Journal of the Korean
Statistical Society
Vol. 26, No. 2, 1997

On a Bayes Criterion for the Goodness-of-Link Test for Binary Response Regression Models : Probit Link versus Logit Link

Hea-Jung Kim [1]

ABSTRACT

In the context of binary response regression, the problem of constructing Bayesian goodness-of-link test for testing logit link versus probit link is considered. Based upon the well known facts that cdf of logistic variate $\approx$ cdf of $t_8/.634$ and, as $\nu \rightarrow \infty$, cdf of t_ν approximates to that of $N(0,1)$, Bayes factor is derived as a test criterion. A synthesis of the Gibbs sampling and a marginal likelihood estimation scheme is also proposed to compute the Bayes factor. Performance of the test is investigated via Monte Carlo study. The new test is also illustrated with an empirical data example.

Key Words : Binary response regression; Probit link versus logit link; Goodness-of-link test; Bayes factor; Marginal likelihood.

[1] Department of Statistics, Dongguk University, Seoul, 100-715, Korea

1. INTRODUCTION

The theory of binary dependent variable regression has its genesis in bioassay(see, for example, Finney, 1971). In that context, Y_i denotes a Bernoulli random variable with the probablilty of success p_i, $i = 1, \cdots, n$, where p_i is related to a set of covariates which may be continuous or discrete, and the binary response regression model is defined as

$$Y_i = F(X_i'\beta) + u_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where the u_i is uncorrelated random error with $Eu_i = 0$, X_i is a vector of, say, q fixed covariates, β is a vector of unknown coefficients, and $F(\cdot)$ is a known cdf linking the probability p_i with linear structure $X_i'\beta$ so that $p_i = F(X_i'\beta)$. In particular, when the link cdf F(having link function $F^{-1}(\cdot)$) is taken to be the standard normal cdf, the resulting model is called the probit model, while a logit model is obtained if F is the logistic cdf. These models are discussed extensively in Nelder and McCullagh(1987) and Collett(1991). For the two models, Press(1982) and Griffiths, Hill and Pope(1987) examined estimation problem from non-Bayesian viewpoint, and Dellaportas and Smith(1993) and Newton, Czado and Chappell(1996) examined the same problem with Bayesian approach.

Specify the model (1.1) by choosing the link cdf F to be the family of t distributions. Then the probit link is a member of the model for $t_\infty = N(0, 1)$. Moreover, due to Albert and Chib(1993) and Soofi, Ebrahimi and Habibullah(1995), the most popular logit link function can be approximately viewed as a member of the t distribution family, because cdf of logistic variate $\approx$ cdf of $t_8/.634$. This specification allows one to investigate the sensitivity of fitted probabilities by the choice of link function. In addition, one can examine which value of the degree-of-freedom parameter for t is best supported by data. A type of the latter procedure is referred to as goodness-of-link test. Collett(1991) gave a procedure for the goodness-of-link test for logit link versus complementary log-log link based on the family of link functions proposed by Aranda-Ordaz(1981). However, due to complexity of the link functions to be compared, a formal goodness-of-link test procedure for probit link versus logit link has not been proposed yet. Thus frequentist goodness-of-link test criterion for this non-nested model comparison mainly depends on informal graphical method such as the index plot and the partial residual plot. The aim of this paper is to suggest a formal test criterion via Bayesian approach.

This paper introduces a Bayesian approach to model selection for choosing the best supported model among the family of t link binary regression models.

Then, by means of the approach, we propose a Bayesian goodness-of-link test procedure for probit link versus logit link based upon the family of t link functions. Performance of the suggested test is also examined with illustrative examples.

2. BAYESIAN TEST CRITERION

Denote the binary regression model (1.1) as M and let $\pi(\beta|M)$, a proper or improper prior density, summarize our prior information about β. The marginal likelihood under the model M with link cdf F is given by

$$m(Y|M) = \int \pi(\beta|M) \prod_{i=1}^{n} F(X_i'\beta)^{Y_i}\{1 - F(X_i'\beta)\}^{1-Y_i} d\beta, \qquad (2.1)$$

where $Y = (Y_1, \cdots, Y_n)'$. Suppose that we are interested in comparing any two models M_k(with $F = F_k$) and M_ℓ(with $F = F_\ell$). Usual Bayesian model choice procedure is based upon the Bayes factor:

$$B_{k\ell} = m(Y|M_k)/m(Y|M_\ell). \qquad (2.2)$$

When $B_{k\ell} \geq 1$ or $\ln B_{k\ell} \geq 0$, we may consider M_k as the better model supported by data. See Jeffreys(1961) for interpretive ranges of the Bayes factor. If we set M_1 and M_2 as probit link model and logit link model, respectively, the Bayes factor B_{12} obtained from (2.2) can be a goodness-of -link test criterion for testing M_1 versus M_2.

It is noted that the problem of analytically calculating each marginal likelihood for the Bayes factor B_{12}, which is the normalizing constant of its posterior density, is extremely challenging(cf. Newton, Czado and Chappell, 1996). Recent development of an MCMC(Markov Chain Monte Carlo) method called the Gibbs sampling approach provides several methods that directly address simulation-based calculation of the marginal likelihood(cf. Newton and Raferty 1994, Gelfand and Dey 1994, and Chib 1995). However, the approach is not generally applicable to the calculation of the Bayes factor B_{12}. This is due to the fact that the marginal likelihood of the logit link model, $m(Y|M_2)$, can only be evaluated under specific assumptions: (i) Sample size at hand is large. (ii) Substantial prior information is available for the Bayesian inference. (iii) Resulting posterior distribution of parameters in the model is asymptotically normal. Therefore, as pointed out by Dellaportas and Smith(1993), if the above-mentioned assumptions are not met, calculation of $m(Y|M_2)$ by posterior simulation will turn out to be misled. Thus we need another criterion for the goodness-of-link test which is applicable without any assumptions. In this section, we will propose a general scheme for calculating the Bayes factor B_{12} with the Gibbs sampler so that, without any assumptions, it may always provide a method for computing the Bayes factor.

Let $M(t)$ be a class of regression models for binary response data having the link cdf T_ν, where cdf T_ν is the family of t_ν distributions. Then the following lemma gives an approximate Bayes factor for testing the probit link model versus the logit link model.

Lemma 1. Suppose $M_{(\nu)} \in M(t)$ and $M_{(8)} \in M(t)$ denote the binary regression models (1.1) having t link cdf's T_ν and T_8, respectively. Then, as $\nu \to \infty$, the Bayes factor B_{12} is approximately equivalent to

$$\begin{aligned} B_{\nu,8} &= \frac{m(Y|M_{(\nu)})}{m(Y|M_{(8)})} \\ &= \frac{\int \pi(\beta|M_{(\nu)}) \prod_{i=1}^{n} T_\nu(X_i'\beta)^{Y_i}\{1 - T_\nu(X_i'\beta)\}^{1-Y_i} d\beta}{\int \pi(\beta_8|M_{(8)}) \prod_{i=1}^{n} T_8(X_i'\beta_8)^{Y_i}\{1 - T_8(X_i'\beta_8)\}^{1-Y_i} d\beta_8} \end{aligned}$$

where $\beta_8 = .634\beta$.

Proof. Noticing that $t_\nu, \nu \to \infty$, corresponds to the unit normal distribution, one can generalize the probit link by choosing t_ν link so that the marginal likelihood $m(Y|M_{(\nu)}) \to m(Y|M_1)$ as $\nu \to \infty$. On the other hand, as mentioned in Section 1, cdf of logistic distribution is approximately equivalent to that of $t_8/.634$, and hence $Pr(U1 \le X_i'\beta) \approx Pr(U2/.634 \le X_i'\beta) = Pr(U2 \le X_i'\beta_8)$, where $U1$ and $U2$ are logistic random variable and t_8 random variable, respectively, and β_8 is equal to $.634\beta$. This leads to the following relation between the logit link model(M_2) and the binary response T_8 link cdf regression model $M_{(8)}$:

$$\begin{aligned} Y_i &= F(X_i'\beta) + u_i, \\ &\approx T_8(X_i'\beta_8) + u_i, \quad i = 1, \cdots, n, \end{aligned}$$

where F is the logistic cdf link. From the relation, we see that $M_{(8)}$ with the coefficient vector β_8 is approximately equivalent to M_2, and hence $m(Y|M_2) \approx m(Y|M_{(8)})$.

3. COMPUTING THE BAYES FACTOR

3.1. Gibbs sampler

Consider a model $M_{(\nu)} \in M(t)$. Let $f(Y|\beta, M_{(\nu)}) = \prod_{i=1}^{n} T_\nu(X_i'\beta)^{Y_i}\{1 - T_\nu(X_i'\beta)\}^{1-Y_i}$ denote the sampling density(likelihood function) where $T_\nu(\cdot)$ is the cdf of t distribution with ν(fixed) degrees of freedom. To allow the possibility that the posterior simulation requires data augmentation, we introduce n latent variables $Z_1, \cdots, Z_n$, where the Z_i are independently distributed as a t with location parameter $X_i'\beta$, scale parameter 1, and degrees of freedom ν, such that

$$Z_i \sim t_\nu(X_i'\beta, 1) \text{ and } Y_i = I(Z_i > 0), \quad i = 1, \cdots, n, \tag{3.1}$$

where $I(A)$ is an indicator function of the event A. The above specification in fact defines the t_ν link model because $Pr(Z_i > 0) = T_\nu(X_i'\beta)$. Let us introduce the additional independent random variables λ_i, and write the distribution of Z_i as the following scale mixture of normal distribution:

$$Z_i|\lambda_i \sim N(X_i'\beta, \lambda_i^{-1}) \text{ and } \lambda_i \sim \text{Gamma}(\nu/2, 2/\nu), \quad i = 1, \cdots, n, \tag{3.2}$$

so that $Z_i \sim t_\nu(X_i'\beta, 1)$. Suppose an informative prior distribution $\beta \sim N_q(\alpha, \Omega^{-1})$ is chosen for the regression parameter. Then the joint posterior density of the unobservable β, $Z = (Z_1, \cdots, Z_n)'$, and $\lambda = (\lambda_1, \cdots, \lambda_n)'$ given the data $Y = (Y_1, \cdots, Y_n)'$ of the model $M_{(\nu)}$ is given by

$$\begin{aligned} p(Z, \lambda, \beta|Y, M_{(\nu)}) \propto & \prod_{i=1}^{n} [\{I(Z_i > 0)I(Y_i = 1) + I(Z_i \leq 0)I(Y_i = 0)\} \\ & \phi_1(Z_i; X_i'\beta, \lambda_i^{-1})\phi_q(\beta; \alpha, \Omega^{-1})c(\nu)\lambda_i^{\nu/2-1}e^{-\nu\lambda_i/2}], \end{aligned} \tag{3.3}$$

where $c(\nu) = [\Gamma(\nu/2)(2/\nu)^{\nu/2}]^{-1}$, and $\phi_p(\ ; \mu, \Sigma)$ is the $N_p(\mu, \Sigma)$ pdf.

Note that this joint distribution is complicated in the sense that it is difficult to normalize and directly sample from. But computation of respective marginal posterior distributions of Z_i's, λ_i's, and β using the Gibbs sampling algorithm requires only fully conditional distributions of them. The required fully conditional distributions are

$$\beta|Z, \lambda, Y, M_{(\nu)} \sim N_q(\hat{\beta}, (\Omega + X'D_\lambda X)^{-1}), \tag{3.4}$$

where $\hat{\beta} = (\Omega + X'D_\lambda X)^{-1}(\alpha\Omega + X'D_\lambda Z)$, $X = (X_1, \cdots, X_n)'$, and $D_\lambda = \text{Diagonal}(\lambda_1, \cdots, \lambda_n)$.

$Z_1, \cdots, Z_n$ are independent with

$$Z_i|\beta, \lambda, Y, M_{(\nu)} \sim N(X_i'\beta, \lambda_i^{-1})I(Z_i > 0) \text{ if } Y_i = 1, \qquad (3.5)$$
$$Z_i|\beta, \lambda, Y, M_{(\nu)} \sim N(X_i'\beta, \lambda_i^{-1})I(Z_i \leq 0) \text{ if } Y_i = 0,$$

where $N(X_i'\beta, \lambda_i^{-1})I(B)$ is the normal distribution truncated to the interval event B.

$\lambda_1, \cdots, \lambda_n$ are independent with

$$\lambda_i|\beta, Z, Y, M_{(\nu)} \sim \text{Gamma}\left(\frac{\nu+1}{2}, \frac{2}{\nu + (Z_i - X_i'\beta)^2}\right).$$

Derivation of (3.4) is given as follows: From (3.3), the posterior density of β given Z, λ is

$$p(\beta|Z, \lambda, Y, M_{(\nu)}) \propto \phi_q(\beta; \alpha, \Omega^{-1}) \prod_{i=1}^{n} \phi_1(Z_i; X_i'\beta, \lambda_i^{-1}). \qquad (3.6)$$

This conditional density is the usual posterior density for the weighted regression parameter in the normal linear model

$$Z = X\beta + \varepsilon,$$

where ε is distributed $N_n(0, D_\lambda^{-1})$. Using standard results(cf. Zellner, 1971) with the informative prior pdf $\phi_q(\beta; \alpha, \Omega^{-1})$, we have (3.4). Note that simulation from the truncated normal distributions in (3.5) can be easily conducted via the algorithm by Devroye(1986).

3.2. Computing Procedure

As given above, the Gibbs sample for the model $M_{(\nu)}$ is defined through the complete conditional posterior densities:

$$p(\beta|Z, \lambda, Y, M_{(\nu)});\ p(\lambda_i|Z, \beta, Y, M_{(\nu)});\ p(Z_i|\beta, \lambda, Y, M_{(\nu)});\ i = 1, \cdots, n. \qquad (3.7)$$

The objective is to compute the marginal likelihood $m(Y|M_{(\nu)})$ from the Gibbs output obtained from the above conditional posteriors.

Lemma 2. Given the complete conditional densities (3.7), the marginal likelihood of the model $M_{(\nu)}$ is given by

$$m(Y|M_{(\nu)}) = \frac{f(Y|\beta, M_{(\nu)})\pi(\beta|M_{(\nu)})}{p(\beta|Y, M_{(\nu)})}, \qquad (3.8)$$

where $\pi(\beta|M_{(\nu)})$ is the prior density of β and

$$p(\beta|Y, M_{(\nu)}) = \int p(\beta|Z, \lambda, Y, M_{(\nu)})p(Z, \lambda|Y, M_{(\nu)})dZd\lambda.$$

Proof. As defined in (2.1), $m(Y|M_{(\nu)}) = \int \pi(\beta|M_{(\nu)})f(Y|\beta, M_{(\nu)})d\beta$. Thus Bayes theorem, i.e.

$$p(\beta|Y, M_{(\nu)}) = f(Y|\beta, M_{(\nu)})\pi(\beta|M_{(\nu)})/m(Y|M_{(\nu)}),$$

gives the result.□

Note that (3.8) holds for any β in the regression parameter space. Let G replicated outputs from the Gibbs algorithm(cf. Gelfand and Smith, 1990) with (3.7) be given by $\{\beta^{(g)}, \lambda^{(g)}, Z^{(g)}\}$, where

$$\lambda^{(g)} = (\lambda_1^{(g)}, \cdots, \lambda_n^{(g)})', Z^{(g)} = (Z_1^{(g)}, \cdots, Z_n^{(g)})',\ g = 1, \cdots, G.$$

Then the posterior density, $p(\beta|Y, M_{(\nu)})$, is appropriately estimated by taking the ergodic average of the full conditional density with the posterior draws of (Z, λ), leading to the estimate

$$\hat{h}_\nu = G^{-1} \sum_{g=1}^{G} h_\nu^{(g)}, \tag{3.9}$$

where $\hat{h}_{(\nu)} = \hat{p}(\beta|Y, M_{(\nu)})$ and $h_\nu^{(g)} = p(\beta|Z^{(g)}, \lambda^{(g)}, Y, M_{(\nu)})$. By virtue of the ergodic theorem(cf. Tierney, 1994), it can be shown that, for any $\beta \in R^q$,

$$\hat{p}(\beta|Y, M_{(\nu)}) \rightarrow p(\beta|Y, M_{(\nu)}), \text{ as } G \rightarrow \infty,$$

almost surely. Thus, for given β at β^*, (3.8) and (3.9) lead to the desired estimated marginal likelihood

$$\hat{m}(Y|M_{(\nu)}) = \frac{f(Y|\beta^*, M_{(\nu)})\pi(\beta^*|M_{(\nu)})}{\hat{p}(\beta^*|Y, M_{(\nu)})}. \tag{3.10}$$

As for the selection of the point β^*, the choice of the point is not critical, because Lemma 2 holds for any β. However, for efficiency of estimation, it may be better to take β^* as posterior mean or posterior mode of β which can be easily obtained from the Gibbs output.

Lemma 3. Let $M_{(\nu 1)}, M_{(\nu 2)} \in M(t)$ be any two t link binary regression models to be compared. Then the simulation-based estimate of the Bayes factor($B_{\nu 1.\nu 2}$) in logarithm scale, for given β at $\beta^*_{\nu i}$, $i = 1, 2$, is

$$\begin{aligned} \ln \hat{B}_{\nu 1.\nu 2} &= \ln f(Y|\beta^*_{\nu 1}, M_{(\nu 1)}) - \ln f(Y|\beta^*_{(\nu 2)}, M_{(\nu 2)}) \\ &+ \ln \pi(\beta^*_{\nu 1}|M_{(\nu 1)}) - \ln \pi(\beta^*_{\nu 2}|M_{(\nu 2)}) - \ln \hat{h}^*_{\nu 1} + \ln \hat{h}^*_{\nu 2} \end{aligned} \tag{3.11}$$

with

$$\ln \hat{B}_{\nu 1.\nu 2} \rightarrow \ln B_{\nu 1.\nu 2} \text{ as } G \rightarrow \infty, \tag{3.12}$$

and

$$\text{var}(\ln \hat{B}_{\nu 1.\nu 2}) = \left(\frac{\partial \ln \hat{B}_{\nu 1.\nu 2}}{\partial \hat{h}^*}\right)' \text{var}(\hat{h}^*) \left(\frac{\partial \ln \hat{B}_{\nu 1.\nu 2}}{\partial \hat{h}^*}\right), \tag{3.13}$$

where the derivative vector consists of elements $-\hat{h}^{*-1}_{\nu 1}$ and $\hat{h}^{*-1}_{\nu 2}$,

$$\text{var}(\hat{h}^*) = G^{-1}\left[\Delta_0 + \sum_{s=1}^{\ell}\left(1 - \frac{s}{\ell+1}\right)(\Delta_s + \Delta_s')\right], \tag{3.14}$$

$$\Delta_s = G^{-1} \sum_{g=s+1}^{G} (h^{*(g)} - \hat{h}^*)(h^{*(g)} - \hat{h}^*)',$$

$\hat{h}^* = (\hat{h}^*_{\nu 1}, \hat{h}^*_{\nu 2})'$, $\hat{h}^*_{\nu i} = \hat{p}(\beta^*_{\nu i}|Y, M_{(\nu i)})$, $h^{*(g)} = (h^{*(g)}_{\nu 1}, h^{*(g)}_{\nu 2})'$,
$h^{*(g)}_{\nu i} = p(\beta^*_{\nu i}|Z^{(g)}, \lambda^{(g)}, Y, M_{(\nu i)})$, $i = 1, 2$, and ℓ is some constant at which the autocorrelation function of $h^{(g)}_{\nu i}$ tapers off.

Proof. The Bayes factor for the two models $M_{(\nu 1)}$ and $M_{(\nu 2)}$, i.e. $m(Y|M_{(\nu 1)}) / m(Y|M_{\nu 2})$, can be estimated by repeating the calculation, described in the derivation of (3.10), for both models. This yields

$$\hat{B}_{\nu 1.\nu 2} = \hat{m}(Y|M_{(\nu 1)})/\hat{m}(Y|M_{(\nu 2)}).$$

Taking logarithmic function on both sides of the equation, we have (3.11). Since h inherits the ergodicity of the Gibbs output, it follows from the ergodic theorem (Tierney, 1994) that (3.12) holds. Using the result of Chip(1995), we have (3.14). Moreover, since $\ln \hat{B}_{\nu 1.\nu 2}$ is a function differentiable at $\hat{h}^*$, its variance (3.13) is found by the delta method(cf. Tanner, 1993).□

The merits of logarithmic expression for the Bayes factor, (3.11), are noted: (i) The estimate does not suffer from any instability occurring from estimating the inverse value in (3.10). (ii) As shown in Lemma 3, the entire estimation error arising from simulation can be easily derived. An estimate of the Bayes factor will be obtained by computing anti-logarithm of (3.11).

When several models are to be compared, Lemma 3 would extend in a natural way.

Corollary 1. Suppose $M_{(\nu 1)}, \cdots, M_{(\nu_K)}$ are K t link binary response regression models to be compared. Then the best fitted model $M_{(\nu k)}$ supported by the data is chosen such that

$$\ln \hat{m}(Y|M_{(\nu k)}) = \text{Max}\{\ln \hat{m}(Y|M_{(\nu j)}); j = 1, \cdots, K\}. \tag{3.15}$$

The criterion in (3.15) may be used for the variable selection procedure for probit, logit, and t link models. Variance of $\ln \hat{m}(Y|M_{(\nu k)})$ can be easily obtained from the result of Lemma 3.

Theorem 1. Given a binary response data set, logarithm of the Bayes factor for the goodness-of-link test criterion for testing probit link model versus logit link model is approximately estimated by

$$\ln \hat{B}_{\nu,8} = \ln \hat{m}(Y|M_{(\nu)}) - \ln \hat{m}(Y|M_{(8)}), \quad \text{as } \nu \to \infty, \tag{3.16}$$

where $\hat{m}(Y|M_{(\nu)})$ and $\hat{m}(Y|M_{(8)})$ are defined by (3.10).

Proof. Lemma 1, Lemma 2, and Lemma 3 give the result.□

Theorem 1 leads to the Bayesian goodness-of-link test criterion: If $\ln \hat{B}_{\nu,8} \geq 0$, the data set is in favor of probit link model; otherwise it is fitted to logit model. Value of ν can be chosen so that it may achieve desired degree of accuracy in the approximation. We refer to Johnson and Kotz(1969) for the formula for deciding the degree of accuracy.

4. ILLUSTRATIVE EXAMPLES : Probit Link Versus Logit Link

4.1. Simulation Study

A simulation study is done to investigate the performance of the Bayesian goodness-of-link test suggested in the previous section. Main concern in our simulation study is to examine whether the suggested test safely chooses the best fitted model between probit and logit link regression models. Two models considered in this simulation study are

$$\begin{aligned} \text{Model I } \Phi^{-1}(p_i) &= \beta_0 + \beta_1 X_i, \quad i = 1, \ldots, n, \\ \text{Model II. logit}(p_i) &= \beta_0 + \beta_1 X_i, \quad i = 1; \ldots, n, \end{aligned}$$

where $p_i = Pr(Y_i = 1)$. Given the values of β_0 and β_1, each model generates the data set $\{Y_i, X_i\}$ of size n using p_i values generated from $U(0,1)$. Based upon the generated data set, we compute Bayes factor ($\hat{B}_{\nu,8}$ in Theorem 1) for testing probit link versus logit link and examine the efficiency of the criterion in the test.

Since we are interested in the goodness-of-link test(not in the estimation), uniform prior is placed on the regression parameter β, i.e. $(\beta_0, \beta_1)'$ are distributed uniform on R^2. Then estimation of the Bayes factor is done by the following steps. To initialize the Gibbs sampling procedure, the ordinary least squares estimate of β obtained from linear regression analysis of current data is used. The initial values of λ_i's are set to one. Replications of the iterative procedure thereafter proceed independently. An assessment of convergence of the procedure is made by using the pragmatic checks on stationarity outlined in Gelfand *et al.* (1990). In particular, a number of summary statistics from the replicated samples (first and second moments and selected percentiles) are monitored every 10 iterations for β, augmented by direct graphical comparison of successive marginal density reconstructions in the final stages once stationarity appears to be achieved.

This process (generating the data set and estimation of the Bayes factor via the Gibbs sampling of G=500 cycles with 30 iterations in each cycle (the Gibbs sequence with 30 iterations)) is repeated 200 times for each model. Table 1 notes the resulting percentages of correct decision attained by the suggested goodness-of-link test based upon the data set generated from true models (Model I and Model II) with various values of β_0, β_1, δ and n. Marginal log-likelihood and its standard deviation of each t_ν link model($\nu = 8, 90, 120$) are also tabulated.

4.2. An Empirical Example

We illustrate the goodness-of-link test on nodal involvement data analyzed by Collett(1991). The data set is listed in Table 2. Two models to be compared are probit and logit link models with covariates fitted by Collett(1991):

$$\Phi^{-1}(p_i) = \beta_0 + \beta_1 \log(X_{1i}) + \beta_2 X_{2i} + \beta_3 X_{3i} \qquad (4.1)$$

and

$$\text{logit}(p_i) = \beta_0 + \beta_1 \log(X_{1i}) + \beta_2 X_{2i} + \beta_3 X_{3i}, \qquad (4.2)$$

where $i = 1, \cdots, 53$. Here X_{1i} is level of serum scid phosphate, X_{2i} is the result of an X-ray examination, coded 0 if negative and 1 if positive, X_{3i} is the size of the tumor, coded 0 if small and 1 if large, and the binary outcome

Table 1. Marginal Log-likelihood, Its Standard Deviation(in parenthesis), and Percentages of Correct Decision via Bayes Factors $B_{120,8}$ and $B_{90,8}$

(β_0, β_1)	n	$\nu = 8$	$\nu = 90$	$\nu = 120$	$B_{90,8}$	$B_{120,8}$
		(True	Model I)			
(1, 2)	10	-5.3386	-3.8252	-3.8655	95.5	96.0
		(.6053)	(.2120)	(.2158)		
	20	-10.0092	-6.2019	-6.2317	97.5	98.0
		(.7717)	(.3054)	(.3382)		
	50	-29.1725	-19.2471	-19.4135	98.5	99.0
		(.7835)	(.5672)	(.5012)		
(0, 4)	10	-4.8423	-3.3836	-3.4691	90.5	89.5
		(.6166)	(.5172)	(.4144)		
	20	-8.8803	-5.8826	-5.9798	98.0	98.5
		(.5057)	(.2684)	(.2610)		
	50	-24.9700	-19.0324	-19.3364	99.5	100
		(.7297)	(.2449)	(.5214)		
(2, 1)	10	-5.1293	-3.7659	-3.7015	94.5	95.0
		(.6258)	(.3625)	(.3391)		
	20	-9.7632	-7.3221	-7.2190	96.5	97.0
		(.7121)	(.4325)	(.4306)		
	50	-26.5625	-19.8962	-19.8773	99.5	99.5
		(.7352)	(.5769)	(.4918)		
		(True	Model II)			
(1, 2)	10	-4.5167	-5.6745	-5.8231	91.5	94.0
		(.7089)	(.4796)	(.3424)		
	20	-7.2349	-8.1453	-8.5168	92.5	94.5
		(.8248)	(.7030)	(.4425)		
	50	-23.9223	-29.2651	-29.4008	98.5	99.0
		(.7835)	(.5672)	(.5012)		
(0, 4)	10	-3.5393	-4.1112	-3.9902	88.5	87.5
		(.4800)	(.3858)	(.6227)		
	20	-5.5740	-6.1728	-6.2386	94.0	96.5
		(.4445)	(.3224)	(.3340)		
	50	-18.5758	-19.6891	-19.9439	95.5	97.0
		(.9997)	(.8749)	(.5127)		
(2, 1)	10	-5.1432	-6.1934	-6.0170	91.5	93.0
		(.6651)	(.5657)	(.5169)		
	20	-9.1326	-9.8241	-10.1209	93.5	94.5
		(.7332)	(.7340)	(.6906)		
	50	-25.4125	-30.6282	-31.2173	95.0	96.5
		(.9823)	(.7980)	(.9163)		

Table 2. Nodal Involvement Data

case	Y	X_1	X_2	X_3	case	Y	X_1	X_2	X_3	case	Y	X_1	X_2	X_3
1	0	.48	0	0	2	0	.56	0	0	3	0	.50	0	0
4	0	.52	0	0	5	0	.50	0	0	6	0	.49	0	0
7	0	.46	1	0	8	0	.62	1	0	9	1	.56	0	0
10	0	.55	1	0	11	0	.62	0	0	12	0	.71	0	0
13	0	.65	0	0	14	1	.67	1	0	15	0	.47	0	0
16	0	.49	0	0	17	0	.50	0	0	18	0	.78	0	0
19	0	.83	0	0	20	0	.98	0	0	21	0	.52	0	0
22	0	.75	0	0	23	1	.99	0	0	24	0	1.87	0	0
25	1	1.36	1	0	26	1	.82	0	0	27	0	.40	0	1
28	0	.50	0	1	29	0	.50	0	1	30	0	.40	0	1
31	0	.55	0	1	32	0	.59	0	1	33	1	.48	1	1
34	1	.51	1	1	35	1	.49	0	1	36	0	.48	0	1
37	0	.63	1	1	38	0	1.01	0	1	39	0	.76	0	1
40	0	.95	0	1	41	0	.66	0	1	42	1	.84	1	1
43	1	.81	1	1	44	1	.76	1	1	45	1	.70	0	1
46	1	.78	1	1	47	1	.70	0	1	48	1	.67	0	1
49	1	.82	0	1	50	1	.67	0	1	51	1	.72	1	1
52	1	.89	1	1	53	1	1.26	1	1	-	-	-	-	-

observed is the occurrence or nonoccurrence of cancer surrounding lymph nodes.

Under the assumption of uniform prior for regression parameters, we undertake the Gibbs sampler for G=5000 cycles with Gibbs sequence 30 after deleting the first 500, and the estimates β_ν^*, $\nu = 8, 120$, are obtained. Logarithm of the Bayes factor for $M_{(120)}$(probit model) versus $M_{(8)}$(logit model) is 8.6991 supporting probit model for the binary response data. The Gibbs sampling yields the following estimated models for (4.1) and (4.2):

$$\Phi^{-1}(\hat{p}_i) = -1.0020 + 2.8276 \log(X_{1i}) + 2.2451 X_{2i} + 1.8457 X_{3i} \tag{4.3}$$

and

$$\text{logit}(\hat{p}_i) = -5.6444 + 8.292 \log(X_{1i}) + 11.0375 X_{2i} + 7.2695 X_{3i}. \tag{4.4}$$

A plot of the standardized deviance residuals against the corresponding observation number, known as the index plot(cf. Collett, 1991), is given in Figure 1 and Figure 2 to show that the probit model is adequate for the data.

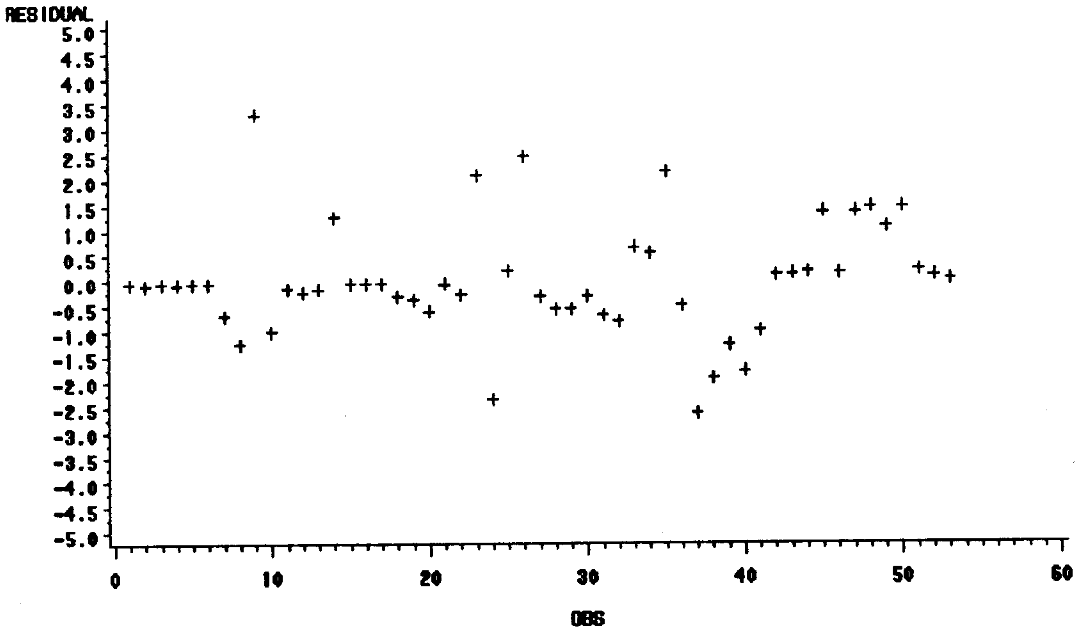

Figure 1. Index plot of the standardized deviance residuals on fitting (4.3) to the data

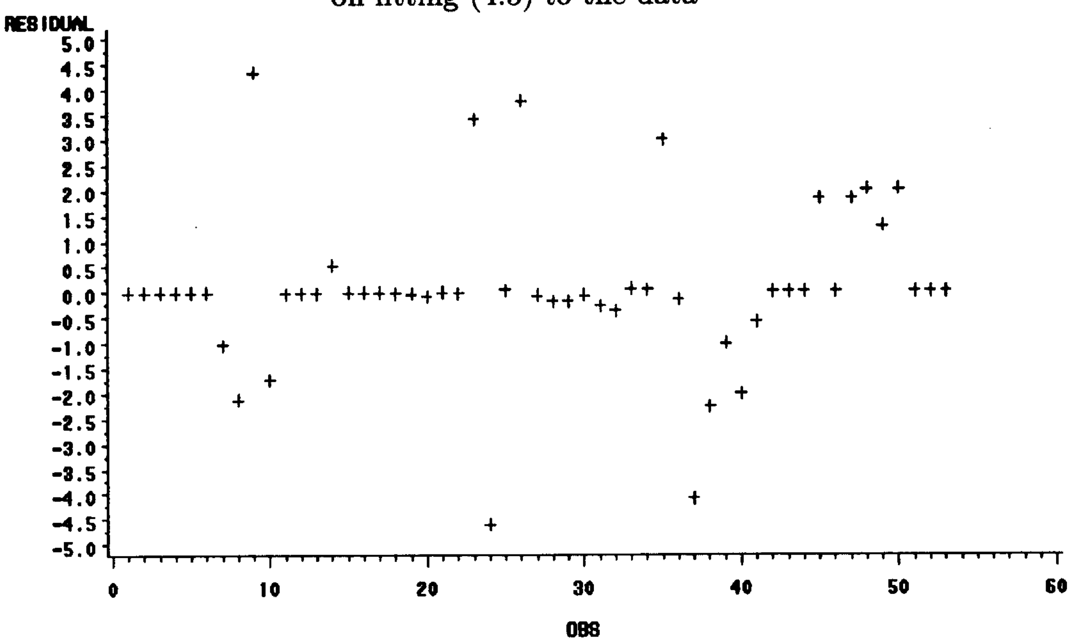

Figure 2. Index plot of the standardized deviance residuals on fitting (4.4) to the data

5. CONCLUDING REMARKS

We have considered the problem inherent in modeling an optimal binary response regression model with unknown link function. Our consideration in this paper concerns with a goodness-of-link test for probit link model versus logit link model. As a Bayesian criterion for the test, an approximate Bayes factor is proposed. It is derived by the well known facts that cdf of logistic and standard normal variates approximately equal to those of $t_8/.634$ and t_ν for a large ν, respectively. Computing method for the Bayes factor is also proposed by means of a synthesis of Gibbs sampling and marginal likelihood estimation scheme. The numerical studies in Section 4 show that the proposed test yields favorable goodness-of-link test results without regard to the sample sizes considered.

The proposed test has a number of advantages. First, it allows one to perform the desired test simply by a Bayes factor designed for t link binary response regression models, where it may be difficult to evaluate the exact Bayes factor for probit model versus logit model. Second, the proposed test will be preferable to the test by sampling approach(even though it has not been proposed) for small samples. Finally, computing the test criterion using the Gibbs sampling requires simulation mainly from standard distributions such as the normal distribution and the gamma distribution and, therefore, is easy to implement using many statistical packages.

The model selection criterion suggested in Corollary 1 is easily applicable to the variable selection procedure for probit model, logit model, and the t link binary response regression model. Especially, the criterion allows one to choose the best model among non-nested models(owing to variable transformations), which might be a merit of the criterion over the usual model selection criteria such as AIC and SBC. A study pertaining to the variable selection using the criterion is worthy of carrying out and is left for future study.

REFERENCES

(1) Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polytomous response data, *Journal of the American Statistical Association*, **88**, 669-679.

(2) Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, 68, 357-363.

(3) Chip, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313-1321.

(4) Collett, D. (1991). *Modeling Binary Data*, Chapman and Hall, London.

(5) Dellaportas, P. and Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazard models via Gibbs sampling, *Applied Statistics*, **42**, 443-459.

(6) Devroy, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

(7) Finney, D. J. (1971). *Probit Analysis*, 3rd ed., Cambridge University Press, Cambridge.

(8) Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotic and exact calculations, *Journal of the Royal Statistical Society*, Ser. B, **56**, 363-375.

(9) Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *Journal of the American Statistical Association*, **85**, 972-985.

(10) Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

(11) Griffiths, W. E., Hill, R. C., and Pope, P. J. (1987). Small sample properties of probit model estimators, *Journal of the American Statistical Association*, **82**, 929-937.

(12) Jeffreys, H. (1961). *Theory of Probability*, 3rd ed., Oxford University Press.

(13) Nelder, J. A. and McCullagh, P. (1989). *Generalized Linear Models*, Chapman and Hall, New York.

(14) Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood Bootstrap, *Journal of the Royal Statistical Society*, Ser. B, **56**, 3-48.

(15) Newton, M. A., Czado, C, and Chappell, R. (1996). Bayesian inference for semiparametric binary regression, *Journal of the American Statistical Association*, **91**, 142-153.

(16) Press, S. J. (1982). *Applied Multivariate Analysis*, 2nd ed., Krieger Publishing Company, Malabar, Florida.

(17) Soofi, S. S., Ebrahimi, N., and Habibullah, M. (1995). Information distinguishability with application to analysis of failure data, *Journal of the American Statistical Association*, **90**, 657-668.

(18) Tanner, M. A. (1993). *Tools for Statistical Inference*, 2nd ed., Springer Verlag, New York.

(19) Tanner, M. A. and Wong, W. (1987). The calculation of posterior distribution by data augmentation, *Journal of the American Statistical Association*, **82**, 528-550.

(20) Tierney, L. (1994). Markov.chains for exploring posterior distributions, *The Annals of Statistics*, **22**, 1701-1762.

(21) Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York.