The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 357-366

Goodness of Link Tests for Binary Response Data

In-Kwon Yeo[1)]

Abstract

The present paper develops a method to check the propriety of link functions for binary data. In order to parameterize a certain type of goodness of the link, a family of link functions indexed by a shape parameter is proposed. I first investigate the maximum likelihood estimation of the shape parameter as well as regression parameters and then derive their large sample behaviors of the estimators. A score test is considered to evaluate the goodness of the current link function. For illustration, I employ two families of power transformations, the modulus transformation by John and Draper (1980) and the extended power transformation by Yeo and Johnson (2000), which are appropriate to detect symmetric and asymmetric inadequacy of the selected link function, respectively.

Keywords : Fisher information matrix; Power transformation; Score test.

1. Introduction

Suppose that $Y_1, \ldots, Y_n$ are independent Bernoulli random variables with success probability $\mu_i = P(Y_i = 1)$. In generalized linear models, it is assumed that a monotone and differentiable function of μ_i satisfies a linear model

$$g(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta},$$

where $g(\cdot)$ is called the link function. The link function describes the dependence of μ_i on the linear predictor $\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$ for a particular regressor $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown parameter vector to be estimated.

To ensure that μ is restricted to the interval [0,1], a cumulative probability distribution

$$\mu = g^{-1}(\eta) = \int_{-\infty}^{\eta} f(s)\, ds,$$

1) Visiting Professor, Department of Control and Instrumentation Engineering, Kangwon National University, Chunchon, 200-701, Korea.
E-mail : inkwon@multimedia.kangwon.ac.kr

is usually employed. The density function $f(s)$ is called the tolerance distribution. When the tolerance is logistic distribution, the corresponding link is logit, $g(\mu)=\log(\mu/(1-\mu))$. The probit link $g(\mu)=\Phi^{-1}(\mu)$ and the complementary log-log link $g(\mu)=\log(-\log(1-\mu))$ are respectively derived from normal and extreme value tolerance distribution. Although the determination of the link function is an important process in modelling, see Wedderburn (1974), the choice of link is usually a matter of taste. Hence, the adequacy of the selected link cannot be taken for granted and should, if possible, be checked.

There are basically two ways of testing the goodness of the link for binary data. One is to extend the link function $g(\mu)$ to a parameterized form $g(\mu,\lambda)$ which includes some standard links such as logit, probit, and complementary log-log for a specific value of λ and then to make inferences about the value of λ. Several authors including Pregibon (1980), Aranda-Ordaz (1981), Guerrero and Johnson (1982), and Stukel (1988) introduce a generalized links and then choose the best link from the class generated by generalizations. A second procedure merely expands the link in a Taylor series in the linear predictor η and leads to the inclusion of a term in $\hat{\eta}^2$ in the linear predictor. A score test is then calculated.

In the literature about generalized link functions, the link to be tested should be included in its own link class. This article provide a method of testing the propriety of an arbitrary link function which there is good grounds for selecting or not. As usual ways, I first consider a class of link functions indexed by a single shape parameter, study the estimation of the shape parameter as well as regression parameters, and investigate the asymptotic properties of the estimators. In section 3, a score test and a bootstrap test are considered to examine the adequacy of the current link function. When the selected link functions is not proper to fit the data, the class may provide an alternative link. For illustration, two well-known data sets are employed in section 4.

2. GENERALIZED LINK FUNCTIONS AND ESTIMATION

Let $g(\cdot)$ be the selected link function to fit the data. Although the link function is carefully chosen, the data may often indicate that the selected link is unsuitable. The general approach to detect unsuitableness is to use the tests based on the deviance statistic or the Pearson chi-squared statistic. However, these tests are generally not very powerful even though the methods are quite popular. In order to find more powerful indication, I first consider a family of generalized link functions. The proposed class of link functions takes the form

$$\phi(\lambda, g(\mu))=\eta=\boldsymbol{x}^T\boldsymbol{\beta} \quad \text{and} \quad \mu=g^{-1}(\phi^{-1}(\lambda,\eta)). \tag{1}$$

I write $h=g^{-1}$ and $\psi=\phi^{-1}$ for convenient notation. Throughout this paper, I assume that

$\psi(\lambda, \eta)$ has following properties:

[P1] $\psi(\lambda, \eta)$ is strictly increasing nonlinear function of η

[P2] $\psi(\lambda, \eta)$ is continuously differentiable in λ and η

[P3] for a specific value $\lambda = \lambda_0$, $\psi(\lambda_0, \eta) = \eta$.

John and Draper(1980), Stukel (1988) and Yeo and Johnson (2000) introduce transformations satisfying properties above.

To satisfy the usual identifiability constraints, I now assume that if $\psi(\cdot, \cdot)$ is the true link function, then there exists a unique set of parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+1})^T = (\lambda, \beta_1, \ldots, \beta_p)^T$ such that (1) holds for all $\boldsymbol{x}$ in the covariate space. Then, the log-likelihood functions for n observations is

$$l_n(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n}\{y_i \log(\mu_i(\boldsymbol{\theta})) + (1-y_i)\log(1-\mu_i(\boldsymbol{\theta}))\}$$

$$= \sum_{i=1}^{n}\{y_i \log(h(\psi(\lambda, \boldsymbol{x}_i^T\boldsymbol{\beta}))) + (1-y_i)\log((1-h(\phi(\lambda, \boldsymbol{x}_i^T\boldsymbol{\beta})))\}.$$

Let $\nabla l_n(\boldsymbol{\theta}_*|\boldsymbol{y})$ be the gradient of the log-likelihood function evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_*$, and let $\nabla^2 l_n(\boldsymbol{\theta}_*|\boldsymbol{y})$ the observed information matrix at $\boldsymbol{\theta} = \boldsymbol{\theta}_*$. Then the score function $S_k(\boldsymbol{\theta}|\boldsymbol{y})$ for θ_k is the k-th element of $\nabla l_n(\boldsymbol{\theta}|\boldsymbol{y})$ and the maximum likelihood estimate (MLE) is obtained by solving the iterative equation

$$\widehat{\boldsymbol{\theta}_{(m+1)}} = \widehat{\boldsymbol{\theta}_{(m)}} - [\nabla^2 l_n(\widehat{\boldsymbol{\theta}_{(m)}}|\boldsymbol{y})]^{-1}\nabla l_n(\widehat{\boldsymbol{\theta}_{(m)}}|\boldsymbol{y}),$$

where $\widehat{\boldsymbol{\theta}_{(m)}}$ is the m-th iterative MLE. Replacing $-\nabla^2 l_n(\boldsymbol{\theta}|\boldsymbol{y})$ by the Fisher information matrix $\boldsymbol{i}_n(\boldsymbol{\theta}) = E[\nabla l_n(\boldsymbol{\theta}|\boldsymbol{Y})(\nabla l_n(\boldsymbol{\theta}|\boldsymbol{Y}))^T] = -E[\nabla^2 l_n(\boldsymbol{\theta}|\boldsymbol{Y})]$, I may also obtain the MLE by the method of scoring. Let $\nabla^2_{k,l} l_n(\boldsymbol{\theta}|\boldsymbol{y})$ be the (k,l)-th entry of $\nabla^2 l_n(\boldsymbol{\theta}|\boldsymbol{y})$ and $\mu_i^{(k)}(\boldsymbol{\theta}) = \partial\mu_i(\boldsymbol{\theta})/\partial\theta_k = \partial h(\psi(\lambda, \boldsymbol{x}_i^T\boldsymbol{\beta}))/\partial\theta_k$. Then I evaluate the (k,l)-th entry of $\boldsymbol{i}_n(\boldsymbol{\theta})$ as

$$\boldsymbol{i}_{k,l} = E[S_k(\boldsymbol{\theta}|\boldsymbol{Y})S_l(\boldsymbol{\theta}|\boldsymbol{Y})] = -E[\nabla^2_{k,l} l_n(\boldsymbol{\theta}|\boldsymbol{Y})] = \sum_{i=1}^{n}\frac{\mu_i^{(k)}(\boldsymbol{\theta})\mu_i^{(l)}(\boldsymbol{\theta})}{\mu_i(\boldsymbol{\theta})(1-\mu_i(\boldsymbol{\theta}))}. \quad (2)$$

Example 1. (Logit) Let $g(\mu) = \log(\mu/(1-\mu))$. Then, $\mu^{(k)}(\boldsymbol{\theta}) = \psi^{(k)}(\lambda, \eta)\mu(\boldsymbol{\theta})(1-\mu(\boldsymbol{\theta}))$ and the (k,l)-th entry of the Fisher information matrix $\boldsymbol{i}_n(\boldsymbol{\theta})$ is given as

$$\boldsymbol{i}_{k,l} = \sum_{i=1}^{n}\psi^{(k)}(\lambda, \eta_i)\psi^{(l)}(\lambda, \eta_i)\mu_i(\boldsymbol{\theta})(1-\mu_i(\boldsymbol{\theta})). \quad \square$$

Example 2. (Probit) Let $g(\mu) = \Phi^{-1}(\mu)$. Then, $\mu^{(k)}(\boldsymbol{\theta}) = \psi^{(k)}(\lambda, \eta)\phi(\psi(\lambda, \eta))$ and the (k,l)-th entry of $\boldsymbol{i}_n(\boldsymbol{\theta})$ is

$$i_{k,l}=\sum_{i=1}^{n}\frac{\psi^{(k)}(\lambda,\eta_i)\psi^{(l)}(\lambda,\eta_i)\phi(\psi(\lambda,\eta_i))\phi(\psi(\lambda,\eta_i))}{\Phi(\psi(\lambda,\eta_i))(1-\Phi(\psi(\lambda,\eta_i)))}. \quad \square$$

Example 3. (Complementary log-log) Assume that $g(\mu)=\log(-\log(1-\mu))$. Then, $\mu^{(k)}(\theta)=\psi^{(k)}(\lambda,\eta)\exp[\psi(\lambda,\eta)-\exp(\psi(\lambda,\eta))]$ and the (k,l)-th entry of $i_n(\theta)$ is

$$i_{k,l}=\sum_{i=1}^{n}\frac{\psi^{(k)}(\lambda,\eta_i)\psi^{(l)}(\lambda,\eta_i)\exp\{2[\psi(\lambda,\eta_i)-\exp(\psi(\lambda,\eta_i))]\}}{\exp(-\exp(\psi(\lambda,\eta_i))[1-\exp(-\exp(\psi(\lambda,\eta_i))]}. \quad \square$$

Suppose that the following conditions are satisfied:

[C1] the parameter space Θ is a compact subset of R^{p+1},

[C2] for all i, $x_i\in X$ where X is a compact subset of R^p.

Then, an application of Rubin's theorem (1956) with the continuity ψ ensures that for $\theta\in\Theta$,

$$\lim_{n\to\infty}\{\max n^{-1}l_n(\theta|Y)-\max n^{-1}E[l_n(\theta|Y)]\}=0$$

with probability one, where

$$E[l_n(\theta|Y)]=\sum_{i=1}^{n}\{\mu_i(\theta)\log(\mu_i(\theta))+(1-\mu_i(\theta))\log(1-\mu_i(\theta))\}.$$

Suppose that $l_0(\theta)=\lim_{n\to\infty}n^{-1}E[l_n(\theta|Y)]$ exists and $\theta_0=\arg\max l_0(\theta)$ for $\theta\in\Theta$. Then, the MLE $\hat{\theta}$ is a strongly consistent estimator of θ_0. Further conditions are required to show the asymptotic normality of $\hat{\theta}$:

[C3] θ_0 is an interior point of Θ,

[C4] both h and ψ with respect to θ are twice continuously differentiable,

[C5] $\lim_{n\to\infty}n^{-1}i_n(\theta_0)$ exists and is positive definite.

If conditions C1 - C5 hold, then

$$i_n^{1/2}(\hat{\theta}_*)(\hat{\theta}-\theta_0)\simeq N_{p+1}(0,I), \tag{3}$$

where $i_n^{1/2}$ is the square root matrix of i_n and $\hat{\theta}_*=a_n\hat{\theta}+(1-a_n)\theta_0$ for $a_n\in[0,1]$.

3. THE GOODNESS OF LINK TESTS

Note that, for a specific value $\lambda=\lambda_0$, $\psi(\lambda_0,\eta)=\eta$. Hence, the adequacy of the current link $g(\cdot)$ can be represented by the hypothesis $H_0:\lambda=\lambda_0$. Although Wald test using (3) and the likelihood ratio test can be used to test the null hypothesis, I only focus on the score test

which is an appealing way to test λ of interest parameter with nuisance parameter β because of computational advantages. Alternative tests which detect global lack of fit of parameteric model for the link function can be found in, for instant White (1982), Su and Wei (1991) and Cheng and Wu (1994). Cox and Snell (1989) include techniques for testing the parameter without nuisances in binary response models.

Let $S_\beta(\theta|y)$ denote the $p\times 1$ vector with k-th element $S_{k+1}(\theta|y)$ and denote the partition of the scores $\nabla l_n(\theta|y)=(S_\lambda(\theta|y), S_\beta(\theta|y)^T)^T$. The partitioned submatrices of the information matrix for $S_\lambda(\theta|Y)$ and $S_\beta(\theta|Y)$ are given as

$$i_{\lambda\lambda}(\theta)=E[(S_\lambda(\theta|Y))^2],$$
$$i_{\lambda\beta}(\theta)=E[S_\lambda(\theta|Y)(S_\beta(\theta|Y))^T],$$
$$i_{\beta\beta}(\theta)=E[S_\beta(\theta|Y)(S_\beta(\theta|Y))^T].$$

Here, each entry of the submatrices can be evaluated with $E[S_k(\theta|Y)S_l(\theta|Y)]$ given in (2).

Let $\hat{\beta}_0$ be the MLE of β under $H_0: \lambda=\lambda_0$ and write $\hat{\theta}_0=(\lambda_0, \hat{\beta}_0^T)^T$. Under C1 - C5, $S_\lambda(\hat{\theta}_0|Y)$ is asymptotically normally distributed with mean 0 and variance

$$\Sigma(\hat{\theta}_0)=i_{\lambda\lambda}(\hat{\theta}_0)-i_{\lambda\beta}(\hat{\theta}_0)\,i_{\beta\beta}^{-1}(\hat{\theta}_0)\,i_{\lambda\beta}(\hat{\theta}_0)^T.$$

The existence of the score function in the neighborhood of $\lambda=\lambda_0$ is due to the continuity of the derivatives of the ψ in λ. The score statistic for testing H_0: $\lambda=\lambda_0$ with a nuisance β is given as

$$T^2=\frac{S_\lambda(\hat{\theta}_0|Y)^2}{\Sigma(\hat{\theta}_0)}.$$

The asymptotic null distribution of T^2 is χ^2 with one degree of freedom and the test based on T^2 is the asymptotically locally most powerful unbiased test (see Bhat and Nagnur, 1965). However, the exact distribution of score statistic for even moderately large samples may be quite different from χ^2 distribution. Moreover, the score test needs to estimate the variance $\Sigma(\theta_0)$. To address these problems, I suggest a bootstrap technique (see Yeo and Johnson, 1998).

4. EXAMPLES

In this section, I illustrate my method by two examples with following two families of power transformations;

$$\psi^{JD}(\lambda,\eta)=\begin{cases}((\eta+1)^{\lambda}-1)/\lambda, & \lambda\neq 0,\eta\geq 0,\\ \log(\eta+1), & \lambda=0,\eta\geq 0,\\ -((-\eta+1)^{\lambda}-1)/\lambda, & \lambda\neq 0,\eta<0,\\ -\log(-\eta+1), & \lambda=0,\eta<0,\end{cases}$$

$$\psi^{YJ}(\lambda,\eta)=\begin{cases}((\eta+1)^{\lambda}-1)/\lambda, & \lambda\neq 0,\eta\geq 0,\\ \log(\eta+1), & \lambda=0,\eta\geq 0,\\ -((-\eta+1)^{2-\lambda}-1)/(2-\lambda), & \lambda\neq 2,\eta<0,\\ -\log(-\eta+1), & \lambda=2,\eta<0.\end{cases}$$

The function ψ^{JD} is known as the modulus transformation introduced by John and Draper (1980) and ψ^{YJ} is the extended power transformation by Yeo and Johnson (2000). Note that both ψ^{JD} and ψ^{YJ} reduce to the identity function when $\lambda=1$. Hence, testing hypothesis H_0: $\lambda=1$ can represent the test of the goodness of the selected link. Since, as depicted in Figure 1, ψ^{JD} changes from convex to concave as η changes sign, it can be applied to detect the under-estimation on one tail and the over-estimation on the other tail. Antithetically, ψ^{YJ} is concave in η for $\lambda<1$ and convex for $\lambda>1$ so the corresponding curves are asymmetrically pulled down in the tails for $\lambda<1$ and pushed up fir $\lambda>1$. Thus, it might be used to detect either under-estimation or over-estimation in both tails.

Although this goodness of links test can be applied to any types of link functions, logit, probit and complementary log-log links are tested in following examples. S-plus on a IBM-PC was run to generate the bootstrap samples $\boldsymbol{Y}_b^*$ and the bootstrap estimators $\widehat{\boldsymbol{\theta}_0^*}$ were computed by glm function with the binomial family in S-Plus. For each example, the bootstrap critical values based on 500 samples were calculated at the 5% level of significance.

4.1 Mortality of adult beetle after exposure to CS_2

Data of adult beetle mortality after exposure to gaseous carbon disulphide, CS_2, reported by Bliss (1935) were investigated by Prentice (1976), Aranda-Ordaz (1981) and Stukel (1988) who proposed generalized models for binary response data. According to these works, analyses on an asymmetric scale for these data are suggested.

Table 1. Score Tests and Bootstrap Critical Values for Beetle Mortality Data.

	ψ	T^2(p-value)	$S_\lambda(\widehat{\boldsymbol{\theta}_0})$	Bootstrap critical values
Logit	ψ^{JD}	0.623 (0.430)	1.810	(-4.441, 4.135)
	ψ^{YJ}	8.038 (0.005)	17.758	(-12.489, 12.038)
Probit	ψ^{JD}	0.234 (0.628)	0.884	(-3.343, 3.384)

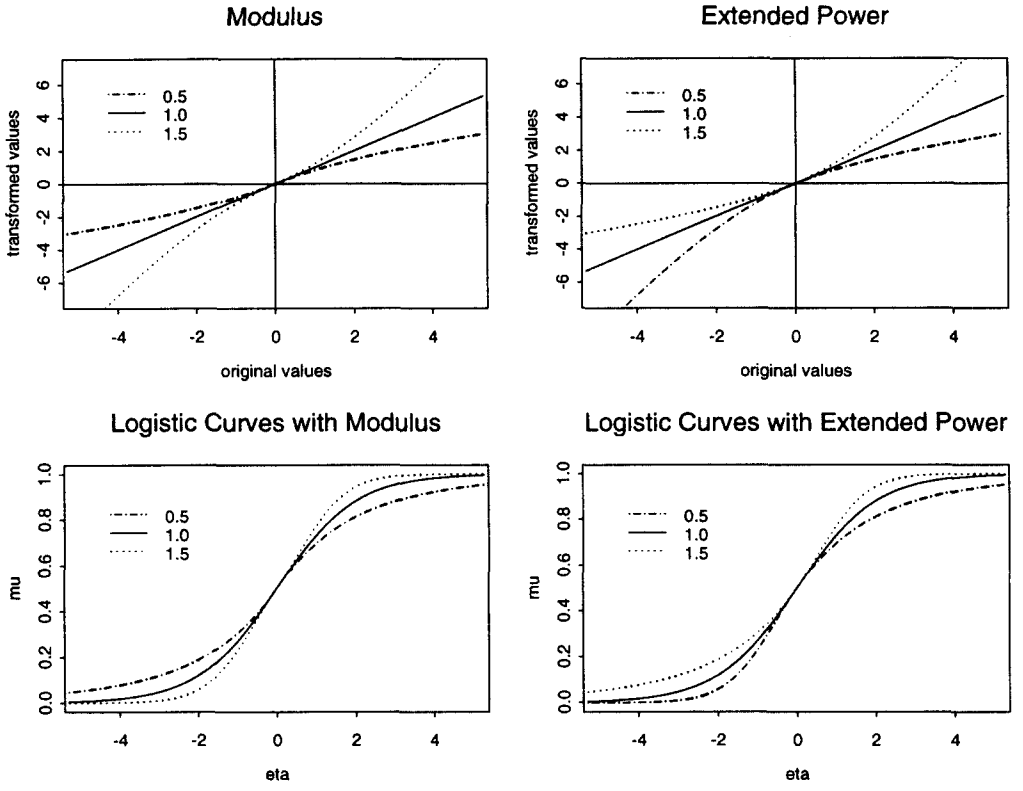

Figure 1. Plots of ψ^{JD} and ψ^{YJ} and the corresponding logistic curves for λ = 0.5 (broken line), 1.0 (solid), 1.5 (dotted).

	ψ^{YJ}	6.806 (0.009)	12.562	(-9.078, 8.896)
Cloglog	ψ^{JD}	0.833 (0.361)	-1.611	(-3.486, 2.999)
	ψ^{YJ}	0.018 (0.893)	0.634	(-8.964, 8.965)

Table 1 shows that for logit and probit, neither the score test nor the bootstrap test by ψ^{YJ} retains the null hypothesis $\lambda = 1$. However, both the score and the bootstrap test for the complementary log-log link function do not give the evidence of lack of fit, when ψ^{JD} and ψ^{YJ} are used. This confirms Aranda-Ordaz's (1981) result that the complementary log-log link is the most appropriate for these data. In fact, Yeo and Johnson (1998) show that the generalized logistic link using ψ^{YJ} with $\lambda = 1.48$ provides the better fit than the complementary log-log link.

4.2 Age of menarche in Warsaw girls

Milicer and Szczotka (1966) analyzed data determining the age of menarche of a sample of 3918 Warsaw girls in 1965. Aranda-Oradaz (1981) claimed that for these data the probit scale is, on practical grounds, the best choice within the class suggested by him. Sometimes a logarithmic transformation of the explanatory variable leads to proper description of data with a symmetric tolerance distribution. Thus, I also consider the models which use the logarithmic scales of age instead of the original values of age. The data are found in Aranda-Oradaz (1981) and Stukel (1988).

Table 2. Score Tests and Bootstrap Critical Values for Age at Menarche Data.

	ψ	T^2(p-value)	$S_\lambda(\widehat{\theta_0})$	Bootstrap critical values
Logit	ψ^{JD}	3.333 (0.068)	13.357	(-13.684, 13.837)
	ψ^{YJ}	4.282 (0.039)	-39.589	(-35.129, 33.243)
	$\psi^{JD}*$	2.862 (0.091)	12.692	(-16.021, 13.786)
	$\psi^{YJ}*$	2.287 (0.130)	-29.309	(-38.614, 39.740)
Probit	ψ^{JD}	0.064 (0.801)	-1.254	(-14.033, 13.458)
	ψ^{YJ}	7.995 (0.005)	-37.599	(-26.558, 27.527)
	$\psi^{JD}*$	0.034 (0.853)	0.949	(-17.123, 12.174)
	$\psi^{YJ}*$	1.094 (0.296)	-14.012	(-26.665, 27.025)
Cloglog	ψ^{JD}	21.347 (0.000)	31.290	(-24.176, 22.493)
	ψ^{YJ}	112.665 (0.000)	-146.734	(-30.756, 35.836)
	$\psi^{JD}*$	11.991 (0.001)	27.852	(-20.833, 19.737)
	$\psi^{YJ}*$	62.586 (0.000)	-116.275	(-29.833, 31.377)

* denotes that the logarithmic transformation of age is used.

Table 2 shows that none of link functions is appropriate to the original scale of age when ψ^{YJ} is employed. In this case, other links should be used to obtain the proper fit and the generalized link function, given in (1), with ψ^{YJ} may be applied. When logarithmic transformed ages are used, the probit is recommendable. Guerrero and Johnson (1982) obtained a remarkable improvement using a probit model where the Box-Cox transformation is taken to age and a normal distribution is assumed for transformed age.

5. CONCLUSIONS

The choice of the link function in the analysis of binary data is usually a matter of taste. In this paper, a generalized link function for testing the propriety of the selected link function has been proposed. The advantage of the proposed method is that an arbitrary link function can be tested, while other generalizations tend to restrict the link function to be tested to special cases such as the logit, probit and complementary log-log link function. For a link function $g(\mu)$, the proposed model takes the form $\phi(\lambda, g(\mu)) = \eta$ where ϕ is a nonlinear function and reduces to $\phi(\lambda_0, g(\mu)) = g(\mu)$ for a specific value $\lambda = \lambda_0$. This implies that a nonlinear relationship between the link function and the linear predictor is established and then, by testing $H_0 : \lambda = \lambda_0$, that is, linearity dependency on the predictor the adequacy of the link function g is checked. Due to computational advantages, the score test and the bootstrap test based on the score function are calculated. I recommend two families of power transformations, ψ^{JD} and ψ^{YJ}, which are appropriate to detect the bias of the selected link function, when they are employed in parallel, as well as satisfy the properties [P1]-[P3].

References

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

[2] Bhat, B. R. and Nagnur, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses, *Biometrika*, vol. 52, 459-468.

[3] Bliss, C. J. (1935). The calculation of the dosage-mortality curve, *Annals of Applied Biology*, vol. 22, 134-167.

[4] Cheng, K. F. and Wu, J. W. (1994). Testing goodness of fit for a parametric family of link functions, *Journal of the American Statistical Association*, vol. 89, 657-664.

[5] Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*, 2nd Ed., Chapman and Hall, New York.

[6] Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box-Cox transformation with binary response models, *Biometrika*, vol 69, 309-314.

[7] John, J. A. and Draper, N. R. (1980). An alternative family of transformations, *Applied Statistics*, vol. 29, 190-197.

[8] Milicer, H. and Szczotka, F. (1966). Age at menarche in Warsaw girls in 1965, *Human Biology*, vol. 38, 199-203.

[9] Pregibon, D. (1980). Goodness of link tests for generalized linear models, *Applied Statistics*, vol. 29, 15-24.

[10] Prentice, R. L. (1976). Generalization of the probit and logit methods for dose response

curves, *Biometrics*, vol 32, 761-768.

[11] Rubin, H. (1956). Uniform convergence of random functions with applications to statistics, *Annals of Mathematical Statistics*, vol. 27, 200-203.

[12] Stukel, T.A. (1988). Generalized logistic models, *Journal of the American Statistical Association*, vol. 83, 426-431.

[13] Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model, *Journal of the American Statistical Association*. vol. 86, 420-426.

[14] Wedderburn, R. W. M. (1974). Qausi-likelihood functions, generalized linear models, and the Gaussian-Newton method, *Biometrika*, vol. 61, 439-447.

[15] White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, vol. 50, 1-25.

[16] Yeo, I. K. and Johnson, R.A. (1998). The generalized logistic models with transformations, *Journal of the Korean Statistical Society*, vol. 27, 495-506.

[17] Yeo, I. K. and Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika,* vol. 87, 954-959.