BIOMETRICS 46, 1171-1178
December 1990

Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression

Rollin Brant

Department of Community Health Sciences, University of Calgary,
3330 Hospital Drive N.W., Calgary, Alberta T2N 4N1, Canada

SUMMARY

The proportional odds model for ordinal logistic regression provides a useful extension of the binary logistic model to situations where the response variable takes on values in a set of ordered categories. The model may be represented by a series of logistic regressions for dependent binary variables, with common regression parameters reflecting the proportional odds assumption. Key to the valid application of the model is the assessment of the proportionality assumption. An approach is described arising from comparisons of the separate (correlated) fits to the binary logistic models underlying the overall model. Based on asymptotic distributional results, formal goodness-of-fit measures are constructed to supplement informal comparisons of the different fits. A number of proposals, including application of bootstrap simulation, are discussed and illustrated with a data example.

1. Introduction

The proportional odds model for ordinal logistic regression described by McCullagh (1980) provides a useful extension of the binary logistic model to situations where the response variable takes on ordered categorical values. This paper considers an approach to assessing the goodness of fit of such models, based on comparing fits to the binary logistic models that are subsidiary to the overall model. Formal goodness-of-fit measures are described, together with informal methods for discerning and describing lack of fit. The paper is organized as follows. In Section 2, the proportional odds model is briefly reviewed. In Section 3, the basic approach is described. Section 4 provides illustration with a data example.

2. McCullagh's Proportional Odds Model

Suppose that one has N sets of independent observations on a response variable y taking ordinal values 1 to k and $\mathbf{x}$ a p-vector of explanatory variables. The proportional odds model (McCullagh, 1980) is based on consideration of the cumulative distribution probabilities $\gamma_j = \Pr\{y \leq j\}$ and takes the form

$$\operatorname{logit}(\gamma_j) = \log[\gamma_j/(1 - \gamma_j)] = \theta_j - \boldsymbol{\beta}^t\mathbf{x}$$

where the p-vector $\boldsymbol{\beta}$ and $\theta_1 < \theta_2 < \cdots < \theta_{k-1}$ represent unknown parameters. The model is also referred to as the grouped continuous model, since it can be constructed by supposing that observations on a latent variable following a logistic distribution with conditional mean $\eta = \boldsymbol{\beta}^t\mathbf{x}$, have been grouped in intervals with cutpoints θ_j ($j = 1, \ldots, k - 1$). For a comparison of this model with some of its main competitors see Greenwood and Farewell (1989).

Key words: Goodness of fit; Ordinal data; Proportional odds model.

While general approaches to goodness of fit, such as those described in Tsiatis (1980) and Lemeshow and Hosmer (1982), can be applied to the model, none of them apply directly to the proportionality assumption implicit in the use of common values of $\boldsymbol{\beta}$. One approach to checking the proportionality assumption is to assume

$$\operatorname{logit}(\gamma_j) = \log[\gamma_j/(1 - \gamma_j)] = \theta_j - \boldsymbol{\beta}_j^{\mathrm{t}}\mathbf{x} \tag{2.1}$$

(i.e., augment the model, incorporating separate $\boldsymbol{\beta}_j$'s) and then apply conventional likelihood methods for assessing the hypothesis

$$H_0: \quad \boldsymbol{\beta}_j = \boldsymbol{\beta}, \qquad j = 1, \ldots, k - 1.$$

Algorithms for fitting (2.1) have been implemented in GLIM as discussed in Hutchison (1985) and Ekholm and Palmgren (1989). Similarly, Armstrong and Sloane (1989) discuss a simple example where it suffices to incorporate the separate estimates in a standard likelihood-based comparison of "saturated" and reduced models to assess the viability of the proportional odds assumption.

3. Assessing Proportionality Based on Separate Fits

The approach proposed here is based on viewing the augmented model as describing a set of $k - 1$ logistic regressions, for variables z_j ($j = 1, \ldots, k - 1$) defined by

$$z_j = \begin{cases} 1 & y > j \\ 0 & y \leqslant j \end{cases}$$

with success probability, $\pi_j = \Pr\{z_j = 1\} = 1 - \gamma_j$ satisfying

$$\operatorname{logit}(\pi_j) = \log\{\pi_j/(1 - \pi_j)\} = -\theta_j + \boldsymbol{\beta}_j^{\mathrm{t}}\mathbf{x}.$$

A natural approach to assessing proportionality is to examine and compare the conventional logistic fits for the dichotomized responses, z_j ($j = 1, \ldots, k - 1$). In most cases these "separate" estimates will not correspond to the overall maximum likelihood estimates for the augmented model (2.1) since they will not necessarily obey the constraints for monotonicity and may yield fitted values of the probabilities $\pi_{j-1} - \pi_j$ that are negative. Thus standard likelihood-based procedures for assessing H_0, such as likelihood ratio tests, cannot be based directly on the separate fits. However, the separate fits do provide a workable foundation for assessing the proportionality assumption based on the following distributional results, proven in the Appendix.

Let $\mathbf{X}$: $n \times p$ denote the matrix of explanatory observations, and $\mathbf{X}_+$ represent the result of augmenting $\mathbf{X}$ on the left with a column of 1's. Denoting the separate maximum likelihood estimates by $\tilde{\boldsymbol{\beta}}_j$ ($j = 1, \ldots, k - 1$), $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1^{\mathrm{t}} \mid \tilde{\boldsymbol{\beta}}_2^{\mathrm{t}} \mid \ldots \mid \tilde{\boldsymbol{\beta}}_{k-1}^{\mathrm{t}})^{\mathrm{t}}$ has a distribution that is asymptotically multivariate normal with $\mathrm{E}\{\tilde{\boldsymbol{\beta}}_j\} \approx \boldsymbol{\beta}_j$. The asymptotic covariances, $\operatorname{cov}\{\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{\beta}}_l\}$, are obtained by deleting the first row and column of

$$(\mathbf{X}_+^{\mathrm{t}}\mathbf{W}_{jj}\mathbf{X}_+)^{-1}\mathbf{X}_+^{\mathrm{t}}\mathbf{W}_{jl}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{t}}\mathbf{W}_{ll}\mathbf{X}_+)^{-1} \tag{3.1}$$

where $\mathbf{W}_{jl}$: $n \times n$ is diagonal with typical entry $\pi_l - \pi_j\pi_l$ for $j \leqslant l$. Similar expressions are also obtainable for the covariance structure of estimates based on other link functions, as shown in the Appendix. Thus the following methods are easily generalizable to forms of McCullagh's model based on other distributions in addition to the logistic.

When constructing sets of H_0, it is natural to substitute maximum likelihood estimates from the proportional odds model into the $\mathbf{W}_{jl}$'s to provide $\hat{V}(\tilde{\boldsymbol{\beta}}_j, \tilde{\boldsymbol{\beta}}_l)$, giving rise in turn to the overall estimated covariance matrix $\hat{\mathbf{V}}(\tilde{\boldsymbol{\beta}})$. Thus one can obtain standard errors for assessing the statistical significance of any apparent discrepancies between the estimated coefficients of the separate fits. To avoid post hoc selection, one can construct an omnibus test by combining contrasts in the $\hat{\boldsymbol{\beta}}_j$'s into a quadratic form, giving rise to an asymptotic χ^2 test. More explicitly, one can take the $(k-2)p \times (k-1)p$ contrast matrix,

$$\mathbf{D} = \begin{bmatrix} I & -I & 0 & \cdots & 0 \\ I & 0 & -I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ I & 0 & 0 & \cdots & -I \end{bmatrix},$$

and form the associated Wald-type goodness-of-fit statistic

$$X^2 = (\mathbf{D}\tilde{\boldsymbol{\beta}})^{\mathrm{t}}[\mathbf{D}\hat{\mathbf{V}}(\tilde{\boldsymbol{\beta}})\mathbf{D}^{\mathrm{t}}]^{-1}(\mathbf{D}\tilde{\boldsymbol{\beta}}),$$

which will be asymptotically χ^2 on $(k-2)p$ degrees of freedom under the null hypothesis. If X^2 is found to be significant, individual differences $\tilde{\boldsymbol{\beta}}_j - \tilde{\boldsymbol{\beta}}_l$ may be considered in relation to their approximate standard errors to elucidate the nature of the lack of fit.

The above test suffers from two defects common to many omnibus goodness-of-fit procedures. First, if either of k or p is large, the degrees of freedom above will be such that one cannot expect the test to be very powerful. Second, even if the test is sufficiently powerful to detect departures from proportionality, inspection of the individual components of the test statistic (i.e., the differences $\tilde{\boldsymbol{\beta}}_j - \tilde{\boldsymbol{\beta}}_l$) may provide no clear indication as to the nature of the discrepancy detected. Thus it is desirable to consider the specific types of departure likely to be encountered, and modify the above approach accordingly.

Many types of departure from the specified model will lead to differential biases in the separate estimates, so a wide range of possibilities can be considered. It is most appropriate to consider departures that still lie in the general framework of an underlying linear model. Within this context, one immediately considers the following standard list of model inadequacies:

1. Misspecification of the linear predictor, $\eta = \boldsymbol{\beta}^{\mathrm{t}}\mathbf{x}$.
2. Nonhomogeneous dispersion of the latent variable with varying $\mathbf{x}$.
3. Misspecification of the distributional form for the latent variable, i.e., nonlogistic link function.

The basic strategy considered here is not likely to lead to sensitive procedures for assessing the first inadequacy, since any biases introduced will affect all the separate estimates more or less equally. Neither is the second likely to be reflected in a simple way in the separate estimates. The approach does hold promise for the third inadequacy, as the following shows. Suppose that a latent variable construction holds with cumulative distribution function F. The logistic transform of the true underlying probabilities takes the form

$$\lambda_j = \operatorname{logit}(\pi_j) = \log\left\{\frac{1 - F(\theta_j - \eta)}{F(\theta_j - \eta)}\right\}.$$

Letting $h(\cdot) = \log[\{1 - F(\cdot)\}/F(\cdot)]$ and expanding about $\eta = \bar{\eta}$, one has

$$\lambda_j \approx h(\theta_j) - h'(\theta_j)(\eta - \bar{\eta}) = -\theta_j' + \phi_j\boldsymbol{\beta}^{\mathrm{t}}\mathbf{x},$$

where $\theta_j' = h(\theta_j) - h'(\theta_j)\bar{\eta}$ and $\phi_j = -h'(\theta_j) > 0$. Thus, misspecification of this type will be reflected in the alternative

$$H_{\mathrm{A}}:\quad \boldsymbol{\beta}_j = \phi_j\boldsymbol{\beta}, \qquad j = 1, \ldots, k-1. \tag{3.2}$$

This model is similar in form to Anderson's (1984) stereotype ordered regression model, but differs in that it applies to cumulative, rather than separate probabilities.

Shifts of this form can be expected to be substantial only when the form of F departs dramatically from the nominal logistic form, such as when F has substantial skewness, as in the case of the extreme value distribution, which leads to use of the complementary loglog link function, $\log(-\log(1 - \pi))$. However, other forms of misspecification will also reflect themselves in similar departures. Consider the problem of differential misclassification in the y observations. If one observes y^* when the true response is y, the success probabilities for the subsidiary binary responses, z_j^*, take the form

$$\pi_j^* = \Pr\{y^* > j \mid y > j\} \times \pi_j + \Pr\{y^* > j \mid y \leq j\} \times (1 - \pi_j).$$

Expanding the logit of the above probability about $\bar{\eta}$, one is led to an attenuated version of the original model of the form (3.2), with $\phi_j \leq 1$ ($j = 1, \ldots, k - 1$). Thus a test with power in the direction of (3.2) will be useful in detecting at least two forms of departure.

The relevant test is motivated by expressing (3.2) as $\mathrm{E}(\tilde{\boldsymbol{\beta}}_j) \approx \boldsymbol{\beta}_1 + \delta_j \boldsymbol{\beta}_1$ ($j = 1, \ldots, k - 1$) with $\delta_1 = 0$. This has the form of a nonlinear regression equation for $\tilde{\boldsymbol{\beta}}$. A test for $\delta_j = 0$ ($j = 2, \ldots, k - 1$) can be constructed by performing a weighted regression of $\tilde{\boldsymbol{\beta}}$ on the $(k - 1)p \times (p + k - 2)$ design matrix

$$\mathbf{D} = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & \hat{\boldsymbol{\beta}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ I & 0 & \cdots & \hat{\boldsymbol{\beta}} \end{bmatrix}$$

weighting inversely by $\hat{\mathbf{V}}(\tilde{\boldsymbol{\beta}})$. The last $(k - 2)$ entries of the resultant regression estimate provide an estimate $\hat{\boldsymbol{\delta}}$ of $(\delta_2, \ldots, \delta_{k-1})$ and an estimated (asymptotic) variance–covariance matrix, $\hat{\mathbf{V}}(\hat{\boldsymbol{\delta}})$, can be obtained in the usual way. This leads to the statistic $X_{\mathrm{a}}^2 = \hat{\boldsymbol{\delta}}^{\mathrm{t}} \hat{\mathbf{V}}(\hat{\boldsymbol{\delta}})^{-1} \hat{\boldsymbol{\delta}}$, which is asymptotically χ^2 on $k - 2$ degrees of freedom, under H_0.

If the number of parameters is large in relation to sample size, the asymptotic distributional approximations may not be reliable. Bootstrap simulation may be performed to obtain the finite-sample null distribution at the proportional odds maximum likelihood estimates. Since each bootstrap replicate requires k iterative fits, this may not be computationally feasible. The computational cost can be mitigated somewhat by replacing each $\tilde{\boldsymbol{\beta}}_j$ by a corresponding "one-step" estimate (see Appendix). Whether or not the fully iterated estimates are used, two approaches present themselves. The simplest, conceptually, is to generate bootstrap replicates of X^2 or X_{a}^2 (or "one-step" versions) based on simulated observations from the original overall maximum likelihood fit. The bootstrap significance level is then obtained directly by referring the observed value to the empirical distribution. A computationally cheaper alternative is to generate only bootstrap replicate values of the basic estimates, $\tilde{\boldsymbol{\beta}}_j - \tilde{\boldsymbol{\beta}}_l$, in the case of X^2, or $\hat{\boldsymbol{\delta}}_{\mathrm{b}}$ in the case of X_{a}^2. The observed estimate can then be compared to the empirical distribution, either directly, if the estimate is scalar, or in terms of its Mahalanobis distance from the mean of the empirical distribution. Both approaches are illustrated in the following example.

4. An Example

The data considered here represent observations on 83 livers donated for transplantation. The response, y, represents a clinical assessment made on the liver shortly after transplantation, and takes values 0 to 2 with larger values corresponding to undesirable outcomes. The variables thought by the investigators to be potentially important in determining outcome were age of the donor (AGE), the cause of death of the donor (DX), cold ischemic time (CIT), which is the length of time between removal and transplantation of the liver, and which of two types of solution was used for maintaining the liver pretransplant (SOLN).

In addition, it was thought that results for livers from young donors might differ, so an indicator for donors younger than 12 years was introduced (PED). As well, an "interaction" term (SOLN $\times$ CIT) was introduced, since the two types of solution were thought to have different effects on the liver over time. The calculations were performed using the "S" language (Becker, Chambers, and Wilks, 1988).

The two resulting logistic fits are given in Table 1, followed by the combined fit under the proportional odds model in Table 2. A number of discrepancies are apparent in the tables, including for example, differences in the PED term. (It was these differences that motivated the present paper.) As previously mentioned, standard errors can be obtained for particular differences in coefficients. For example, the difference in the PED coefficients is 3.17 with a standard error of 1.15, suggesting possible lack of fit. To assess overall significance, the more formal goodness-of-fit statistics were computed, yielding $X^2 = 11.2$, with observed level of significance .08 when referred to $\chi^2_{(6)}$.

Table 1
Fitted coefficients, separate binary logistic fits

Response	Intercept	AGE	PED	CIT	SOLN	SOLN $\times$ CIT	DX
$\mathbf{z}_1$	−3.383	.08748	1.612	.2414	2.283	−.3032	−.9513
(s.e.)	1.68	.03338	1.026	.2084	1.611	.2675	.5507
$\mathbf{z}_2$	−12.09	.157	4.784	1.019	6.962	−1.012	−1.968
(s.e.)	3.287	.0474	1.512	.3963	3.046	.4514	.8052
Ratio (2/1)	—	1.7947	2.9677	4.2212	3.0495	3.3377	2.0687
(s.e.)	—	.9345	2.465	4.6788	2.9246	3.8807	2.5207

Table 2
Fitted coefficients from proportional odds model

	θ_1	θ_2	AGE	PED	CIT	SOLN	SOLN $\times$ CIT	DX
Estimate	5.36	7.2	.115	2.79	.492	3.37	−.501	−1.33
s.e.	1.64	1.73	.0319	1.05	.201	1.51	.251	.536

The ratios of the estimated β's are also given in Table 1, together with asymptotic standard errors derived by applying the delta method to $\mathbf{V}(\tilde{\boldsymbol{\beta}})$. The estimates all exceed 1 and are of a comparable magnitude, taking into account their relative imprecision as indicated by their large standard errors. Thus, the relevance of (3.2) as an alternative is indicated. The observed value of X^2_{a} is 5.04 with observed significance level .025 based on the $\chi^2_{(1)}$ approximation. Due to the number of parameters involved (eight parameters in the ordinal logistic model plus one proportionality parameter), the viability of the asymptotic approximation is suspect. Two hundred bootstrap replications of the one-step version of X^2_{a} were computed, simulating from the fit in Table 2. The QQ plot in Figure 1 of the sorted values of X^2_{a} versus $\chi^2_{(1)}$ quantiles suggests that the asymptotic approximation is roughly adequate, at least for the one-step test. The observed value of the one-step statistic is 4.11, with nominal significance level .042. The bootstrap level of significance is .055 if one refers to the bootstrap distribution of X^2_{a} or .065 if one compares the observed $\hat{\boldsymbol{\delta}}$ to its bootstrap distribution.

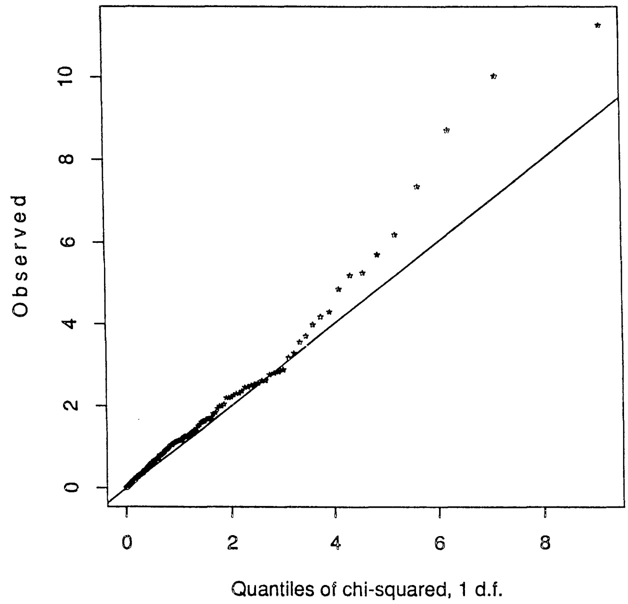

Figure 1. QQ plot comparing results of 200 bootstrap replications of X_a^2 (one-step version) to quantiles of $\chi^2_{(1)}$.

5. Discussion

As previously mentioned, assessment of the proportionality assumption can also be based on fitting the augmented models (2.1), as in Hutchison (1985) and Ekholm and Palmgren (1989). Similarly, a more directed approach can be based on fitting (3.2). The augmented model approach is attractive in that it provides a more standard theoretical framework for developing tests. One drawback, however, is that specialized algorithms must be developed to fit the augmented models. A more serious problem is inherent in the models themselves. For example, if one wishes to extend the use of model (2.1) beyond values of **x**'s actually observed, the $\boldsymbol{\beta}_j$'s must be constrained to ensure monotonicity of the extrapolated γ_j's. Similar difficulties pertain to (3.2). Depending on the range of admissible values of **x**, this can lead to technical difficulties in fitting and the need for nonstandard likelihood theory to allow for the possibility of estimates falling on the boundary of the parameter space. It may be best then to view (2.1) and (3.2) not as scientifically meaningful models, but as directional alternatives helpful in validating the simpler proportional odds model.

The approach suggested here has the advantage of requiring only the software for fitting the logistic and proportional odds model, plus some basic matrix manipulations. Distributional results based on asymptotic calculations are easy to obtain and apply, while bootstrap simulations are also feasible. By working through the underlying binary models, the approach naturally augments the more informal examination of the data, so that users of the proportional odds model can evaluate the aptness of the model while gaining added insight into the complexities of the data.

RÉSUMÉ

Le modèle des chances proportionnelles pour la régression logistique ordinale permet d'étendre avec profit le modèle logistique binaire à des situations où la variable réponse prend ses valeurs dans une suite de catégories ordonnées. Le modèle peut être représenté par différentes régressions logistiques de variables dépendantes binaires, avec des paramètres de régression communs traduisant l'hypothèse de proportionnalité des chances. La vérification de l'hypothèse de proportionnalité est essentielle à une application valable du modèle. L'approche qui est décrite repose sur les comparaisons des ajustements (corrélés) obtenus séparément pour chaque modèle logistique binaire définissant le modèle global. Des résultats concernant les propriétés asymptotiques des distributions permettent de construire explicitement des mesures d'adéquation pour compléter les comparaisons informelles des différents ajustements. Différents développements, dont l'utilisation de simulations par bootstrap, sont discutés et illustrés par un exemple réel.

REFERENCES

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B* **46,** 1–40.

Armstrong, B. and Sloane, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology* **129,** 191–204.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language*. Pacific Grove, California: Wadsworth.

Ekholm, A. and Palmgren, J. (1989). Regression models for an ordinal response are best handled as nonlinear models. *GLIM Newsletter* **18,** 31–35.

Greenwood, C. and Farewell, V. (1989). A comparison of regression models for ordinal data in analysis of transplanted kidney function. *Canadian Journal of Statistics* **16,** 325–336.

Hutchison, D. (1985). Ordinal variable regression using the McCullagh (proportional odds) model. *GLIM Newsletter* **9,** 9–17.

Lemeshow, S. and Hosmer, D. W. (1982). A review of goodness-of-fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology* **115,** 92–106.

McCullagh, P. (1980). Regression models for ordinal data (with Discussion). *Journal of the Royal Statistical Society, Series B* **42,** 109–142.

Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika* **67,** 250–251.

Received June 1989; *revised April* 1990; *accepted July* 1990.

APPENDIX

Let

$$\boldsymbol{\beta}_{+j} = \begin{pmatrix} -\theta_j \\ \boldsymbol{\beta}_j \end{pmatrix}$$

represent the unknown parameters of the dichotomous logistic regression model for $\mathbf{z}_j$. Let $\mathbf{Z}_j$, $(j = 1, \ldots, k-1)$ represent the $k-1$ separate N-vectors of independent $\mathbf{z}_j$ observations and denote the corresponding mean vector $\mathrm{E}\{\mathbf{Z}_j\}$ by $\boldsymbol{\Pi}_j$, with entries defined by (2.1). Applying standard arguments, the separately fitted maximum likelihood estimates,

$$\tilde{\boldsymbol{\beta}}_{+j} = \begin{pmatrix} -\tilde{\theta}_j \\ \tilde{\boldsymbol{\beta}}_j \end{pmatrix}$$

are asymptotically unbiased and satisfy

$$(\tilde{\boldsymbol{\beta}}_{+j} - \boldsymbol{\beta}_{+j}) = (\mathbf{X}_+^t \mathbf{W}_{jj} \mathbf{X}_+)^{-1} \mathbf{X}_+^t (\mathbf{Z}_j - \boldsymbol{\Pi}_j) + o_p(n^{-1/2}),$$

where $\mathbf{W}_{jj}$ is defined in Section 3. The asymptotic covariance of $\tilde{\boldsymbol{\beta}}_j$ and $\tilde{\boldsymbol{\beta}}_l$, $\mathbf{V}_{jl}$, is obtained by deleting the first row and column of

$$(\mathbf{X}_+^t \mathbf{W}_{jj} \mathbf{X}_+)^{-1} \mathbf{X}_+^t \operatorname{cov}(\mathbf{Z}_j, \mathbf{Z}_l) \mathbf{X}_+ (\mathbf{X}_+^t \mathbf{W}_{ll} \mathbf{X}_+)^{-1},$$

which gives the result (3.1) after noting that $\mathbf{W}_{jl} = \operatorname{cov}(\mathbf{Z}_j, \mathbf{Z}_l)$.

If H_0 holds, then a set of estimates asymptotically equivalent to $\tilde{\boldsymbol{\beta}}_{+j}$ $(j = 1, \ldots, k-1)$ are the so-called "one-step" estimates given by

$$\hat{\boldsymbol{\beta}}_{+j} + (\mathbf{X}_+^t \hat{\mathbf{W}}_{jj} \mathbf{X}_+)^{-1} \mathbf{X}_+^t (\mathbf{Z}_j - \hat{\boldsymbol{\Pi}}_j),$$

where

$$\hat{\boldsymbol{\beta}}_{+j} = \begin{pmatrix} -\hat{\theta}_j \\ \hat{\boldsymbol{\beta}} \end{pmatrix}$$

and $\hat{\mathbf{W}}_{jj}$ are maximum likelihood estimates under the proportional odds model.

In the event that a nonlogistic link is used, and $g(\pi) = -\theta_j + \boldsymbol{\beta}^t\mathbf{x}$ is assumed, results corresponding to those above can be obtained by replacing $\mathbf{X}_+$ by $\mathbf{U} = \mathbf{H}\mathbf{X}_+$, where $\mathbf{H}$ is diagonal with typical element $\{dg/d\lambda\}^{-1}$, where $\lambda = \text{logit}(\pi)$.