Goodness-of-Fit Tests for the Ordinal Response Models with Misspecified Links

Kwang Mo Jeong[a], Hyun Yung Lee[1,a]

[a]Department of Statistics, Pusan National University

Abstract

The Pearson chi-squared statistic or the deviance statistic is widely used in assessing the goodness-of-fit of the generalized linear models. But these statistics are not proper in the situation of continuous explanatory variables which results in the sparseness of cell frequencies. We propose a goodness-of-fit test statistic for the cumulative logit models with ordinal responses. We consider the grouping of a dataset based on the ordinal scores obtained by fitting the assumed model. We propose the Pearson chi-squared type test statistic, which is obtained from the cross-classified table formed by the subgroups of ordinal scores and the response categories. Because the limiting distribution of the chi-squared type statistic is intractable we suggest the parametric bootstrap testing procedure to approximate the distribution of the proposed test statistic.

Keywords: Ordinal response, proportional odds model, goodness-of-fit, generalized link, ordinal scores, bootstrap.

1. Introduction

To assess the goodness of the assumed model we usually perform the goodness-of-fit(GOF) test in addition to the check of residuals in the generalized linear models. In the binary logistic model including continuous explanatory variables the GOF test by Hosmer and Lemeshow (1980), the so called Hosmer-Lemeshow test, is very popular. The Hosmer-Lemeshow test statistic is a Pearson chi-squared type statistic applied to a $g \times 2$ table formed by subgroups based on the predicted probabilities and the two response categories. The Pearson chi-squared type statistic calculated from this kind of table with random cell boundaries does not have the limiting chi-squared distribution in general. So we encounter the difficulty of determining the critical points of GOF test statistic. We study on the GOF test statistic and the approximation to its distribution in the ordinal response cumulative logit including the binary logistic model as a special case.

The grouping method by Hosmer and Lemeshow (1980) in the binary logistic regression models to be helpful for the ordinal response model. A $g \times 2$ table is formed based on the sorted values of predicted probabilities in such a manner that the number of cases in all subgroups are approximately equal. The Hosmer-Lemeshow statistic is the usual Pearson chi-squared type statistic applied to this $g \times 2$ table.

But for the case of ordinal response model with K response categories the use of predicted probabilities of only a certain response category does not seem to be reasonable. We instead consider a linear combination of the response probabilities $\pi_k(\mathbf{x})$ which is called the ordinal scores by Lipsitz et al. (1996). We partition the whole subjects into g subgroups using the percentiles of these ordinal

---

scores, which results in a $g \times K$ cross-classified table. We routinely compute the chi-squared type statistic from this $g \times K$ table to perform the GOF test for the assumed model.

Chernoff and Lehmann (1954) discussed the limiting distribution of the chi-squared type statistic from the table with random cell boundaries, and later Moore and Spruill (1975) discussed it more generally. Under some regularity conditions the limiting distribution of the chi-squared type statistic from the random table is a weighted sum of independent chi-squared random variables with one degrees of freedom. Because of the complexity of the limiting distribution it is difficult to apply it directly.

It leads us to consider an alternative approach to find the distribution of the proposed test statistic. We suggest a parametric bootstrap testing procedure to perform the GOF test for the ordinal response model.

2. Ordinal Response Models

To formulate the ordinal response model we let $P(Y \leq j|\mathbf{x})$ be the cumulative of response probabilities $\pi_k(\mathbf{x})$'s until the jth category at the covariate value $\mathbf{x}$. That is

$$P(Y \leq j|\mathbf{x}) = \sum_{k=1}^{j} \pi_k(\mathbf{x}), \tag{2.1}$$

where $j = 1, \ldots, K - 1$. The cumulative logit model is defined as

$$\log\left[\frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})}\right] = \alpha_j + \boldsymbol{\beta}_j' \mathbf{x}, \quad j = 1, \ldots, K. \tag{2.2}$$

The left hand side of (2.2) is simply written as $\text{logit}[P(Y \leq j|\mathbf{x})]$. In particular when $\boldsymbol{\beta}_j'$ s are all equal to a common $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, we obtain the so called proportional odds model of the form

$$\text{logit}\left[P(Y \leq j|\mathbf{x})\right] = \alpha_j + \boldsymbol{\beta}' \mathbf{x}. \tag{2.3}$$

We note that the model (2.3) has the same effects $\boldsymbol{\beta}$ for each logit but each cumulative logit has its own intercept α_j satisfying $\alpha_1 \leq \cdots \leq \alpha_{K-1}$. The proportional odds model constrains the $K - 1$ response curves to have the same shape. The probability of response category $\pi_k(\mathbf{x})$ can be recursively obtained from the Equations (2.1) and (2.3), and hence we can find the MLEs of the α_j and $\boldsymbol{\beta}$ based on the likelihood function of multinomial trials.

The overall fit of the model can be significantly improved by using asymmetric links in many applications. A family of generalized link functions which includes somewhat important and commonly used symmetric and asymmetric links has been introduced by Stukel (1988). The most popular links such as logit, probit, and complementary log-log link are approximated by the members of this family.

Hereafter for simplification we denote the linear predictor as

$$\eta_j(\mathbf{x}) = \alpha_j + \boldsymbol{\beta}_j' \mathbf{x}. \tag{2.4}$$

The generalized link model is given in terms of $h_\gamma(\eta_j(\mathbf{x}))$ by

$$\text{logit}\left[P(Y \leq j|\mathbf{x})\right] = h_\gamma(\eta_j(\mathbf{x})), \tag{2.5}$$

where $h_\gamma(t)$ is a non-decreasing function depending on $\gamma = (\gamma_1, \gamma_2)$. The generalized link functions will be explained in detail in Section 4.1.

Table 1: Grouping structure by ordinal scores

Group	1	$\cdots$	k	$\cdots$	K	Total
1	$\hat{O}_{11}$	$\cdots$	$\hat{O}_{1k}$	$\cdots$	$\hat{O}_{1K}$	n_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
j	$\hat{O}_{j1}$	$\cdots$	$\hat{O}_{jk}$	$\cdots$	$\hat{O}_{jK}$	n_j
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
g	$\hat{O}_{g1}$	$\cdots$	$\hat{O}_{gk}$	$\cdots$	$\hat{O}_{gK}$	n_g

3. Goodness-of-Fit Tests

3.1. Grouping by ordinal scores

The ordinal scores can be used as a means of partitioning subjects into subgroups in the ordinal response models. Lipsitz *et al.* (1996) defined the ordinal score $s(\mathbf{x})$ at the covariate value $\mathbf{x}$ by

$$s(\mathbf{x}) = \pi_1(\mathbf{x}) + 2\pi_2(\mathbf{x}) + \cdots + K\pi_K(\mathbf{x}). \tag{3.1}$$

Specially, when $K = 2$ the ordinal score $s(\mathbf{x})$ is equivalent to the binary response probability. The ordinal score $s(\mathbf{x})$ depends on the parameters $\alpha_j's$ and $\boldsymbol{\beta}$ and we simply denote $\hat{s}(\mathbf{x})$ when the estimated parameters are substituted.

The grouping scheme consists of the following steps. We firstly sort the ordinal scores in an increasing (or decreasing) order and nextly partition the whole observations so that approximately the same number of ordinal scores belongs to each group. If we let $[a]$ be the largest integer smaller than or equal to a, the first group consists of approximately $[n/g]$ number of subjects having the ordinal scores from the smallest. The second group is similarly formed by the subjects having the next $[n/g]$ smallest ordinal scores. In this fashion the group boundaries $[c_j, c_j + 1)$, $j = 1, \ldots, g$, are determined with $c_1 = 0$ and $c_{g+1} = \infty$. Hence if a subject with k^{th} response category has ordinal score $s(\mathbf{x})$ such that $c_j \leq s(\mathbf{x}) < c_{j+1}$ then the subject belongs to the (j, k) cell. Thus we can build $g \times K$ cross-classified table with random cell boundaries as shown in Table 1. The cell counts $\hat{O}_{jk}'s$ are obtained by counting the number of subjects in the (j, k) cell.

3.2. Test statistic

In a similar manner to obtain $\hat{O}_{jk}$ we can determine the expected frequency $\hat{E}_{jk}$ by summing the estimated probabilities $\hat{\pi}_k(\mathbf{x}_i)$ for all subjects in the (j, k) cell as follows

$$\hat{E}_{jk} = \sum_{\hat{s}(\mathbf{x}_i) \in [c_j, c_j+1)} \hat{\pi}_k(\mathbf{x}_i), \tag{3.2}$$

where the summation is taken over all the subjects having the ordinal scores $c_j \leq s(\mathbf{x}) < c_{j+1}$ and the kth response category.

As a GOF test on the proportional odds model (2.3) we consider the Pearson chi-squared type statistic of the form, which has been investigated by many researchers in many testing situations including the test of independence,

$$T = \sum_{j=1}^{g} \sum_{k=1}^{K} \frac{\left(\hat{O}_{jk} - \hat{E}_{jk}\right)^2}{\hat{E}_{jk}}. \tag{3.3}$$

As discussed by Moore and Spruill (1975) the limiting distribution of the chi-squared type statistic can be expressed as a weighted sum of independent chi-squared variables with one degrees of freedom. Since this limiting distribution is infeasible except for special cases we need an alternative approach to find the distribution of T. On the other hand Bull (1994) suggested the limiting distribution of T be chi-squared with $g(K-1) - 2$ degrees of freedom.

As an alternative approach we propose the bootstrap based testing procedure which consists of the following steps. Let $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, n$, be a random sample with y_i denoting the ordinal response.

Step 1: Fit the assumed model to find the $\hat{\pi}_k(\mathbf{x}_i)$ and the ordinal scores $\hat{s}(\mathbf{x}_i)$.

Step 2: Calculate the value of T from the $g \times K$ table as given in Table 1.

Step 3: Generate the ordinal responses y_i^* from the multinomial distribution having probabilities $\hat{\pi}_k(\mathbf{x}_i)$ to form a bootstrap sample $(\mathbf{x}_i, y_i^*)$, $i = 1, 2, \ldots, n$.

Step 4: In a similar way as Step 2 calculate the statistic T^* based on the bootstrap sample $(\mathbf{x}_i, y_i^*)$.

Step 5: Step 3 and Step 4 are repeated B times to find the empirical significance probability of the statistic T.

3.3. Other tests

In this section we explain other testing procedures for the GOF of ordinal response models with the grouping methods on which the test statistics depend. Firstly, we comment on the test by Pigeon and Heyse (1999), denoted as χ_{PH}^2, which is in the same vein of the proposed statistic in the sense that it is computed based on the Table 1. The statistic χ_{PH}^2 is of the form

$$\chi_{PH}^2 = \sum_{j=1}^{g} \sum_{k=1}^{K} \frac{\left(\hat{O}_{jk} - \hat{E}_{jk}\right)^2}{\phi_{jk}\hat{E}_{jk}}, \tag{3.4}$$

where $\phi_{jk} = \sum_{i=1}^{n_j} \hat{\pi}_k(\mathbf{x}_i)\{1 - \hat{\pi}_k(\mathbf{x}_i)\}/\{n_j\bar{\pi}_{jk}(1 - \bar{\pi}_{jk})\}$ with $\bar{\pi}_{jk}$ denoting the average of $\hat{\pi}_k(\mathbf{x}_i)$ in the (j,k) cell. Pigeon and Heyse (1999) suggested the limiting distribution of χ_{PH}^2 be chi-squared with $(g-1)(K-1)$ degrees of freedom.

Nextly, before we introduce the test statistic by Pulkstenis and Robinson (2004) we briefly explain their grouping strategy which consists of the following two stages. All subjects are partitioned into subgroups according to the patterns of categorical covariates, and then each subgroup is split into two parts according to the median of ordinal scores within each subgroup. We denote the number of all possible covariate patterns by I to discriminate from the number of groups g mentioned in Section 3.1. The resulting cross-classified table appears as in Table 2. We note that the additional stratification based on the medians of ordinal scores doubles the number of covariate patterns to incorporate information related to all continuous covariates in the model.

The cell count n_{jhk} or the expected count $\hat{E}_{jhk}$ is routinely computed by summing the number of subjects or predicted probabilities within each cell. Pulkstenis and Robinson (2004) suggested two types of test statistics, the one is the chi-squared type statistic defined by

$$\chi_{PR}^2 = \sum_{j=1}^{I} \sum_{h=1}^{2} \sum_{k=1}^{K} \frac{\left(n_{jhk} - \hat{E}_{jhk}\right)^2}{\hat{E}_{jhk}} \tag{3.5}$$

Table 2: Grouping structure by categorical covariate patterns

Group	Subgroup	1	$\cdots$	k	$\cdots$	K
1	1	n_{111}	$\cdots$	n_{11k}	$\cdots$	n_{11K}
	2	n_{121}	$\cdots$	n_{12k}	$\cdots$	n_{12K}
$\vdots$		$\vdots$		$\vdots$		$\vdots$
I	1	n_{I11}	$\cdots$	n_{I1k}	$\cdots$	n_{I1K}
	2	n_{I21}	$\cdots$	n_{I2k}	$\cdots$	n_{I2K}

Table 3: Dataset of mental impairment

Y	X_1	X_2	Y	X_1	X_2	Y	X_1	X_2	Y	X_1	X_2
Well	1	1	Well	1	0	Mild	9	1	Moderate	3	0
Well	9	1	Well	2	0	Mild	3	0	Impaired	8	1
Well	4	1	Mild	5	1	Mild	3	1	Impaired	2	1
Well	3	1	Mild	6	0	Mild	1	1	Impaired	7	1
Well	2	0	Mild	3	1	Moderate	0	0	Impaired	5	0
Well	0	1	Mild	1	0	Moderate	4	1	Impaired	4	0
Well	1	0	Mild	8	1	Moderate	3	0	Impaired	4	0
Well	3	1	Mild	2	1	Moderate	9	0	Impaired	8	1
Well	3	1	Mild	5	0	Moderate	6	1	Impaired	8	0
Well	7	1	Mild	5	1	Moderate	4	0	Impaired	9	0

and the other is the deviance type of the form

$$D_{PR}^2 = 2 \sum_{j=1}^{I} \sum_{h=1}^{2} \sum_{k=1}^{K} n_{jhk} \log \frac{n_{jhk}}{\hat{E}_{jhk}}. \tag{3.6}$$

According to Pulkstenis and Robinson (2004) the limiting distribution of the test statistics χ_{PR}^2 and D_{PR}^2 is chi-squared with $(2I - 1)(K - 1) - q - 1$ degrees of freedom, where q is the number of categorical covariates. As commented by Kuss (2002) the tests by Pulkstenis and Robinson (2004) cannot be applied in the models with no categorical covariates. Furthermore we note that this grouping method incurs the sparseness of the table due to subpartition of groups.

3.4. Practical example

Here we explain the various GOF tests through a practical example having an ordinal response. The dataset given in Table 3 comes from Agresti (2002) on the study of mental health for a random sample of adult residents of Alachua County, Florida. The variable mental impairment is ordinal with four categories; well, mild symptom formation, moderate symptom formation, impaired. We are interested in modelling the mental impairment in terms of two explanatory variables. The one is the life events index, which is a composite measure of the number and severity of important life events that occurred to the subject within the past 3 years. The other is the socioeconomic status of binary measurement high or low. We regard the mental impairment as an ordinal response variable Y, and also we denote the life event and the socioeconomic status as X_1 and X_2, respectively.

We fit the proportional odds model

$$\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 X_1 + \beta_2 X_2, \quad j = 1, 2, 3. \tag{3.7}$$

The tests χ_{PR}^2 and D_{PR}^2 by Pulkstenis and Robinson (2004) have the respective significances 0.530 and 0.290 independently of the value of g. But as we see in Table 4 the significances of the statistic T and the χ_{PH}^2 vary according to the number of groups g.

Table 4: Significance of GOF tests

g	T		χ^2_{PH}
	bootstrap	Bull	
4	0.210	0.490	0.390
5	0.320	0.410	0.330
6	0.240	0.340	0.270

Table 5: Generalized link functions

Link Types	Values of γ_1, γ_2
logit link	$\gamma_1 = 0, \quad \gamma_2 = 0$
complementary log-log link	$\gamma_1 = 0.620, \ \gamma_2 = -0.037$
asymmetric long-short link	$\gamma_1 = -1, \quad \gamma_2 = 1$
long link	$\gamma_1 = -1, \quad \gamma_2 = -1$
short link	$\gamma_1 = 1, \quad \gamma_2 = 1$

All of the tests compared here say that the GOF the assumed model (3.7) seems to be proper because none of the significance is smaller than the usual significance level $\alpha = 0.05$. But we are confronted with the determination of g. The rule-of-thumb in the chi-squared type statistics requires $\hat{E}_{jk}$ to be greater than or equal to 5. According to this rule we can roughly approximate g to satisfy $g \leq n/(5K)$ when the sample size is moderately large. In this example from $n = 40$ and $K = 4$ we find $g = 2$, very small number, which results in the decrease of powers of GOF tests. It would be desirable to consider the tradeoff between the magnitude of expected counts and the number of groups according to the values of n and K. Conclusively we prefer the value of $g = 4$ in this example.

4. A Monte Carlo Study

4.1. Simulation design

We consider generalized links to compare empirical powers of various GOF tests discussed in this paper. The generalized link function $h_\gamma(\eta_j(\mathbf{x}))$ is defined in two ways according as $\eta_j(\mathbf{x}) \geq 0$ or not. When $\eta_j(\mathbf{x}) \geq 0$ the $h_\gamma(\eta_j(\mathbf{x}))$ is of the form

$$h_\gamma\left(\eta_j(\mathbf{x})\right) = \begin{cases} \gamma_1^{-1}\left\{\exp(\gamma_1\eta_j(\mathbf{x})) - 1\right\}, & \gamma_1 > 0, \\ \eta_j(\mathbf{x}), & \gamma_1 = 0, \\ -\gamma_1^{-1}\log\left(1 - \gamma_1\eta_j(\mathbf{x})\right), & \gamma_1 < 0. \end{cases} \qquad (4.1)$$

On the other hand when $\eta_j(\mathbf{x}) < 0$, the function $h_\gamma(\eta_j(\mathbf{x}))$ is defined by

$$h_\gamma\left(\eta_j(\mathbf{x})\right) = \begin{cases} -\gamma_2^{-1}\exp\left(-\gamma_2\eta_j(\mathbf{x})\right) - 1, & \gamma_2 > 0, \\ \eta_j(\mathbf{x}), & \gamma_2 = 0, \\ \gamma_2^{-1}\log\left(1 + \gamma_2\eta_j(\mathbf{x})\right), & \gamma_2 < 0. \end{cases} \qquad (4.2)$$

The function $h_\gamma(\eta_j(\mathbf{x}))$ is strictly increasing and nonlinear. The values of γ_1 and γ_2 controls the shape of the lower and upper tails of the link function. Several links listed in Table 5 are used in this simulation study to compare the empirical powers of the GOF tests discussed in this paper for the ordinal response models with misspecified links.

To assess the powers of detecting misspecified links we fit the proportional odds model

$$H_0 : \text{logit}[P(Y \leq j|\mathbf{x})] = \eta_j(\mathbf{x}) \qquad (4.3)$$

against the generalized link model

$$H_1 : \text{logit}[P(Y \leq j|\mathbf{x})] = h_\gamma(\eta_j(\mathbf{x})). \tag{4.4}$$

Firstly for the null model (4.3) we assume the linear predictor

$$\eta_j(\mathbf{x}) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z.$$

Here the variables X_i, $i = 1, 2$, are generated from the Bernoulli distribution with probability $1/2$, and Z is from the uniform distribution over $(1, 3)$. We simply let the covariate values as $\mathbf{x} = (x_1, x_2, z)$. For example, when $X_1 = 0$, $X_2 = 0$, $Z = 1$ we have $\mathbf{x} = (0, 0, 1)$. The intercept α_j, and the coefficients of X_i and Z are determined in the following way. We arbitrarily fix $\alpha_1 = -1.25$ and $\beta_2 = -1.0$, and then simultaneously solve the equations

$$P[Y = 1|\mathbf{x} = (0, 0, 1)] = 0.27, \qquad P[Y \leq 2|\mathbf{x} = (1, 0, 3)] = 0.56$$

to find other coefficients. In this way we determined the linear predictor $\eta_j(\mathbf{x})$ of (4.1) as

$$M_0 : \eta_j(\mathbf{x}) = \alpha_j + 0.75X_1 - 1.0X_2 + 0.25Z; \quad \alpha_1 = -1.25, \ \alpha_2 = 0.25.$$

When we fit the proportional odds model (4.1) to the pseudo dataset generated from the model itself the GOF tests are required to satisfy the significance level α in the respect of Type I error rate.

Furthermore to compare the powers of various GOF tests for the pseudo dataset generated from the model (4.2) we fit the proportional odds model (4.1). The coefficients of $\eta_j(\mathbf{x})$ in the model (4.2) are determined in a similar manner discussed before. The first of the generalized link models we are to consider is the complementary log-log model which have $\gamma_1 = 0.620$ and $\gamma_2 = -0.037$. In this model we take the linear predictor as

$$M_1 : \eta_j(\mathbf{x}) = \alpha_j - 0.67X_1 + 0.69Z; \quad \alpha_1 = -0.5, \ \alpha_2 = 1.0.$$

The categorical variable X_1 is assumed as before to be Bernoulli with probability $1/2$ and Z is uniform over $(-3, 3)$. For the case of asymmetric long-short link having $\gamma_1 = -1$ and $\gamma_2 = 1$, we take $\eta_j(\mathbf{x})$ to be

$$M_2 : \eta_j(\mathbf{x}) = \alpha_j - 0.15X_1 + 2.0Z; \quad \alpha_1 = -1.5, \ \alpha_2 = 1.5.$$

Similarly for the long-link with $\gamma_1 = -1$ and $\gamma_2 = -1$, the model M_3, the $\eta_j(\mathbf{x})$ is of the form

$$M_3 : \eta_j(\mathbf{x}) = \alpha_j - 7.0X_1 + 2.0Z; \quad \alpha_1 = 3.0, \ \alpha_2 = 10.0.$$

Lastly, for the short-link model having $\gamma_1 = 1$ and $\gamma_2 = 1$, the model M_4, we take $\eta_j(\mathbf{x})$ as

$$M_4 : \eta_j(\mathbf{x}) = \alpha_j - 1.5X_1 + 1.5Z; \quad \alpha_1 = -1.5, \ \alpha_2 = 1.5.$$

4.2. Simulation results

In this section we tabulate the empirical powers for the various GOF tests obtained from 200 iterations and 300 bootstrap replications. We will discuss the performance of the proposed test T compared to other tests discussed in this paper. We also include the usual Pearson chi-squared statistic χ^2 and the deviance D^2 for comparison in the last two columns of Table 6 and Table 7. As we see in Table 6

Table 6: Empirical significance levels for the true model

Model	α	N	T		χ^2_{PH}	χ^2_{PR}	D^2_{PR}	χ^2	D^2
			bootstrap	Bull					
M_0	0.05	100	0.030	0.005	0.005	0.060	0.115	0.055	0.065
		150	0.055	0.025	0.025	0.050	0.070	0.060	0.065
		200	0.060	0.030	0.035	0.095	0.115	0.080	0.085
	0.10	100	0.075	0.030	0.030	0.135	0.190	0.105	0.135
		150	0.110	0.065	0.065	0.160	0.185	0.125	0.125
		200	0.115	0.065	0.065	0.150	0.165	0.110	0.115

Table 7: Empirical powers of the misspecified models

Model	α	N	T		χ^2_{PH}	χ^2_{PR}	D^2_{PR}	χ^2	D^2
			bootstrap	Bull					
M_1	0.05	100	0.090	0.040	0.060	0.060	0.095	0.070	0.080
		150	0.080	0.065	0.065	0.075	0.070	0.045	0.045
		200	0.105	0.055	0.055	0.075	0.075	0.050	0.050
	0.10	100	0.190	0.110	0.110	0.150	0.205	0.150	0.155
		150	0.135	0.100	0.100	0.140	0.145	0.105	0.110
		200	0.185	0.130	0.130	0.145	0.150	0.120	0.125
M_2	0.05	100	0.510	0.510	0.520	0.155	0.180	0.025	0.030
		150	0.730	0.785	0.795	0.265	0.360	0.025	0.025
		200	0.845	0.885	0.890	0.395	0.525	0.035	0.0300
	0.10	100	0.655	0.570	0.585	0.245	0.335	0.075	0.075
		150	0.880	0.865	0.870	0.375	0.540	0.075	0.080
		200	0.915	0.915	0.915	0.520	0.650	0.060	0.060
M_3	0.05	100	0.455	0.410	0.410	0.395	0.310	0.110	0.100
		150	0.650	0.615	0.615	0.530	0.415	0.195	0.180
		200	0.690	0.670	0.680	0.640	0.550	0.275	0.250
	0.10	100	0.555	0.495	0.495	0.475	0.395	0.215	0.180
		150	0.755	0.715	0.715	0.610	0.520	0.285	0.265
		200	0.810	0.780	0.780	0.735	0.670	0.375	0.360
M_4	0.05	100	0.610	0.575	0.590	0.260	0.465	0.090	0.090
		150	0.790	0.800	0.805	0.450	0.635	0.155	0.155
		200	0.885	0.900	0.905	0.610	0.835	0.175	0.175
	0.10	100	0.750	0.695	0.715	0.410	0.595	0.145	0.150
		150	0.895	0.870	0.885	0.565	0.790	0.240	0.240
		200	0.930	0.935	0.935	0.760	0.920	0.310	0.310

for the true model M_0 the proposed tests have approximately correct the Type I errors that range from 0.03 to 0.06 at the nominal significance level $\alpha = 0.05$. The tendency is similar at $\alpha = 0.10$. On the other hand the Type I error of χ^2_{PR} is 0.095 at $\alpha = 0.05$, and also 0.160 at $\alpha = 0.10$. Both χ^2_B and χ^2_{PH} are seem to be conservative.

The empirical powers for the misspecified link models M_1 through M_4 are given in Table 6. In the complementary log-log link model M_1 all tests have lower powers than the other models of misspecified links. But for the asymmetric long-short model M_2, our proposed test attains its power as high as 91.5% at $\alpha = 0.10$. We can see similar tendency in other links models of M_1 and M_4.

5. Conclusion and Further Research

The Hosmer-Lemeshow GOF test for the binary logistic model has been extended to the ordinal response cumulative logit model. We suggest a grouping strategy using the ordinal scores to form a cross-classified table from which the chi-squared type GOF test statistic can be computed. As a special case for the binary logistic model this method coincides with the Hosmer-Lemeshow GOF

test. As have been studied by many researchers the chi-squared type statistic on the table with random boundaries does not have the limiting chi-squared distribution but is a weighted sum of independent chi-squared random variables with one degree of freedom.

This is a difficulty in performing the GOF test because we cannot easily find the significance of the test statistic. A bootstrap based testing procedure is a good alternative to the limiting distribution of GOF test. We explained the proposed GOF test through a practical example. A Monte Carlo study has been performed to compare various GOF tests for ordinal response models with misspecified links.

Extensive study on the limiting distribution of the proposed test statistic and also on the consistency of the parametric bootstrapped test statistic is remained as a further research.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., Wiley, New York.

Bull, S. (1994). *Analysis of Attitudes toward Workplace Smoking Restrictions*, In: N. Lange, L. Ryan, D. Billard, L. Conquest, and J. Greenhouse (eds.): Case Studies in Biometry. Wiley, New York, 249–271.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimation in Chi-square test of goodness of fit, *The Annals of Mathematical Statistics*, **25**, 579–586.

Hosmer, D. W. and Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model, *Communications in Statistics, Theory and Methods*, **9**, 1043–1069.

Kuss, O. (2002), Global goodness-of-fit tests in logistic regression with sparse data, *Statistics in Medicine*, **21**, 3789–3801.

Lipsitz, S. R., Fitzmaurice, G. M. and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models, *Applied Statistics*, **45**, 175–190.

Moore, D. S. and Spruill, M. C. (1975). Unified large sample theory of general Chi-squared statistics for tests of fit, *The Annals of Statistics*, **3**, 599–616.

Pigeon, J. G. and Heyse, J. F. (1999). An improved goodness-of-fit statistic for probability prediction models, *Biometrical Journal*, **41**, 71–82.

Pulkstenis, E. and Robinson, T. J. (2004). Goodness-of-fit tests for ordinal response regression models, *Statistics in Medicine*, **23**, 999–1014.

Stukel, T. A. (1988). Generalized logistic models, *Journal of the American Statistical Association*, **83**, 426–431.