Semiparametric Prediction with Retrospective Logit Models

Cheng Peng

Department of Mathematics, West Chester University
25 University Avenue, West Chester, PA 19383, USA
cpeng@wcupa.edu

October 7, 2023

Abstract

We unify the four commonly used polychotomous logistic regression (PLR) models, namely, baseline-logit, adjacent-logit, cumulative logit and continuation ratio, and define a super structure for polychotomous logistic regression models under case control data. Zhang's (2000) information matrix test of goodness-of-fit for retrospective binary logistic regression models has been extended to assess the goodness-of-fit of the proposed polychotomous logistic models. Some numerical and simulation will also be presented.

AMS 2000 subject classifications. Primary 62F12; 62J12; 62P10

Keywords polychotomous logistic model, case-control study, semiparametric model, goodness-of-fit, information matrix test, noncentral chi-square.

1. Introduction

The binary logistic regression model has been extremely attractive and played an important role in many practical areas such as biomedical, social sciences, political science, etc. It model the log odds of observing the outcome of interest from the possible outcome adjusted (controlled) by some risk factors. Let Y be the response variable taking either value 2 (disease, success, etc.) or value 1 (non-disease, failure, etc) and $\mathbf{X}$ be the p-dimensional vector of risk factors affecting the probabilities of observing $Y = 2$. Let α and $\boldsymbol{\beta}$ be the regression coefficients. The p-dimensional vector $\boldsymbol{\beta}$ usually called the log-odds ratio parameter associated with the p-risk factors. The binary logistic regression model is defined as

$$\log \frac{P(Y=2|\mathbf{X}=\mathbf{x})}{1-P(Y=1|\mathbf{X}=\mathbf{x})} = \alpha + \boldsymbol{\beta}^T \mathbf{x} \tag{1·1}$$

Let $\pi_i(\mathbf{x}) = P(Y = i|\mathbf{X} = \mathbf{x})$ for $i = 1, 2$. Clearly $\pi_0(\mathbf{x}) + \pi_1(\mathbf{x}) = 1$. Using the notation of cell probability $\pi_i(\mathbf{x})$, we can re-express (1·1) as

$$\log \frac{\pi_2(\mathbf{x})}{1-\pi_2(\mathbf{x})} = \log \frac{\pi_2(\mathbf{x})}{\pi_1(\mathbf{x})} = \alpha + \boldsymbol{\beta}^T \mathbf{x} \tag{1·2}$$

That is, the binary logistic regression model links the log odds of observing $Y = 1$ (disease) with a linear combination risk factors $\alpha + \boldsymbol{\beta}^T\mathbf{x}$ and α and $\boldsymbol{\beta}$ are naturally called odds ratio parameters. $\boldsymbol{\beta}$ reflect the contribution of associated risk factors to the odds ratio. Note that if we link the log of odds of observing $Y = 0$ with a linear predictor, then we have

$$\log \frac{\pi_1(\mathbf{x})}{1-\pi_2(\mathbf{x})} = \log \frac{\pi_1(\mathbf{x})}{\pi_2(\mathbf{x})} = \alpha^* + \boldsymbol{\beta}^{*T} \mathbf{x} \tag{1·3}$$

Since $\alpha^* = -\alpha$ and $\boldsymbol{\beta}^* = -\boldsymbol{\beta}^*$, model (1·1) and model (1·3) are equivalent.

In many practical situations, there are more than two categories for the response variable Y. For instance, in epidemiologic study, the variable of interest may be the disease severity taking values non-disease (1), mild (2), moderate (3), and severe (4). We can use similar approach to link the odds of observing certain disease severity with a linear predictor. There are several ways to define logits according the ordinality of the response variable Y. For the nominal response Y with J categories, we can select any category, say the category J, as the reference and define log odds (restricted to the two selected categories) similar to (1·1).

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x}, \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·4}$$

which is usually called *generalized logit model or baseline logit model*. Since Y has J categories, we need $J - 1$ log odds to fully specify the model. When $J = 2$, (1·4) reduces to (1·1). This model has been studied and applied to different subject areas, to name a few, by McCullagh (1980), Anderson (1984), Begg and Gray (1984), Fienberg (1978), and among others.

For ordinal variable Y, three models with practical meaning are defined by incorporating the ordinal information. The first of these models is called *adjacent logit model* having the following definition

$$\log \frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x}, \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·5}$$

The adjacent logit model uses the ordinal information in Y, but it can be, in fact, expressed into a baseline logit model (Agresti 2002). In other words, we can fit generalized logit model to a data and then use relationship between the regression coefficients of the two models to get fitted adjacent logit model, and vice versa. Since Y is ordinal, it is meaningful to define cumulative probability $\gamma_j(\mathbf{x}) = P(Y \leqslant j) = \sum_{k=1}^{j} \pi(\mathbf{x})$. The next ordinal logit model is defined based on $\gamma_j(\mathbf{x})$ with the following form

$$\log \frac{\gamma_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x}, \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·6}$$

where $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_{J-1}$ which guarantees that the fitted probabilities are non-negative. This model link the log of odds of observing response Y beeing lower than the j-th category with a linear predictor. It is essentially a binary logistic model for each specific choice of j. Since it is defined based on cumulative response probability, it is naturally called *cumulative logit model* in literature. McCullagh (1980) pointed this model to be a *proportional odds* model when the log-odds ratio parameters are assumed to be constant across all $J-1$ logits ($\boldsymbol{\beta}_j = \boldsymbol{\beta}$) since

$$\frac{\gamma_j(\mathbf{x}_1)/(1-\gamma_j(\mathbf{x}_1))}{\gamma_j(\mathbf{x}_2)/(1-\gamma_j(\mathbf{x}_2))} = \exp(-\boldsymbol{\beta}^T(\mathbf{x}_1 - \mathbf{x}_2)), \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·7}$$

Note that if Y represents the discrete the survival time, $(1-\gamma_j(\mathbf{x}_1))$ becomes survival function. Therefore, model (1·7) can be used to model discrete survival time. Recently, Marshall and Olkin (1997) used the idea of the definition of proportional odds model to proposed a general methods to expanding existing parametric families. The last model we will cover in this paper is usually called *continuation ratio logistic regression* in literature, see for example, Fienberg (1978), McCullagh (1980), Agresti (2002), and among others. The definiton for the model is given by

$$\log \frac{\pi_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x}, \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·8}$$

or

$$\log \frac{\pi_{j+1}(\mathbf{x})}{\gamma_j(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x}, \quad \text{for} \quad j = 1, 2, \cdots, J-1\cdot \tag{1·9}$$

Unlike ordinal PLR models (1·5) and (1·6) having *palindromic invariant* property, the continuation ratio logistic model does not possesses this property. Reversing the response Y in continuation ratio model yields inequivalent models. Therefore, (1·8) and (1·9) are not equivalent even though the inferential procedures for them are the same. Because of the irreversibility of the response Y in the continuation ratio model, it has a special attraction to the cases in which the response has natural special hierarchical structure. In the rest of the discussion, we only focus on continuation ratio logistic model (1·8). The continuation ratio logistic model has been receiving considerably attention from researchers in different areas, such as Hemker et al. (2001), Tutz (1991), Kvist and Thyregod (2000), Rindorf and Lewy (2001) and Cox (1988).

Since these polychotomous logistic regression models consist of a sequence of binary logits according to the ways of using the ordinality of the response, fitting these models to prospective datasets is similar to fitting a binary logistic regression model using any general purpose statistical packages. Pearson and deviance χ^2 are the standard tests for goodness-of-fit. However, in many situations, it is either impractical or impossible to collect data prospectively. For instance, in epidemiologic study, if the disease under investigation is rare (one out one million)or has a extremely long latency (30 years), take data prospectively will end up with either very few cases or 30 years of waiting time for the disease development. Farewell (1979) and Prentice and Pyke (1979) systematically investigated the ways of fitting the logistic regression model to retrospective data and concluded that it is valid to fit the logistic regression model to the retrospective data as if it were collected prospectively, but the inference on the intercept parameter is not possible unless the sampling fractions are given. Scott and Wild (1986) pointed out that such approach is sensitive to model misspecification.

Using empirical likelihood method (Owen 1988, 1990), Qin and Zhang (1997) proposed a profiled semiparametric empirical likelihood method combining estimation and a Kolmogorov-Smirnov type goodness-of-fit test for the retrospective loistic regression model. In the same direction, Zhang (2002, 2004) and Peng and Zhang (2008) studied generalized logit, proportional odds and continuation ratio logistic regression based on case-control data respectively. Since there is no analytical expression for proposed Kolmogorov-Smirnov test statistic, a Bootstrap procedure was used to establish the decision rule.

In this paper, we first define a unified structure of the aforementioned four retrospective polychotomous logistic regression models based on the work of Zhang (2002, 2004) and Peng and Zhang (2008) and then extend Zhang (2001) information matrix test of goodness of fit to the unified structure of retrospective polychotomous logistic regression models. In Section 2, we propose a unified structure of retrospective polychotomous logistic regression models and the asymptotic results of the semipaarametric empirical likelihood estimators. In Section 3, we establish the information matrix equality. Construction of information matrix test statistic (χ^2) will be presented in Section 4. Some numerical results based on real-life data and power analysis based on a local alternative via simulation study are given in Section 5, Section 6 includes summary and concluding remarks. Technical details are given in the appendix.

2. Unified Structure of Retrospective PLR and Summary of Existing Asymptotic Results

In this section, propose a unified structure for retrospective PLR models. Since the the adjacent logit and the generalized logit models are equivalent in the sense that one can be converted to the other since the regression coefficients in both models have a one-to-one linear relationship (see, for example, Agresti, 2002, pages 286-287). Throughout this paper PLR means generalized logit, proportional odds and continuation ratio logistic regression models.

2·1 The Unified Structure of PLR

Let $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}$ be the random sample collected from the i-th population (i-th category of Y) for $i = 1, \cdots, I$. Assume further that these samples are jointly independent. Denote $P(Y = i) = \tau_i$, the population proportion of i-th category, for $i = 1, \cdots, I$. According to Bayes Theorem, we have

$$P(X|Y = i) = \frac{P(Y = i|X) \cdot P(X)}{P(Y = i)} = \frac{\pi_i(X)P(X)}{P(Y = i)}. \tag{2·1}$$

Let $f_i(x)$ and $f(x)$ be the density functions of X given that $Y = i$ and X respectively. For $i = 1, \cdots, I - 1$, we define $\omega_i = \log[P(Y = i)/P(Y = I)]$ to be the sampling fraction based on i-th and I-th subpopulations. Let $\theta_i = (\alpha_i, \beta_i^T)^T$, $\theta = (\theta_1^T, \cdots, \theta_{I-1}^T)^T$, $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_{I-1})^T$. Re-expressing (1·4), (1·6) and (1·8) in terms of $\pi_i(x)$ and solving for $\pi_i(x)$ by using the fact that $\sum_{i=1}^{I} \pi_i(x) = 1$, and plugging $\pi_i(x)$ in (2·1), we get the following I-sample semiparametric model (see Zhang, 1999, 2001 and Peng & Zhang 2008 for details),

$$\begin{cases} X_{I1}, \cdots, X_{In_I} \overset{\text{i·i·d·}}{\sim} f_I(x), \\ X_{i1}, \cdots, X_{in_i} \overset{\text{i·i·d·}}{\sim} \exp\big(\gamma_i + g_i(x, \theta)\big) \cdot f_I(x), \quad \text{for} \quad i = 1, \cdots, I - 1, \end{cases} \tag{2·2}$$

where, for the generalized logit model,

$$g_i(x) = \beta_i^T x \tag{2·3}$$

and $\gamma_i = \alpha_i + \omega_i$; for the proportional odds model,

$$g_i(x) = \log\Big(\frac{s_i(x,\alpha,\beta) - s_{i-1}(x,\alpha,\beta)}{s_I(x,\alpha,\beta) - s_{I-1}(x,\alpha,\beta)}\Big) \tag{2·4}$$

with $s_i(x,\alpha,\beta) = \exp(\alpha_i + \beta^T x)/[1+\exp(\alpha_i + \beta^T x)]$ and $\gamma_i = \omega_i$; for the continuation model,

$$g_i(x,\theta) = \begin{cases} \beta_1^T x + \sum_{l=2}^{I-1} \log[1+\exp(\alpha_l + \beta^T x)], & i = 1, \\ \alpha_i + \beta_i^T x + \sum_{l=i+1}^{I-1} \log[1+\exp(\alpha_l + \beta^T x)] & i = 2, \cdots, I-2, \\ \alpha_{I-1} + \beta_{I-1}^T x, & \text{otherwise}\cdot \end{cases} \tag{2·5}$$

and $\gamma_1 = \alpha_1 + \omega_1$ and $\gamma_i = \omega_i$ for $i > 1$.

Remark 2.1. It is customarily assumed the equal odds ratio parameter (across all logits) in the proportional odds model. That is $\beta = \beta_1^T = \beta_2^T = \cdots = \beta_{I-1}^T$ (McCullagh 1980, Zhang 2004).

Remark 2.2. It is seen that, all intercepts in the retrospective generalized logit model are inestimable ($\gamma_i = \alpha_i + \omega_i$ for $i = 1, \cdots, I-1$); only α_1 in retrospective continuation ratio models is inestimable ($\gamma_1 = \alpha_1 + \omega_1$), all intercepts in the retrospective proportional odds models are estimable ($\gamma_i = \omega_i$ for $i = 1, \cdots, I-1$).

Remark 2.3. The parameters in the vector θ vary depending on the correct model. In retrospective generalized logit model, $\theta_i = \beta_i$ for all $i = 1, 2, \cdots, I-1$; in retrospective proportional odds model, $\theta_i = (\alpha_i, \beta_i^T)^T$ for all $i = 1, 2, \cdots, I-1$; in retrospective continuation ratio model, $\theta_1 = \beta_1$ and $\theta_i = (\alpha_i, \beta_i^T)^T$ for all $i = 2, \cdots, I-1$.

2·2 Model Estimation and Some Asymptotic Results

For $i = 1, \cdots, I$, let $F_i(x)$ be the corresponding cumulative distribution of $f_i(x)$, $\{T_1, \cdots, T_n\}$ be the pooled sample $\{X_{11}, \cdots, X_{1n_1}; \cdots; X_{I1}, \cdots, X_{In_I}\}$ with $n = \sum_{i=1}^I n_i$, Then from model (2·2) , we have following likelihood function

$$\begin{aligned} L(\gamma,\theta,F_I|\mathbf{X}) &= \Big[\prod_{i=1}^{I-1}\prod_{j=1}^{n_i} \exp\Big(\gamma_i + g_i(X_{ij};\theta)\Big) dF_I(X_{ij})\Big]\Big[\prod_{j=1}^{n_I} dF_I(X_{Ij})\Big] \\ &= \prod_{s=1}^{n} p_s \prod_{i=1}^{I-1}\prod_{j=1}^{n_i} \Big\{\exp\Big[\gamma_i + g_i(X_{ij};\theta)\Big]\Big\}, \end{aligned} \tag{2·6}$$

where $p_s = dF_I(t_s)$ is the empirical likelihood of $F_I(X)$ at point t_s with $\sum_{s=1}^{n} p_s = 1$, and $g_i(x;\theta)$ is specified respectively in (2·3), (2·4) and (2·5) for $i = 1, 2, \cdots, I$. The corresponding log-likelihood function is

$$l(\gamma, \theta, F_I) = \sum_{s=1}^{n} \log p_s + \sum_{i=1}^{I-1} n_i \gamma_i + \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} g_i(X_{ij}, \theta) \tag{2·7}$$

The estimates of parameters θ, γ_i and the distribution $F_I(x)$ corresponding to the density function $f_I(x)$ will be obtained by maximizing $l(\gamma, \theta, F_I)$ subject to the following constraints, for $s = 1, 2, \cdots, n, i = 1, 2, \cdots, I-1$,

1. $p_s \geqslant 0$, 2. $\sum_{s=1}^{n} p_s = 1$, 3. $\sum_{s=1}^{n} p_s \big\{ \exp[\gamma_i + g_i(T_s;\theta)] - 1 \big\} = 0$. We first use *Lagrange Multipliers* to maximize the log-likelihood function (2·7) with given constraints by fixing parameters first and get profile MLE of p_s

$$\hat{p}_s = \frac{1}{n_I} \cdot \frac{1}{1 + \sum_{i=1}^{I-1} \rho_i \exp[\gamma_i + g_i(T_s;\theta)]}, \tag{2·8}$$

where $\rho_i = n_i/n_I$ for $i = 1, 2, \cdots, I-1$. One can see that $\hat{p}_s$ is a function of data values and unknown parameters γ and θ. In (2·7) we substitute $\hat{p}_s$ for p_s and get the following semiparametric profile empirical log likelihood function

$$l(\gamma, \theta) = -n \log n_I - \sum_{s=1}^{n} \log \left\{ 1 + \sum_{i=1}^{I-1} \rho_i \exp[\gamma_i + g_i(T_s;\theta)] \right\} + \sum_{i=1}^{I-1} \sum_{j=1}^{n_i} \Big[g_i(X_{ij}, \theta) + \gamma_i \Big] \cdot$$

The corresponding score equations are given by, for $u = 1, 2, 3, \cdots, I-1$,

$$n_u - \sum_{s=1}^{n} \frac{\rho_u \exp[\gamma_u + g_u(T_s;\theta)]}{1 + \sum_{m=1}^{I-1} \rho_m \exp[\gamma_m + g_m(T_s, \theta)]} = 0, \tag{2·9}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{n_i} \Big[\frac{\partial g_i(X_{ij};\theta)}{\partial \theta} - \frac{\sum_{m=1}^{I-1} \rho_m \exp[\gamma_m + g_m(X_{ij};\theta)] \cdot \partial g_m(X_{ij};\theta)/\partial\theta}{1 + \sum_{m=1}^{I-1} \rho_m \exp[\gamma_m + g_m(X_{ij};\theta)]} \Big] = 0 \cdot \tag{2·10}$$

The semiparametric MLE of (γ, θ), denoted by $(\tilde{\gamma}, \tilde{\theta})$, is a solution to the above system of score equations (2·9) and (2·10). Hence, the semiparametric empirical likelihood estimate of p is given by

$$\tilde{p}_s = \frac{1}{n_I} \cdot \frac{1}{1 + \sum_{i=1}^{I-1} \rho_i \exp[\tilde{\gamma}_i + g_i(T_s;\tilde{\theta})]}, \tag{2·11}$$

The corresponding semiparametric estimate of $F_i(x)$ under retrospective polychotomous logistic regression model (2·2) are, for $i = 1, 2, \cdots, I-1$,

$$\tilde{F}_I(t) = \sum_{s=1}^{n} \tilde{p} I_{[T_s \leqslant t]}, \quad \tilde{F}_i(t) = \sum_{s=1}^{n} \tilde{p} \exp(\tilde{\gamma}_i + g_i(T_s : \tilde{\theta})) I_{[T_s \leqslant t]} \cdot \tag{2·12}$$

We assume that n_i/n approaches to a constant when $n = \sum_{i=1}^{I} n_i \to \infty$, for $i = 1, \cdots, I$. Denote $\rho_u = \lim_{n\to\infty} n_u/n_I, \rho = \lim_{n\to\infty} \sum_{i=1}^{I-1} n_i/n_I, \gamma_I = 0, g_I(x) = 0$. Let (γ_0, θ_0) be the true value of (γ, θ). Furthermore, we define, for $u, v = 1, 2, \cdots, I$,

$$e_u(x) = \rho_u \exp[\gamma_u + g_u(x,\theta)]\Big|_{\gamma=\gamma_0,\theta=\theta_0}; \tag{2·13}$$

$$g_u^{\theta^T}(x) = \frac{\partial g_u(x,\theta)}{\partial \theta^T}\Big|_{\theta=\theta_0}; \quad g_u^{\theta}(x) = \frac{\partial g_u(x,\theta)}{\partial \theta}\Big|_{\theta=\theta_0}; \quad g_u^{\theta\theta^T}(x) = \frac{\partial^2 g_u(x,\theta)}{\partial \theta \partial \theta^T}\Big|_{\theta=\theta_0}; \tag{2·14}$$

$$e_{uv}(x) = \rho_u \exp[\gamma_u + g_u(x,\theta)]\rho_v \exp[\gamma_v + g_v(x,\theta)]|_{\gamma=\gamma_0,\theta=\theta_0}, \quad u \neq v;$$

$$e_{uu}(x) = -\sum_{u\neq v, v=1}^{I} e_{uv}(x), \quad \text{for} \quad u = 1, 2, \cdots, I-1;$$

$$P(x) = \{1 + \sum_{m=1}^{I-1} \rho_m \exp[\gamma_m + g_m(x,\theta)]\}^{-1}\big|_{\gamma=\gamma_0,\theta=\theta_0}; \tag{2·15}$$

$$a_{uv} = -\frac{1}{1+\rho}\int P(x)e_{uv}(x)dF_I(x); \quad \text{for} \quad u \neq v;$$

$$a_{uu} = -\sum_{u\neq v, v=1}^{I} a_{uv}, \quad \text{for} \quad u = 1, 2, \cdots, I-1;$$

Consequently,

$$a_{uI} = -\frac{1}{1+\rho}\int P(x)e_u(x)dG_I(x) = \sum_{v=1}^{I-1} a_{uv} \quad \text{for} \quad u = 1, 2, \cdots, I-1\cdot$$

$$S_1 = \Big(a_{uv}\Big)_{u,v=1,2,\cdots,I-1}, \quad S_2 = \big((\boldsymbol{a}_1^{(\theta)})^T, (\boldsymbol{a}_2^{(\theta)})^T, \cdots, (\boldsymbol{a}_{I-1}^{(\theta)})^T\big)^T \tag{2·16}$$

where

$$\boldsymbol{a}_u^{(\theta)} = \frac{1}{1+\rho}\int e_u(x)\Big(g_u^{\theta^T}(x) - P(x)\sum_{m=1}^{I-1} e_m(x) g_m^{\theta^T}(x)\Big)dF_I(x)$$

$$S_3 = \frac{1}{1+\rho}\int \Big(\sum_{m=1}^{I-1} g_m^{\theta}(x) g_m^{\theta^T}(x) e_m(x) - P(x)\sum_{m=1}^{I-1} e_m(x) g_m^{\theta}(x) \sum_{m=1}^{I-1} e_m(x) g_m^{\theta^T}(x)\Big)dF_I(x)\cdot \tag{2·17}$$

$$D = \operatorname{diag}\big(\tfrac{1}{\rho_1}, \cdots, \tfrac{1}{\rho_I}\big), \quad S = \tfrac{1}{1+\rho}\begin{pmatrix} S_1 & S_2 \\ S_2^T & S_3 \end{pmatrix}, \quad \Omega = \begin{pmatrix} D+U & O_1 \\ O_1 & O_2 \end{pmatrix}\cdot \tag{2·18}$$

where U is the $(I-1)\times(I-1)$ matrix with all elements being 1, O_1 is an $\big[(I-1)(p+1)-1\big]\times(I-1)$ zero matrix, O_2 is an $\big[(I-1)(p+1)-1\big]\times\big[(I-1)(p+1)-1\big]$ zero matrix, p is the number of parameters in the corresponding retrospective PLR model. With the above notations, we state the asymptotic normality of the estimators of the regression coefficients as follows

Theorem 1. *(Zhang 2002, 2004, Peng and Zhang 2008) Under retrospective polychotomous logistic regression model (2·2), we have*

$$n^{-1/2}\left(\frac{\partial l(\gamma_0,\theta_0))}{\partial\gamma^T},\frac{\partial l(\gamma_0,\theta_0)}{\partial\theta^T}\right)^T \to_P N(0,V)$$

consequently,

$$\sqrt{n}\left(\tilde{\gamma}^T-\gamma_0^T \quad \tilde{\theta}^T-\theta_0^T\right)^T \to_d N(0,\Gamma)$$

where $V=S-(1+\rho)(S_1^T,S_2^T)(D+U)(S_1^T,S_2^T)^T$ *and* $\Gamma=S^{-1}-(1+\rho)\Omega$. S_1,S_2,S *and* Ω *are specified in (2·16) and (2·18).*

3. Information Matrix Equality

In this section, we proposed a information matrix based generalized moment χ^2 test to assess the fit of the retrospective PLR model (2·2). White (1982) proposed a parametric information test using the fact that, under standard regularity conditions, the expectation of the score derivation matrix $U_n(\xi)=-n^{-1}\{\partial^2 l(\xi)/\partial\xi\partial\xi^T\}$ and the score squared matrix $V_n(\xi)=n^{-1}\sum_{i=1}^n\{\partial l(\xi)/\partial\xi\}\{\partial l(\xi)/\partial\xi^T\}$ are equal to the the information matrix if the model under study is correctly specified (ξ is the vector of model parameters). Hausman (1978), Hausman and McFadden (1984), Holly (1982), Newey (1985a,b), and among authors studied this information test and applied it to econometrics. Lin and Wei (1991) extend the White (1982) to partial likelihood set-up and use it to test goodness-of-fit for Cox models. Zhang (2001) extends to semiparametric empirical likelihood case and use it to assess the model fit for retrospective binary logistic regression models. Here we generalize Zhang (2001) goodness-of-fit test to retrospective PLR models.

Recall that $\theta_i=(\alpha_i,\beta_i^T)^T$, $\theta=(\theta_1^T,\cdots,\theta_{I-1}^T)^T$, $\gamma=(\gamma_1,\gamma_2,\cdots,\gamma_{I-1})^T$. For $i=1,2,\cdots,I$, and $j=1,2,\cdots,n_i$, the log-likelihood of α and β based on retrospective model (2·2) at point X_{ij} is defined to be

$$l_{ij}(\gamma,\theta)=\log\{1+\sum_{m=1}^{I-1}\rho_m\exp[\gamma_m+g_m(X_{ij},\theta)]\}+I_{[i<I]}\left[\gamma_i+g_i(X_{ij},\theta)\right]\cdot \qquad (3\cdot1)$$

Then the semiparametric profile log-likelihood can be written as

$$l(\gamma,\theta)=n\log n_I-\sum_{i=1}^{I}\sum_{j=1}^{n_i}l_{ij}(\gamma,\theta)\cdot$$

For the notational simplicity, denote $\phi = (\gamma^\tau, \theta^\tau)$ and $\phi_0 = (\gamma_0^\tau, \theta_0^\tau)$. Furthermore, denote

$$U_n(\phi_0) = U_n(\gamma_0, \theta_0) = \frac{1}{n}\frac{\partial^2 l(\phi_0)}{\partial\phi\phi^T} = \frac{1}{n}\begin{pmatrix} \frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\gamma\partial\gamma^\tau} & \frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\gamma\partial\theta^\tau} \\ \frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\theta\partial\gamma^\tau} & \frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\theta\partial\theta^\tau} \end{pmatrix} \tag{3·2}$$

and

$$\begin{aligned} V_n(\phi_0) = V_n(\gamma_0, \theta_0) &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\{\frac{\partial l_{ij}(\phi_0)}{\partial\phi}\}\{\frac{\partial l_{ij}(\phi_0)}{\partial\phi^\tau}\} \\ &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\begin{pmatrix} \frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma^\tau} & \frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\theta^\tau} \\ \frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\theta}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma^\tau} & \frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\theta}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\theta^\tau} \end{pmatrix}, \end{aligned} \tag{3·3}$$

where The cell elements of (3·2) and (3·3) are specified in the proof of the following information matrix equality given in the appendix.

Theorem 2. *Under model retrospective polychotomous logistic regression model (2·2), we have*

$$E\Big[U_n(\gamma_0, \theta_0)\Big] = -\begin{pmatrix} S_{11} & S_{21}^T \\ S_{21}^T & S_{33} \end{pmatrix} = -S \tag{3·4}$$

and

$$E\Big[V_n(\gamma_0, \theta_0)\Big] = \begin{pmatrix} S_{11} & S_{21}^T \\ S_{21}^T & S_{33} \end{pmatrix} = S \tag{3·5}$$

Consequently, the following Information Matrix Equality *holds.*

$$E\big(U_n(\gamma_0, \theta_0) + V_n(\gamma_0, \theta_0)\big) = 0 \tag{3·6}$$

where $U_n(\gamma_0, \theta_0)$ *and* $V_n(\gamma_0, \theta_0)$ *are score derivative matrix and squared score matrix respectfully.*

Since the matrix $U_n(\gamma_0, \theta_0) + V_n(\gamma_0, \theta_0)$ is unobservable, we can estimate it by substituting values of the parameters γ_0 and θ_0 with the consistent semiparametric estimates $\tilde{\gamma}$ and $\tilde{\theta}$. It is expected that the matrix is close to a zero matrix if our model fits the data well. By using this fact, we propose a a Wald-type test statistic to assess the global fit of the adjacent category logit model based on case-control data based on the difference between the estimated consistent semiparametric score derivative matrix $U_n(\tilde{\gamma}, \tilde{\theta})$ and the squared score matrix $V_n(\tilde{\gamma}, \tilde{\theta})$ to assess the fit of the model in the next section.

4. Construction of IM χ^2 Test Statistics

We will follow the approach of White (1982) and Zhang (2001) to construct the test statistic. Since the matrix $U_n(\tilde{\gamma}, \tilde{\theta}) + V_n(\tilde{\gamma}, \tilde{\theta})$ is symmetric, we now construct a test statistic based on the lower triangular elements of the matrix. Let q be the number of parameters in the retrospective polychotomous logistic regression model. Denote $\phi = (\phi_1, \phi_2, \cdots, \phi_{q-1}, \phi_q) = (\gamma_1, \cdots, \gamma_I, \theta_1^T, \theta_2^T, \cdots, \theta_I^T)^T$ where θ_i is specified in remark 2.3. Furthermore, let $k = v + (u-1)q) - u(u-1)/2, 1 \leqslant u \leqslant v \leqslant q$. It is seen that k and the ordered pair (u, v) have one-to-one correspondence. Define

$$d_k(\phi : X_{ij}) = d_{uv}(\phi : X_{ij}) = \frac{\partial^2 l_{ij}(\phi : X_{ij})}{\partial \phi_u \partial \phi_v} - \frac{\partial l_{ij}(\phi : X_{ij})}{\partial \phi_u} \frac{\partial l_{ij}(\phi : X_{ij})}{\partial \phi_v} \tag{4·1}$$

$$D_{kn}(\phi : X_{ij}) = \frac{1}{n} \sum_{i=0}^{I} \sum_{j=1}^{n_i} d_{uv}(\phi, X_{ij}), \quad \text{where} \quad k = v + (u-1)q) - u(u-1)/2\cdot$$

The explicit expression of $d_k(\phi : X_{ij})$ can be found in the proof of the information matrix equality theorem in the appendix. Using the one-to-one correspondence between $k = v + (u-1)q) - u(u-1)/2, 1 \leqslant u \leqslant v \leqslant q$ and the ordered pair u, v, we define the $s = q(q+1)/2$ dimensional random vector based on the elements in the upper triangular portion (including the main diagonal elements) of $U_n(\phi) + V_n(\phi)$ as follows

$$D_n(\phi : X_{ij}) = \big(D_{1n}(\phi : X_{ij}), D_{2n}(\phi : X_{ij}), \cdots, D_{sn}(\phi : X_{ij})\big), \quad \tilde{D}_n(\phi) = D_n(\tilde{\phi} : X_{ij}) \tag{4·2}$$

where $\tilde{\phi}$ is the semiparametric empirical likelihood estimate derived from (2·9) and (2·10). Since the log likelihood function function is third order differentiable, the partial derivatives and the expected matrix exist. For notational convenience, we suppress the notation in (3·1) as $l_0(\phi : X_{ij}) = l_{ij}(\gamma, \theta)$. We define,

$$\nabla D_{nk}(\alpha, \beta) = \frac{\partial D_{nk}(\alpha, \beta)}{\partial \phi} = \left(\frac{\partial D_{kn}(\phi : X_{ij})}{\partial \phi_1}, \cdots, \frac{\partial D_{kn}(\phi : X_{ij})}{\partial \phi_q} \right) \tag{4·3}$$

where, for $l = 1, 2, \cdots, q$,

$$\frac{\partial D_{kn}(\phi : X_{ij})}{\partial \phi_l} = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(\frac{\partial^3 l_0(\phi, X_{ij})}{\partial \phi_l \partial \phi_u \partial \phi_v} - \frac{\partial^2 l_0(\phi, X_{ij})}{\partial \phi_l \partial \phi_u} \frac{\partial l_0(\phi, X_{ij})}{\partial \phi_v} - \frac{\partial l_0(\phi, X_{ij})}{\partial \phi_u} \frac{\partial^2 l_0(\phi, X_{ij})}{\partial \phi_l \partial \phi_v} \right)$$

The above first, second and third order (partial) derivatives of $l_0(\phi : X_{ij})$ can be explicitly expressed in terms of data values and the values of parameters. Define

$$b = E\left(\frac{\partial D_n(\phi, X_{ij})}{\partial \phi} \right) = (b_{kl}) \quad \text{for} \quad 1 \leqslant k \leqslant s, 1 \leqslant l \leqslant q\cdot \tag{4·4}$$

Clearly, $\nabla D(\phi)$ is an $s \times q$ matrix with cell elements specified by

$$b_{kl} = \frac{1}{1+\rho} \int \sum_{i=1}^{I} \left(\frac{\partial^3 l_0(\phi,x)}{\partial\phi_l \partial\phi_u \partial\phi_v} - \frac{\partial^2 l_0(\phi,x)}{\partial\phi_l \partial\phi_u} \frac{\partial l_0(\phi,x)}{\partial\phi_v} - \frac{\partial l_0(\phi,x)}{\partial\phi_u} \frac{\partial^2 l_0(\phi,x)}{\partial\phi_l \partial\phi_v} \right) dF_i(x) x \cdot$$

We need few more notations before presenting the main result. Denote

$$\Psi_1 = (\varphi_{kk'}), \quad \Lambda = (\pi_{kw}), \quad \text{for} \quad 1 \leqslant k, k' \leqslant s, 1 \leqslant w \leqslant q$$

be $s \times s$ and $s \times q$ matrices with, as usual, $k = v+(u-1)q-u(u-1)/2, k' = v'+(u'-1)q-u'(u'-1)/2$ and

$$\varphi_{kk'} = \sum_{i=1}^{I} \left(\frac{\rho_i}{1+\rho} \right)^2 \int d_{uv}(\phi : x) d_{u'v'}(\phi : x) dF_i(x)$$

$$+ \sum_{1 \leqslant i \neq i' \leqslant I} \frac{\rho_i \rho_{i'}}{(1+\rho)^2} \int d_{uv}(\phi_0 : x) dG_i(x) \int d_{u'v'}(\phi_0 : x) dF_{i'}(x)$$

$$\pi_{kw} = \sum_{i=1}^{I} \left(\frac{\rho_i}{1+\rho} \right)^2 \int \rho_i d_{uv}(\phi : x) \frac{\partial l(\phi : x)}{\partial\phi_w} dF_i(x)$$

$$+ \sum_{1 \leqslant i \neq i' \leqslant I} \frac{\rho_i \rho_{i'}}{(1+\rho)^2} \int d_{uv}(\phi_0 : x) dG_i(x) \int \frac{\partial l(\phi_0 : x))}{\partial\phi_w} dF_{i'}(x)$$

Define

$$\Sigma = \Psi_1 + b^T [S^{-1} - (1+\rho)\Omega] b - 2\Lambda S^{-1} b$$

Based on the above notations, we state the main result as follows

Theorem 3. *Assume the retrospective polychotomous logistic regression model (2·2) and* $\tilde{\phi} = (\tilde{\gamma}, \tilde{\theta})$ *is the semiparametric likelihood estimate of* $\phi = (\gamma, \theta)$*. We have*

$$\sqrt{n} D_n(\tilde{\phi}) = \sqrt{n} \left[D_n(\phi_0) + \frac{1}{n} b^T S^{-1} \frac{\partial l(\phi_0)}{\partial\phi} \right] + o_p(1) \to N_s(0, \Sigma) \cdot \tag{4·5}$$

Furthermore, if Σ^{-1} *exists, we have*

$$n \big[D_n(\tilde{\phi}) \big]^T \Sigma^{-1}(\tilde{\phi}) D_n(\tilde{\phi}) \to \chi_s^2 \tag{4·6}$$

Proof See the appendix.

The one dimensional statistic $n \big[D_n(\tilde{\phi}) \big]^T \Sigma^{-1}(\tilde{\phi}) D_n(\tilde{\phi})$ measures the discrepancy between the true model (2·2) and any incorrect model. The level of significance can be evaluated through the asymptotic distribution.

Since the covariance matrix Σ is unobservable, we can use the sample version $\tilde{\Sigma}$ by substituting the true value of parameter (ϕ) with the semiparametric estimate ($\tilde{\phi}$), in the calculation of the covariance matrix, we need the cumulative distribution functions $F_i(x), i = 1, 2, \cdots, I-1$. We can replace $F_i(x)$ by our semiparametric estimators specified in (2·12). If we replace the consistent semiparametric estimate $\tilde{\Sigma}$ of Σ and assume further that the inverse of $\tilde{\Sigma}(\tilde{\phi})$ exists, then

$$\left[\tilde{D}_n(\tilde{\phi})\right]^T \tilde{\Sigma}^{-1}(\tilde{\phi}) \tilde{D}_n(\tilde{\alpha}, \tilde{\beta}) \rightarrow \chi^2_s$$

Remark 4.1. If some of of the elements in the matrix $\tilde{Q}_n$ are linear combinations of the others, then the estimated covariance matrix would be singular. In this case, we replace $\tilde{\Sigma}^{-1}$ with the generalized inverse matrix $\tilde{\Sigma}^{+1}$. So we will lose some degrees of freedom. The asymptotic distribution of $D_n(\tilde{\phi})$ is still a χ^2_r distribution with degrees of freedom $r \leqslant s$.

5. A Case Study and Numerical Results

In this section we study the power of the proposed information matrix test under a local alternative. Consider the following model alternative to PLR (2·2)

$$\begin{cases} X_{I1}, \cdots, X_{In_I} \overset{\text{i·i·d·}}{\sim} f_I(x), \\ X_{i1}, \cdots, X_{in_i} \overset{\text{i·i·d·}}{\sim} \exp\big(\gamma_i + g_i(x, \theta) + \kappa_i(\xi, x)\big) \cdot f_I(x), \quad \text{for} \quad i = 1, \cdots, I-1, \end{cases} \tag{5·1}$$

where $\kappa_i(\xi, x)$ is a pre-specified function. Furthermore, assume that there exists a unique ξ_0 such that $\kappa_i(\xi_0, x) = 0$. Therefore, testing that H_0 : *PLR (2·2) is valid* is equivalent to testing $H_0 : \xi = \xi_0$ under model (5·1).

As a special case, we choose the null model as the generalized logit model

$$\begin{cases} X_{I1}, \cdots, X_{1n_1} \overset{\text{i·i·d·}}{\sim} f_1(x) = \exp(\alpha_1 + \beta_1 x), \\ X_{I1}, \cdots, X_{2n_2} \overset{\text{i·i·d·}}{\sim} f_2(x) = \exp(\alpha_3 + \beta_3 x), \\ X_{i1}, \cdots, X_{in_i} \overset{\text{i·i·d·}}{\sim} f_3(x) \end{cases} \tag{5·2}$$

Consider the following alternative to (5·2)

$$\begin{cases} X_{I1}, \cdots, X_{1n_1} \overset{\text{i·i·d·}}{\sim} f_1^a(x) = \exp(\alpha_1 + \beta_1 x + \gamma x^2), \\ X_{I1}, \cdots, X_{2n_2} \overset{\text{i·i·d·}}{\sim} f_2(x) = \exp(\alpha_3 + \beta_3 x), \\ X_{i1}, \cdots, X_{in_i} \overset{\text{i·i·d·}}{\sim} f_3(x) \end{cases} \tag{5·3}$$

Similar to the approach used in Zhang (2002) to get theoretical procedure in assessing the asymptotic power at the local alternative. Since the multiple approximations have been used in this large sample test, we will not focus on the power. In stead, we present the following numerical example to illustrate the implementation of the proposed goodness-of-fit test.

Example Table 5.2 in McCullagh and Nelder (1989) contains data concerning the degree of pneumoconiosis in coalface workers as a function of exposure x measured in years. McCullagh and Nelder (1989) analyzed this data set by employing the proportional odds model and the continuation-ratio logit model. Let X denote "Period spent (years)" and Y represent "prevalence of pneumoconiosis" in which $Y = 0, 1$, and 2 stand for three categories: Normal, Mild pneumoconiosis, and Severe pneumoconiosis. Since the sample data $(X_i, Y_i), i = 1, \cdots, 371$, can be thought as being drawn independently and identically from the joint distribution of (X, Y).

We fit model (5·2) to this data and obtain semiparametric empirical likelihood point estimates $\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\alpha}_2, \tilde{\beta}_2) = (2) = (-2{\cdot}2628, 0{\cdot}0836, -3{\cdot}1776, 0{\cdot}1093)$. The information matrix χ^2 statistic is 7·988 with 10 degrees of freedom. The corresponding observed p-value is 0.37 indicating the goodness-of-fit for of the model. This also agrees with the result obtained by Zhang (2002) via the Kolmogorov-Smirnov test. Since model (5·2) fits the data appropriately, all empirical likelihood estimates of the parameters are statistically valid for making inference.

6. Summary

In this paper, an information matrix goodness-of-fit test has been proposed for testing the adequacy of the retrospective polychotomous logistic regression models. Since the proposed test statistic has a χ^2 distribution under the null hypothesis that the retrospective polychonomous logistic regression is correctly specified, the bootstrap procedure is not needed to find the p-value in order to draw conclusions.

This procedure is, in fact, a nonparametric methods. It is also a computational intensive method and involves inversion of high dimension matrix.

Finally, the method proposed in this paper can be applied directly to all semiparametric models with the general form specified in (2·2).

Appendix

Proof Theorem 2. (Information Matrix Equality). We still use the notation defined in (2·13), (2·14) and (2·15). First of all, we calculate expectations of these second order derivatives of the semiparametric log-likelihood. For $u, v = 1, 2, \cdots, I-1$,

$$E\Big(\frac{1}{n}\frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\gamma_u\partial\gamma_v}\Big) = E\Big(\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{e_u(X_{ij})e_v(X_{ij})}{\big[1+\sum_{i=1}^{I-1}e_i(X_{ij})\big]^2}\Big) = \frac{1}{1+\rho}\int P(x)e_u(x)e_v(x)dF_I(x) \quad \text{(A·1)}$$

if $u \neq v$. For $u = v$, we have

$$E\Big(\frac{1}{n}\frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\gamma_u^2}\Big) = -\frac{1}{1+\rho}\int P(x)e_u(x)\sum_{u\neq v, v=1}^{I-1}e_v(x)dF_I(x), \quad \text{(A·2)}$$

$$E\Big(\frac{1}{n}\frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\gamma_u\partial\theta^\tau}\Big) = -E\Big\{\frac{1}{n}\sum_{s=1}^{n}\frac{e_u(T_s)g_u^\theta(T_s)}{1+\sum_{m=1}^{I-1}e_u(T_s)} - \frac{1}{n}\sum_{s=1}^{n}\frac{e_u(T_s)\sum_{m=1}^{I-1}e_m(T_s)g_m^\theta(T_s)}{\{1+\sum_{m=1}^{I-1}e_m(T_s)\}^2}\Big\}$$
$$= -\frac{1}{1+\rho}\int e_u(x)\Bigg(g_u^\theta(x) - \sum_{m=1}^{I-1}P(x)e_m(x)g_m^\theta(x)\Bigg)dF_I(x), \quad \text{(A·3)}$$

and $$E\left(\frac{1}{n}\frac{\partial^2 l(\gamma_0,\theta_0)}{\partial\theta\partial\theta^\tau}\right) = \frac{1}{n}E\left(\sum_{i=1}^{I}\sum_{j=1}^{n_i}g_i^{\theta\theta^T}(X_{ij}) - \sum_{s=1}^{n}\frac{\sum_{m=1}^{I-1}e_m(T_s)g_m^\theta(T_s)g_m^{\theta^T}(T_s)}{1+\sum_{m=1}^{I-1}e_m(T_s)}\right)$$
$$+\frac{1}{n}E\left(\sum_{s=1}^{n}\frac{\sum_{m=1}^{I-1}e_m(T_s)g_m^{\theta\theta^T}(T_s)}{1+\sum_{m=1}^{I-1}e_m(T_s)} + \sum_{s=1}^{n}\frac{\sum_{m=1}^{I-1}e_m(T_s)g_m^\theta(T_s)}{1+\sum_{m=1}^{I-1}e_m(T_s)}\times\frac{\sum_{m=1}^{I-1}e_m(T_s)g_m^{\theta^T}(T_s)}{1+\sum_{m=1}^{I-1}e_m(T_s)}\right)$$
$$= -\frac{1}{1+\rho}\int\left(\sum_{m=1}^{I-1}e_m(x)g_m^\theta(x)g_m^{\theta^T}(x) - \frac{[\sum_{m=1}^{I-1}e_m(x)g_m^\theta(x)][\sum_{m=1}^{I-1}e_m(x)g_m^{\theta^T}(x)]}{1+\sum_{m=1}^{I-1}e_m(x)}\right)dF_I(x)\cdot \quad \text{(A·4)}$$

From equations (A·1), (A·2), (A·3)and (A·4) we obtain

$$E\Big[U_n(\gamma_0,\theta_0)\Big] = -\begin{pmatrix} S_1 & S_2 \\ S_2^T & S_3\end{pmatrix} = -S\cdot \quad \text{(A·5)}$$

Next, we find the expectation of the squared score matrix. Observe that, for $u = 1, 2, \cdots, I-1$,

$$\frac{\partial l_{ij}(\gamma_0,\theta_))}{\partial\gamma_u} = \frac{e_u(X_{ij})}{1+\sum_{m=1}^{I-1}e_m(X_{ij})} - I_{[i<I]}I_{[i=u]}\cdot$$

where $I_{[\cdot]}$ is the indicator function. Taking the product of the above derivatives, and we have

$$\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma_u}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma_v} = \frac{e_u(X_{ij})e_v(X_{ij})}{\Big[1+\sum_{m=1}^{I-1}e_m(X_{ij})\Big]^2} - \frac{e_u(X_{ij})I_{[i<I]}I_{[i=v]}}{1+\sum_{m=1}^{I-1}e_m(X_{ij})}$$

$$-\frac{e_v(X_{ij})I_{[i<I]}I_{[i=u]}}{1+\sum_{m=1}^{I-1}e_m(X_{ij})}+I_{[i<I]}I_{[i=v]}I_{[i=u]}\cdot$$

Note that if $u\neq v$, then $I_{[i<I]}I_{[i=v]}I_{[i=u]}=0$ and

$$E\Big(\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{e_v(X_{ij})I_{[i<I]}I_{[i=u]}}{1+\sum_{m=1}^{I-1}e_m(X_{ij})}\Big)=\frac{1}{1+\rho}\int\frac{e_u(x)e_v(x)}{1+\sum_{m=1}^{I-1}e_m(x)}dF_I(x)\cdot \tag{A·6}$$

After some algebra, we have

$$E\Big(\frac{1}{n}\sum_{i=1}^{I-1}\sum_{j=1}^{n_i}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma_u}\frac{\partial l_{ij}(\gamma_),\theta_0)}{\partial\gamma_v}\Big)=-\frac{1}{1+\rho}\int\frac{e_u(x)e_v(x)}{1+\sum_{m=1}^{I-1}e_m(x)}dF_I(x)\cdot \tag{A·7}$$

If $u=v$, then $E\big(I_{[i<I]}I_{[i=v]}I_{[i=u]}\big)=\int e_u(:x)dF_I(x)$, along with (A·6), we have

$$E\big(\frac{1}{n}\Big[\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma_u}\Big]^2\big)=\frac{1}{1+\rho}\int\frac{e_u(x)\sum_{m=1,m\neq u}^{I-1}e_m(x)}{1+\sum_{m=1}^{I-1}e_m(x)}dF_I(x)\cdot \tag{A·8}$$

Since

$$\Big(\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\gamma}\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\theta^\tau}\Big)^\tau=\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\theta}\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\gamma^\tau}\cdot$$

We only need to calculate the expectation of the matrix on the right side of the above equation. The explicit expression of the right hand side of the above equation is, for $u=1,2,\cdots,I-1$,

$$\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\theta}\frac{\partial l_{ij}(\gamma_0,\theta_0)}{\partial\gamma_u}=\frac{e_u(X_{ij})\big(\sum_{m-1}^{I-1}e_m(X_{ij})g_m^\theta(X_{ij})\big)}{\big[1+\sum_{m=1}^{I-1}e_m(:X_{ij})\big]^2}$$

$$-\frac{\sum_{m-1}^{I-1}e_m(X_{ij})g_m^\theta(X_{ij})I_{[i=u]}}{1+\sum_{m=1}^{I-1}e_m(X_{ij})}-\frac{e_u(X_{ij})g_i^\theta(X_{ij})}{1+\sum_{m=1}^{I-1}e_m(X_{ij})}+g_i^\theta(X_{ij})I_{[i=u]}\cdot$$

With the above expression we obtain

$$E\Big(\frac{1}{n}\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\theta}\frac{\partial l_{ij}(\gamma,\theta_0)}{\partial\gamma^T}\Big)=\frac{1}{1+\rho}\int e_u(x)\Big(g_m^\theta(x)-\frac{\sum_{m-1}^{I-1}e_m(x)g_m^\theta(x)}{1+\sum_{m=1}^{I-1}e_m(x)}\Big)dF_I(x)\cdot \tag{A·9}$$

Finally, Notice that

$$\frac{\partial l_{ij}(\gamma,\theta)}{\partial\theta}\frac{\partial l_{ij}(\gamma,\theta)}{\partial\theta^\tau}=\frac{[\sum_{m=1}^{I-1}e_m(x)g_m^\theta(X_{ij})][\sum_{m=1}^{I-1}e_m(x)g_m^{\theta^T}(X_{ij})]}{[1+\sum_{m=1}^{I-1}e_m(x)]^2}+\sum_{j=1}^{n_i}g_i^\theta(X_{ij})g_i^{\theta^T}(X_{ij})I_{[i<I]}$$

$$-\frac{[\sum_{m=1}^{I-1}e_m(x)g_m^\theta(X_{ij})]g_i^{\theta^T}(X_{ij})I_{[i<I]}}{1+\sum_{m=1}^{I-1}e_m(x)}-\frac{[\sum_{m=1}^{I-1}e_m(x)g_m^{\theta^T}(X_{ij})]g_i^\theta(X_{ij})I_{[i<I]}}{1+\sum_{m=1}^{I-1}e_m(x)}$$

Therefore,

$$E\Big(\frac{1}{n}\frac{\partial l_{ij}(\gamma,\theta)}{\partial\theta}\frac{\partial l_{ij}(\gamma,\theta)}{\partial\theta^\tau}\Big)$$

$$= \frac{1}{1+\rho} \int \Big(\sum_{m=1}^{I-1} e_m(x) g_m^\theta(x) g_m^{\theta T} - \frac{\sum_{m=1}^{I-1} e_m(x) g_m^\theta(x) \sum_{m=1}^{I-1} e_m(x) g_m^\theta(x)}{1+\sum_{m=1}^{I-1} e_m(x)} \Big) dF_I(x) \cdot \qquad (\text{A}\cdot 10)$$

Combine the results in equations (A·7), (A·8), (A·9) and (A·10), we have

$$E\Big(V_n(\phi)\Big) = E\left(\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\{\frac{\partial l_{ij}(\phi_0)}{\partial \phi}\}\{\frac{\partial l_{ij}(\phi_0)}{\partial \phi^\tau}\}\right) = \begin{pmatrix} S_1 & S_2 \\ S_2^\tau & S_3 \end{pmatrix} = S\cdot \qquad (\text{A}\cdot 11)$$

Combining Equations (A·5) and (A·11) we get $E\Big(V_n(\phi)+U_n(\phi)\Big) = 0$ which proves the theorem.

Proof Theorem 3. Assume that the polychotomous logistic regression model (2·2) holds under case-control data. We have proved that $\tilde{\phi} = (\tilde{\gamma}, \tilde{\theta})$ is the consistent empirical likelihood estimate of $\phi = (\gamma, \theta)$. Let $\phi_0 = (\phi_{10}, \phi_{20}, \cdots, \phi_{q0})$ be the true value of the vector of parameters ϕ. Applying the weak law of large numbers, we have $b = \partial D_n(\phi_0 : X_{ij})/\partial\phi + o_P(1)$. Using the first order Taylor expansion $D(\tilde{\phi})$ at a neighborhood of the true value of the parameter ϕ_0, we have

$$D(\tilde{\phi} : X_{ij}) = D(\phi_0 : X_{ij}) + \frac{\partial D(\phi_0 : X_{ij})}{\partial \phi}(\tilde{\phi} - \phi_0) + o_P(||\tilde{\phi} - \phi_0||) \qquad (\text{A}\cdot 12)$$

where $||\tilde{\phi} - \phi_0|| = O_P(n^{-1/2})$ (by Theorem 1). Hence

$$\sqrt{n} D(\tilde{\phi} : X_{ij}) = \sqrt{n}\left(D(\phi_0 : X_{ij}) + \frac{1}{n} b S^{-1} \frac{\partial l \phi_0 : X_{ij}}{\partial \phi}\right) + o_P(1) \qquad (\text{A}\cdot 13)$$

From Theorems 1 and 2, we can see that

$$E\left(D(\phi_0 : X_{ij}) + \frac{1}{n} b S^{-1} \frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right) = E\left(D(\phi_0 : X_{ij})\right) + \frac{1}{n} b S^{-1} E\left(\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right) = 0$$

and

$$\text{var}\left[D(\phi_0 : X_{ij}) + \frac{1}{n} b S^{-1} \frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right] = \text{var}\left[D(\phi_0 : X_{ij})\right] + \text{var}\left[\frac{1}{n} b S^{-1} \frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right]$$
$$+2\text{E}\left[D(\phi_0 : X_{ij})\right]\left[\frac{1}{n}\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi^T} S^{-1} b^T\right] \qquad (\text{A}\cdot 14)$$

Let

$$\Psi_1 = \text{var}\left[D(\phi_0 : X_{ij})\right], \quad \Lambda = \text{E}\left[D(\phi_0 : X_{ij})\right]\left[\frac{1}{n}\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi^T}\right]$$

Note that

$$\text{var}\left[\frac{1}{n} b S^{-1} \frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right] = \text{var}\left[b(\tilde{\phi} - \phi_0)\right] = b[S^{-1} - (1+\rho)\Omega] b^T$$

Next we calculate the elements in $\Psi_1 = (\varphi^1_{kk'})$ where $1 \leqslant k, k' \leqslant s$, $k = v+(u-1)q-(u-1)u/2$, $k' = v' + (u'-1)q - (u'-1)u'/2$ and $s = q(q+1)/2$. Note that, for $i \neq i'$, X_{ij} and $X_{i'j'}$ are jointly independent; X_{ij} and $X_{ij'}$ are identical and independent. Using these facts, we have

$$\varphi_{kk'} = \operatorname{var}\Big(D_{nk}(\phi_0 : X_{ij}), D_{nk'(\phi_0 : X_{ij})}\Big) = E\left\{\left[\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} d_{uv}(\phi_0 : X_{ij})\right]\left[\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} d_{u'v'}(\phi_0 : X_{ij})\right]\right\}$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{I}\left(\sum_{j=1}^{n_i} d_{uv}(\phi_0 : X_{ij})\sum_{j=1}^{n_1} d_{u'v'}(\phi_0 : X_{ij})\right) + \sum_{1\leqslant i\neq i'\leqslant I}\left(\sum_{j=1}^{n_i} d_{uv}(\phi_0 : X_{ij})\sum_{j=1}^{n_{i'}} d_{u'v'}(\phi_0 : X_{i'j})\right)\right]$$

$$= \sum_{i=1}^{I}\left(\frac{\rho_i}{1+\rho}\right)^2 \int d_{uv}(\phi_0 : x) d_{u'v'}(\phi_0 : x) dF_i(x)$$

$$+ \sum_{1\leqslant i\neq i'\leqslant I}^{I} \frac{\rho_i\rho_{i'}}{(1+\rho)^2}\int d_{uv}(\phi_0 : x) dG_i(x)\int d_{u'v'}(\phi_0 : x) dF_{i'}(x) \qquad \text{(A·15)}$$

Similarly, we can calculate the elements in $\Lambda = (\pi_{kw})$ for $1 \leqslant k = v + (u-1)q - u(u-1)/2 \leqslant s$ and $1 \leqslant w \leqslant q$ as follows

$$\pi_{kw} = \frac{1}{n^2}E\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i} d_{uv}(\phi_0 : X_{ij})\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{\partial l(\phi_0 : X_{ij})}{\partial w}\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{I}\left(\sum_{j=1}^{n_i} d_{uv}(\phi_0, X_{ij})\sum_{j=1}^{n_i}\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi_w}\right) + \sum_{1\leqslant i\neq i'\leqslant I}\left(\sum_{j=1}^{n_i} d_{uv}(\phi_0 : X_{ij})\sum_{j=1}^{n_{i'}}\frac{\partial l(\phi_0 : X_{i'j})}{\partial \phi_w}\right)\right]$$

$$= \sum_{i=1}^{I}\left(\frac{\rho_i}{1+\rho}\right)^2 \int d_{uv}(\phi_0 : x)\frac{\partial l(\phi_0 : x)}{\partial \phi_w} dF_i(x)$$

$$+ \sum_{1\leqslant i\neq i'\leqslant I}^{I} \frac{\rho_i\rho_{i'}}{(1+\rho)^2}\int d_{uv}(\phi_0 : x) dG_i(x)\int \frac{\partial l(\phi_0 : x)}{\partial \phi_w} dF_{i'}(x) \qquad \text{(A·16)}$$

Therefore,

$$\Sigma \equiv \operatorname{var}\left[D(\phi_0 : X_{ij}) + \frac{1}{n}bS^{-1}\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right] = \Psi_1 + b[S^{-1} - (1+\rho)\Omega + 2\Lambda S^{-1}]b^T \qquad \text{(A·17)}$$

Since $\partial l(\alpha_0, \beta_0)/\partial\phi$ is a multivariate normal random variable, using multivariate central limit theorem and Slutsky's theorem, we have

$$\sqrt{n}D(\tilde{\phi} : X_{ij}) = \sqrt{n}\left[D(\phi_0) + \frac{1}{n}bS^{-1}\frac{\partial l(\phi_0 : X_{ij})}{\partial \phi}\right] + o_p(1) \to N_s(0, \Sigma)\cdot \qquad \text{(A·18)}$$

As a consequence, if Σ^{-1} exists,

$$nD^T(\tilde{\phi})\Sigma^{-1}D(\tilde{\phi}) \to_d \chi^2_s$$

which completes the proof of theorem 3.

References

Agresti, A.: 2002, *Categorical Data Analysis*, Wiley & Sons, Inc.

Anderson, J. A.: 1984, Regression and ordered categorical variables, *J. Roy. Statist. Soc. Ser. B* **46**(1), 1–30.

Begg, C. B. and Gray, R.: 1984, Calculation of polychotomous logistic regression parameters using individualized regressions, *Biometrika* **71**(1), 11–18.

Cox, A.: 1988, Multinomial regression models based on continuation ratios, *Statistics in Medicine* **7**, 435–441.

Farewell, V. T.: 1979, Some results on the estimation of logistic models based on retrospective data, *Biometrika* **66**(1), 27–32.

Fienberg, S. E.: 1978, *The Analysis of cross clasified categorical data*, MIT-Press, Cambridge.

Hausman, J. A.: 1978, Specification tests in econometrics, *Econometrica* **46**(6), 1251–1271.

Hausman, J. and McFadden, D.: 1984, Specification tests for the multinomial logit model, *Econometrica* **52**(5), 1219–1240.

Hemker, B. T., van der Ark, L. A. and Sijtsma, K.: 2001, On measurement properties of continuation ratio models, *Psychometrika* **66**(4), 487–506.

Holly, A.: 1982, A remark on Hausman's specification test, *Econometrica* **50**(3), 749–759.

Kvist, T., G. H. and Thyregod, P.: 2000, Using continuation ratio logits to analyze the variation of the age composition of fish catches, *Journal of Applied Statistics* **27**, 303–319.

Lin, D. Y. and Wei, L. J.: 1991, Goodness-of-fit tests for the general Cox regression model, *Statist. Sinica* **1**(1), 1–17.

Marshall, A. W. and Olkin, I.: 1997, A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika* **84**(3), 641–652.

McCullagh, P.: 1980, Regression models for ordinal data, *J. Roy. Statist. Soc. Ser. B* **42**(2), 109–142.

McCullagh, P. and Nelder, J. A.: 1989, *Generalized Linear Models*, Chapman and Hall.

Newey, W. K.: 1985a, Generalized method of moments specification testing, *J. Econometrics* **29**(3), 229–256.

Newey, W. K.: 1985b, Maximum likelihood specification testing and conditional moment tests, *Econometrica* **53**(5), 1047–1070.

Owen, A.: 1990, Empirical likelihood ratio confidence regions, *Ann. Statist.* **18**(1), 90–120.

Owen, A. B.: 1988, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika* **75**(2), 237–249.

Peng, C. and Zhang, B.: 2008, A note on goodness-of-fit test for continuation ratio logistic regression models under case-control data, *JSPI* **138**, 2355–2365.

Prentice, R. L. and Pyke, R.: 1979, Logistic disease incidence models and case-control studies, *Biometrika* **66**(3), 403–411.

Qin, J. and Zhang, B.: 1997, A goodness-of-fit test for logistic regression models based on case-control data, *Biometrika* **84**(3), 609–618.

Rindorf, T. and Lewy, P.: 2001, Analysis of length and age distributions using continuation ratiuo logits, *Canadian Journal of Fishery and Aquatic Sciences* **58**, 1141–1152.

Scott, A. J. and Wild, C. J.: 1986, Fitting logistic models under case-control or choice based sampling, *J. Roy. Statist. Soc. Ser. B* **48**(2), 170–182.

Tutz, G.: 1991, Sequential models in categorical regression, *Comput. Statist. Data Anal.* **11**(3), 275–295.

White, H.: 1982, Maximum likelihood estimation of misspecified models, *Econometrica* **50**, 1–26.

Zhang, B.: 2001, An information matrix test for logistic regression models based on case-control data, *Biometrika* **88**(4), 921–932.

Zhang, B.: 2002, Assessing goodness-of-fit of generalized logit models based on case-control data, *J. Multivariate Anal.* **82**(1), 17–38.

Zhang, B.: 2004, Assessing goodness-of-fit of categorical regression models based on case-control data, *Aust. N. Z. J. Stat.* **46**(3), 407–423.