribution $F = (p_1, \ldots, p_s)$ over S. The corresponding cumulative probabilities are $\gamma_j = \sum_{k \leqslant j} p_k$, with $\gamma_s = 1$. For convenience, we define $\gamma_0 = 0$. Then for category $j \in S, r(j, F) = \gamma_{j-1} - (1 - \gamma_j) = \gamma_{j-1} + \gamma_j - 1$. The following properties hold:

Property 1. $-1 \leqslant r(1, F) \leqslant \cdots \leqslant r(s, F) \leqslant 1$;

Property 2. when $s = 1, r(1, F) = r(1, 1) = 0$, where $F = (1)$ is a point mass;

Property 3. $r(j, F) = -r(s - j + 1, G)$, where $G = (q_1, \ldots, q_s) = (p_s, \ldots, p_1)$ with $q_j = p_{s-j+1}$;

Property 4. as functions of $(\gamma_1, \ldots, \gamma_{s-1})$, $\partial r(j, F)/\partial \gamma_j = \partial r(j + 1, F)/\partial \gamma_j$.

Property 1 implies (c) and (d). Property 3 means that when the order of the categories is reversed the residual has the same magnitude but the opposite sign. Property 4 implies that when all probabilities are fixed except for p_j and p_{j+1}, the residuals for categories j and $j + 1$ change at the same rate as γ_j, or equivalently, p_j, changes.

If two adjacent categories t and $t + 1$ are merged, with distribution $G = (q_1, \ldots, q_{s-1})$, where $q_j = p_j$ for $j < t, q_t = p_t + p_{t+1}$ and $q_j = p_{j+1}$ for $j > t$, then the branching property (Brockett & Levine, 1977),

Property 5. $r(j, G) = \begin{cases} r(j, F) & (j < t), \\ r(j + 1, F) & (j > t), \\ \{p_t r(t, F) + p_{t+1} r(t + 1, F)\}/(p_t + p_{t+1}) & (j = t), \end{cases}$

ensures robustness to the number of categories, that is, the residual for the new category is a weighted average of those of the merged categories with weights proportional to their probabilities, and the residual remains the same for the other categories.

Properties 1–5 are sensible for a residual measure for ordinal outcomes. Our residual is the only measure, up to a constant factor, that satisfies Properties 1–5. In fact, Properties 2, 5, and simpler versions of Properties 3 and 4, are sufficient for deriving our residual. Specifically, when $s = 2$, consider:

Property 3'. $r\{1, (p, 1 - p)\} = -r\{2, (1 - p, p)\}$;

Property 4'. $dr\{1, (p, 1 - p)\}/dp = dr\{2, (p, 1 - p)\}/dp$.

Then the following uniqueness property holds.

Property 6. The function $r(j, F)$ satisfies Properties 2, 3', 4' and 5 if and only if $r(j, F) = c\{\gamma_{j-1} - (1 - \gamma_j)\}$, where c is an arbitrary constant.

A similar uniqueness property was proved by Brockett & Levine (1977) while studying the properties of ridits (Bross, 1958; Agresti, 1984). In fact, our residual is closely related to ridits, which have been used for scoring levels of an ordinal variable. The ridit for level j is $\text{ridit}_j = \gamma_{j-1} + p_j/2 = (\gamma_{j-1} + \gamma_j)/2$, and the mean ridit is $1/2$. The following property holds:

Property 7. $r(j, F)/2 = \text{ridit}_j - 1/2$.

Now consider a random variable Y over S, with distribution $F = (p_1, \ldots, p_s)$. Then $R = r(Y, F)$ is a random variable, for which the following hold:

Property 8. $E(R) = 0$;

Property 9. $\text{var}(R) = (1 - \sum_{j=1}^{s} p_j^3)/3$, or alternatively, $\text{var}(R) = \sum_{j=1}^{s} p_j \gamma_{j-1} \gamma_j$;

Property 10. when $p_1 = \cdots = p_s = 1/s$ $\text{var}(R)$ reaches its maximum $(1 - 1/s^2)/3$.

Property 9 provides alternative ways of calculating $\text{var}(R)$ and implies that $\text{var}(R)$ does not depend on the order of the probabilities. The maximum of $\text{var}(R)$ is an increasing function of s and it approaches the cap $1/3$ fairly quickly, being 0·25, 0·30, 0·32, 0·33, for $s = 2, 3, 5, 10$, respectively.

Now consider a random sample of n subjects, with n_j subjects in category j $(j = 1, \ldots, s)$. Their empirical distribution is $\hat{F} = (\hat{p}_1, \ldots, \hat{p}_s)$, where $\hat{p}_j = n_j/n$, with cumulative probabilities $\hat{\gamma}_j = \sum_{k \leqslant j} n_k/n$. Since there are no covariates, we can think of a constant predictor for all subjects. Then $\hat{F}$ is the fitted distribution and for subjects in category j, their residual is $r_j = r(j, \hat{F}) = \hat{\gamma}_{j-1} + \hat{\gamma}_j - 1 = (\sum_{k<j} 2n_k + n_j - n)/n$. If we rank these subjects their midrank is $\text{rank}_j = \sum_{k<j} n_k + (n_j + 1)/2$ and,

Property 11. $\text{rank}_j = T(r_j)$, where $T(r) = (n/2)r + (n + 1)/2$.

The function $T(r)$ can be viewed as a translation from the residual scale to the rank scale. Li & Shepherd (2010) presented statistics for testing the association between two ordinal variables, X and Y, while adjusting for covariates Z. One statistic was the correlation coefficient between the residuals from models for $X \mid Z$ and $Y \mid Z$. Property 11 implies that when there are no covariates, our statistic is Spearman's rank correlation coefficient between X and Y. When covariates exist, $T(r)$ will yield adjusted ranks of the subjects, and our statistic can be interpreted as an adjusted rank correlation.

We now focus on models for an ordinal outcome Y on covariates Z with parameters θ. For subject i $(i = 1, \ldots, n)$ and category j $(j = 1, \ldots, s)$, let $\gamma_{i,j} = \text{pr}(Y_i \leqslant j \mid Z_i; \theta)$; for convenience, we define $\gamma_{i,0} = 0$. Let $p_{i,j} = \text{pr}(Y_i = j \mid Z_i) = \gamma_{i,j} - \gamma_{i,j-1}$. The following moment properties hold:

Property 12. $E(R_i \mid Z_i; \theta) = 0$;

Property 13. $\text{var}(R_i \mid Z_i; \theta) = E(R_i^2 \mid Z_i; \theta) = \sum_{j=1}^{s} p_{i,j} (\gamma_{i,j-1} + \gamma_{i,j} - 1)^2$.

Let $\hat{p}_{i,j}$ and $\hat{\gamma}_{i,j}$ be the maximum likelihood estimates of $p_{i,j}$ and $\gamma_{i,j}$. Then $\hat{r}_i = \hat{\gamma}_{i,y_i-1} + \hat{\gamma}_{i,y_i} - 1$. The variance of R_i can be consistently estimated by inserting these estimates into Property 13, $\hat{\text{var}}(R_i) = \text{var}(R_i \mid z_i; \hat{\theta}) = \sum_{j=1}^{s} \hat{p}_{i,j} (\hat{\gamma}_{i,j-1} + \hat{\gamma}_{i,j} - 1)^2$. One could therefore calculate a standardized residual as $\hat{r}_i/\{\hat{\text{var}}(R_i)\}^{1/2}$, which for binary Y is the Pearson residual $(I_{\{Y_i=2\}} - \hat{p}_{i,2})(\hat{p}_{i,1}\hat{p}_{i,2})^{-1/2}$ (McCullagh & Nelder, 1989).

Now consider a proportional odds model (McCullagh, 1980), $\text{logit}\{\text{pr}(Y \leqslant j \mid Z)\} = \alpha_j + Z^{\mathrm{T}}\beta$ $(j = 1, \ldots, s - 1)$, with parameters $\theta = (\alpha_1, \ldots, \alpha_{s-1}, \beta)$. Under this model, our residuals are related to score residuals (Therneau et al., 1990). Let l_i be the loglikelihood for subject i, and $U_i = U_i(\theta) = \partial l_i/\partial\theta$. Because $E_\theta(U_i) = 0$, U_i is called the score residual. For proportional odds models, the following properties hold:

Property 14. $\hat{r}_i = -\sum_{j=1}^{s-1}(\partial l_i/\partial\alpha_j)|_{\hat{\theta}}$, where $\hat{\theta}$ is the maximum likelihood estimate;

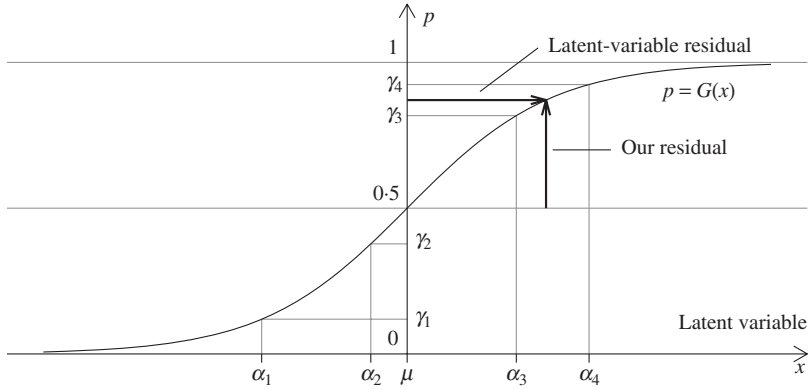*Property* 15. $\sum_{i=1}^{n} \hat{r}_i = 0$.

Fig. 1. Connection between our residual and a latent-variable residual.

Property 14 implies that our residual is a partially aggregated score residual over the α components of $U_i(\hat{\theta})$, and Property 15 is analogous to that of linear regression residuals.

2·3. *Connection to residuals on a latent variable scale*

In cumulative link regression models, the cumulative probability $\gamma_{i,j}$ is modelled through a link function $G^{-1}(\gamma_{i,j}) = \alpha_j + z_i^{\mathrm{T}}\beta$, where G is a cumulative distribution function over the real line. The ordinal outcome Y_i can be viewed as the result of applying thresholds $\alpha_1, \ldots, \alpha_{s-1}$ to a latent random variable U_i that has cumulative distribution function $G_i(u) = G(u + z_i^{\mathrm{T}}\beta)$. Then $E(U_i) = \mu - z_i^{\mathrm{T}}\beta$, where μ is the mean of G. For convenience, let $\alpha_0 = -\infty$ and $\alpha_s = +\infty$.

If we observed U_i, the usual residual on the latent variable scale would be $U_{i,\mathrm{res}} = U_i - E(U_i) = U_i + z_i^{\mathrm{T}}\beta - \mu$. Since $U_i + z_i^{\mathrm{T}}\beta \sim G$, $U_{i,\mathrm{res}}$ follow the same distribution for all subjects and $E(U_{i,\mathrm{res}}) = 0$. As U_i is latent, we do not know its value but only that $\alpha_{j-1} < U_i < \alpha_j$ if $Y_i = j$. One may want to replace U_i with

$$E(U_i \mid Y_i = j) = \frac{1}{p_{i,j}} \int_{\alpha_{j-1}}^{\alpha_j} u \, \mathrm{d}G_i(u) = \frac{1}{p_{i,j}} \int_{\gamma_{i,j-1}}^{\gamma_{i,j}} G_i^{-1}(p) \, \mathrm{d}p = \frac{1}{p_{i,j}} \int_{\gamma_{i,j-1}}^{\gamma_{i,j}} G^{-1}(p) \, \mathrm{d}p - z_i^{\mathrm{T}}\beta,$$

and define a residual on the latent variable scale as

$$L_{i,\mathrm{res}} = E(U_i \mid Y_i = j) + z_i^{\mathrm{T}}\beta - \mu = \frac{1}{p_{i,j}} \int_{\gamma_{i,j-1}}^{\gamma_{i,j}} G^{-1}(p) \, \mathrm{d}p - \mu,$$

where $E(L_{i,\mathrm{res}}) = E\{E(U_{i,\mathrm{res}} \mid Y_i)\} = E(U_{i,\mathrm{res}}) = 0$.

If the mean of G is its median, which is true for logit and probit link functions, then $\mu = G^{-1}(1/2)$. And if the interval $(\gamma_{i,j-1}, \gamma_{i,j})$ is small, then

$$\frac{1}{p_{i,j}} \int_{\gamma_{i,j-1}}^{\gamma_{i,j}} G^{-1}(p) \, \mathrm{d}p - \mu \approx G^{-1}(m_{i,j}) - G^{-1}\left(\frac{1}{2}\right),$$

where $m_{i,j} = (\gamma_{i,j-1} + \gamma_{i,j})/2$. Since $R_i/2 = m_{i,j} - \frac{1}{2}$, our residual is equivalent to comparing $m_{i,j}$ with $1/2$ on the probability scale. Therefore, our residual captures information similar to that of a latent-variable residual, but on a probability scale irrespective of the choice of link function; see Fig. 1.
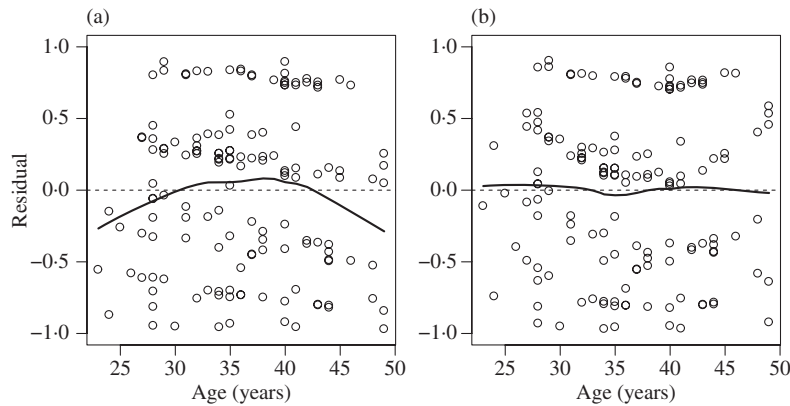
Fig. 2. Residual-by-predictor plots with age included in proportional odds models
with a linear term (a) and with linear and quadratic terms (b). Lowess curves (solid)
are added. A horizontal line (dashed) at zero is included for reference.

3. USE OF THE RESIDUAL IN MODEL DIAGNOSTICS

3·1. *Residual-by-predictor plots*

When an ordinal regression model is correct, $E(L_{i,\text{res}}) = 0$ and $L_{i,\text{res}} = E(U_{i,\text{res}} \mid Y_i)$ is a categorization of $U_{i,\text{res}}$, which follows the same distribution across all subjects. Therefore, $L_{i,\text{res}}$ may have similar distributions across subjects. In addition, we have shown that $E(R_i) = 0$, the range of R_i is symmetric about zero, and R_i and $L_{i,\text{res}}$ are closely related. Therefore, a plot of $\hat{r}$, or its latent-variable version $\hat{l}_{\text{res}}$, versus a predictor can be useful for visually detecting if there is any additional effect of that predictor, such as nonlinearity, on the outcome. This plot is referred to as a residual-by-predictor plot.

We use data from a study of HIV-infected women in Zambia (Parham et al., 2006) to demonstrate an application of residual-by-predictor plots. Cervical specimens from 145 women were examined using cytology and categorized into five ordered stages. We fit proportional odds models to assess the association between stage of cervical lesions and age after adjusting for CD4 count. Under the assumption of a linear relationship of age with the log-odds of severity of lesions, the model tends to overpredict severity of lesions at low and high ages, as shown in Fig. 2. If age is put into the model with linear and quadratic terms, the residuals are much more uniform across ages. The plots are similar using standardized or latent-variable residuals, shown in the Supplementary Material.

3·2. *Partial regression plots*

When fitting a regression model of an ordinal outcome Y on a covariate X and other covariates Z, we may want to examine whether X is associated with Y after adjusting for the effects of Z. To do this, we first fit an ordinal regression model of Y on Z to obtain residuals $\hat{r}_y$, then fit an appropriate regression model of X on Z to obtain residuals $\hat{r}_x$, and plot $\hat{r}_y$ versus $\hat{r}_x$. This plot is called a partial regression plot. Let R_Y be the random variable as defined in (1) and R_X be the residual random variable for the model for $X \mid Z$, of which $\hat{r}_x$ is a realization. When the model for $Y \mid Z$ is correct and Y and X are independent conditional on Z, R_Y and R_X are independent given Z. Thus, $E(R_Y R_X) = E\{E(R_Y R_X \mid Z)\} = E\{E(R_Y \mid Z)E(R_X \mid Z)\} = 0$ by Property 12, $E(R_Y) = E\{E(R_Y \mid Z)\} = 0$ and $\text{cov}(R_Y, R_X) = E(R_Y R_X) - E(R_Y)E(R_X) = 0$; that is, R_Y and R_X are uncorrelated. Hence, the partial regression plot can provide a visual inspection on whether X remains a useful predictor of Y after the effects of Z have been adjusted for. Although the validity of this plot does not depend on the type of residual for X, its effectiveness might.

We use the same dataset as that in § 3·1 to demonstrate partial regression plots. Figure 3(a) is a partial regression plot for the association between cervical lesions and CD4 T-cell count. Residuals from a proportional odds regression of cervical lesions on age and squared age are plotted against residuals from
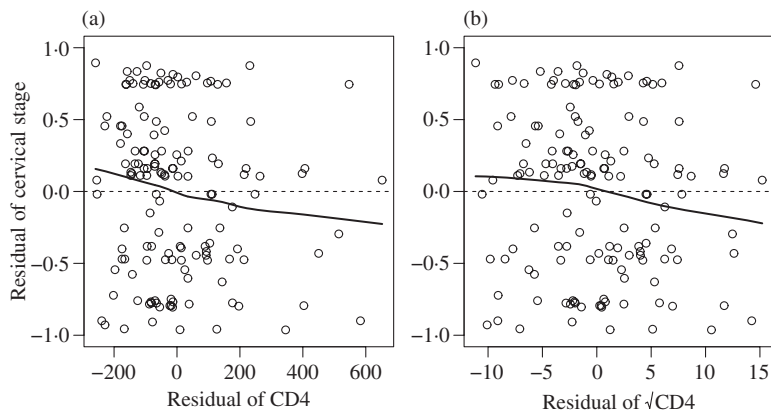
Fig. 3. Partial regression plots looking at the residual association of cervical
lesions with (a) CD4 and (b) square-root transformed CD4 after adjusting for
age, with both linear and quadratic terms. Lowess curves (solid) are added. A
horizontal line (dashed) at zero is included for reference.

a linear regression of CD4 count on age and squared age. There appears to be correlation between these
residuals, so it may be useful to include CD4 count in the model for cervical lesions. Partial regression
plots may also be useful for detecting outliers: no single observation appears to be overly influential in this
analysis, although the skewed nature of the CD4 residuals suggests that a square root transformation of
this variable may lead to a better fit, as suggested by Fig. 3(b). Additional examples of partial regression
plots using our residuals are in Li & Shepherd (2010).

4. DISCUSSION

Our residual is effectively defined on the probability scale of the fitted distribution, and can be
extended to any regression analysis in which fitted distributions are calculated. One potential advantage
of probability-scale residuals is that they can be defined for regression analyses in which the fitted distri-
butions are not completely determined, such as in models of censored data or quantile regression.

Our new residual offers a general solution for how to include an ordinal predictor in a regression model.
Traditional approaches treat an ordinal predictor as either numerical, enforcing a linearity assumption, or
categorical, ignoring order information. An alternative approach would be to fit an appropriate regression
model of the outcome on other covariates Z and an ordinal regression model of the ordinal predictor on
Z, and test for correlation between the residuals of the models. The partial regression plot is the graphical
counterpart of this approach. We have shown this approach to be robust and powerful when the outcome
variable is also ordinal (Li & Shepherd, 2010), and we are studying this approach for other outcome types.

One limitation of our residual is that it seems not useful for checking the proportional odds assumption
in proportional odds models, as the cell-wise information necessary for investigating this assumption is
collapsed into a single value.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes additional figures and a discussion of
residuals in Liu et al. (2009).

APPENDIX

Proofs of properties

Properties 2, 4, 7, 11–13 are obvious.

Proof of Property 1. Since $r(j + 1, F) - r(j, F) = \gamma_{j+1} - \gamma_{j-1} \geqslant 0$, we have $r(j, F) \leqslant r(j + 1, F)$. Since $\gamma_1 \geqslant 0$ and $\gamma_{s-1} \leqslant 1, r(1, F) = \gamma_1 - 1 \geqslant -1$ and $r(s, F) = \gamma_{s-1} \leqslant 1$. □

Proof of Property 3. Let $\gamma_{j|F}$ and $\gamma_{j|G}$ be the cumulative probabilities for distributions F and G, respectively. Let $t = s - j + 1$. Then $\gamma_{t-1|G} = 1 - \gamma_{j|F}$ and $\gamma_{t|G} = 1 - \gamma_{j-1|F}$. Therefore, $r(t, G) = \gamma_{t-1|G} + \gamma_{t|G} - 1 = 1 - \gamma_{j|F} - \gamma_{j-1|F} = -r(j, F)$. □

Proof of Property 5. The results for $j < t$ and $j > t$ are obvious. For $j = t$, since $r(t + 1, F) = r(t, F) + (p_t + p_{t+1})$ and $r(t, F) = 2\gamma_{t-1} + p_t - 1$, we have $\{p_t r(t, F) + p_{t+1} r(t + 1, F)\}/(p_t + p_{t+1}) = r(t, F) + p_{t+1} = 2\gamma_{t-1} + (p_t + p_{t+1}) - 1 = r(t, G)$. □

Proof of Property 6. Let $f(p) = r\{2, (p, 1 - p)\}$ for $0 < p < 1$. We first show that $f(p)/p$ is a constant. By Property 3′, $r\{1, (p, 1 - p)\} = -f(1 - p)$. By Property 4′, $f'(1 - p) = f'(p)$ and thus $f(1 - p) + f(p) \equiv c$ is a constant as its derivative is always zero. By Properties 2 and 5, $0 = r(1, p + 1 - p) = -pf(1 - p) + (1 - p)f(p)$, and thus $f(1 - p) = f(p)(1 - p)/p$. Then, $c \equiv f(1 - p) + f(p) = f(p)/p$ and $f(p) = cp$.

By Property 5, $r(j, F) = r\{2, (a, p_j, b)\}$, where $a = \gamma_{j-1}$ and $b = 1 - \gamma_j$. By Properties 2 and 5,

$$0 = a \times r\{1, (a, p_j, b)\} + p_j \times r\{2, (a, p_j, b)\} + b \times r\{3, (a, p_j, b)\}$$

$$= a \times r\{1, (a, 1 - a)\} + p_j \times r\{2, (a, p_j, b)\} + b \times r\{2, (1 - b, b)\}$$

$$= -af(1 - a) + p_j r(j, F) + bf(1 - b).$$

Thus, $r(j, F) = \{af(1 - a) - bf(1 - b)\}/p_j = c\{a(1 - a) - b(1 - b)\}/p_j = c\{a(p_j + b) - b(a + p_j)\}/p_j = c(a - b) = c\{\gamma_{j-1} - (1 - \gamma_j)\}$. It is easy to show the reverse. □

Proof of Property 8. Since $\sum_j p_j \gamma_{j-1} = \sum_j p_j(\sum_{k<j} p_k) = \sum_{j_1 < j_2} p_{j_1} p_{j_2}$ and $\sum_j p_j(1 - \gamma_j) = \sum_j p_j(\sum_{k>j} p_k) = \sum_{j_1 < j_2} p_{j_1} p_{j_2}$, we have $E(R) = \sum_j p_j\{\gamma_{j-1} - (1 - \gamma_j)\} = 0$. □

Proof of Property 9. Since $E(R) = 0$ and $\gamma_{a-1} - (1 - \gamma_a) = \sum_{b:b \neq a} p_b \times \text{sign}(a - b)$,

$$\text{var}(R) = E(R^2) = \sum_{a=1}^{s} p_a\{\gamma_{a-1} - (1 - \gamma_a)\}^2$$

$$= \sum_{a=1}^{s} p_a \left\{ \sum_{b:b \neq a} p_b^2 + \sum_{b,c:b<c,b \neq a,c \neq a} 2p_b p_c \times \text{sign}(a - b)\text{sign}(a - c) \right\}$$

$$= \sum_{a \neq b} p_a p_b^2 + 2 \sum_{a<b<c} p_a p_b p_c.$$

Since $1 = (\sum_a p_a)^3 = \sum_a p_a^3 + 3 \sum_{a \neq b} p_a p_b^2 + 6 \sum_{a<b<c} p_a p_b p_c$, $\text{var}(R) = (1 - \sum_a p_a^3)/3$.

For the alternative expression, since $\sum_{a<b} p_a p_b = \sum_{a<b} p_a p_b \sum_c p_c = \sum_{a<b}(p_a p_b^2 + p_a^2 p_b + \sum_{c \neq a,b} p_a p_b p_c) = \sum_{a \neq b} p_a p_b^2 + 3 \sum_{a<b<c} p_a p_b p_c$,

$$\text{var}(R) = \sum_{a<b} p_a p_b - \sum_{a<b<c} p_a p_b p_c = \sum_{a<b} p_a p_b \gamma_b = \sum_{b=2}^{s} p_b \gamma_{b-1} \gamma_b = \sum_{b=1}^{s} p_b \gamma_{b-1} \gamma_b.$$

□

Proof of Property 10. To maximize $\mathrm{var}(R)$ subject to the constraint $\sum p_j = 1$, we employ a Lagrange multiplier in the function

$$f(p_1, \ldots, p_s, \lambda) = \mathrm{var}(R) - \lambda \left(\sum p_j - 1 \right) = \frac{1}{3} \left(1 - \sum p_j^3 \right) - \lambda \left(\sum p_j - 1 \right).$$

Setting $\partial f / \partial \lambda = 0$ gives the constraint $\sum p_j - 1$. Setting $\partial f / \partial p_j = -p_j^2 - \lambda = 0$ for all j leads to $p_1 = \cdots = p_s = 1/s$ and $\mathrm{var}(R) = (1 - 1/s^2)/3$. $\qquad\square$

Proof of Properties 14 *and* 15. For proportional odds models, it can be shown that

$$\frac{\partial \gamma_{i,j}}{\partial \alpha_j} = \gamma_{i,j}(1 - \gamma_{i,j}), \quad \frac{\partial \gamma_{i,j}}{\partial \alpha_k} = 0 \quad (j \neq k).$$

For convenience, let $\gamma_{i,0} = 0$. For subject i, let $l_i = \log(p_{i,y_i}) = \log(\gamma_{i,y_i} - \gamma_{i,y_i-1})$ be the loglikelihood. Let $l = \sum_i l_i$ and $\hat{\theta}$ be the maximum likelihood estimate. Then

$$\sum_{j=1}^{s-1} \frac{\partial l_i}{\partial \alpha_j} \bigg|_{\hat{\theta}} = \frac{\hat{\gamma}_{i,y_i}(1 - \hat{\gamma}_{i,y_i})}{\hat{p}_{i,y_i}} - \frac{\hat{\gamma}_{i,y_i-1}(1 - \hat{\gamma}_{i,y_i-1})}{\hat{p}_{i,y_i}}$$

$$= \frac{(\hat{\gamma}_{i,y_i-1} + \hat{p}_{y_i})(1 - \hat{\gamma}_{i,y_i}) - \hat{\gamma}_{i,y_i-1}(\hat{p}_{y_i} + 1 - \hat{\gamma}_{i,y_i})}{\hat{p}_{i,y_i}}$$

$$= (1 - \hat{\gamma}_{i,y_i}) - \hat{\gamma}_{i,y_i-1} = -\hat{r}_i$$

and

$$\sum_{i=1}^{n} \hat{r}_i = -\sum_{i=1}^{n} \sum_{j=1}^{s-1} \frac{\partial l_i}{\partial \alpha_j} \bigg|_{\hat{\theta}} = -\sum_{j=1}^{s-1} \frac{\partial l}{\partial \alpha_j} \bigg|_{\hat{\theta}} = 0.$$

$\qquad\square$

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.

AGRESTI, A. (2002). *Categorical Data Analysis*, 2nd edn. Hoboken, NJ: Wiley.

BROCKETT, P. L. & LEVINE, A. (1977). On a characterization of ridits. *Ann. Statist.* **5**, 1245–8.

BROSS, I. D. J. (1958). How to use ridit analysis. *Biometrics* **14**, 18–38.

LI, C. & SHEPHERD, B. E. (2010). Test of association between two ordinal variables while adjusting for covariates. *J. Am. Statist. Assoc.* **105**, 612–20.

LIU, I., MUKHERJEE, B., SUESSE, T., SPARROW, D. & PARK, S. K. (2009). Graphical diagnostics to check model misspecification for the proportional odds regression model. *Statist. Med.* **28**, 412–29.

MCCULLAGH, P. (1980). Regression models for ordinal data (with discussions). *J. R. Statist. Soc.* B **42**, 109–42.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd edn. London: Chapman & Hall.

PARHAM, G. P., SAHASRABUDDHE, V. V., MWANAHAMUNTU, M. H., SHEPHERD, B. E., HICKS, M. L., STRINGER, E. M. & VERMUND, S. H. (2006). Prevalence and predictors of squamous intraepithelial lesions of the cervix in HIV-infected women in Lusaka, Zambia. *Gynecol. Oncol.* **103**, 1017–22.

THERNEAU, T. M., GRAMBSCH, P. M. & FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–60.