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

Goodness-of-Fit Test for Mean and Variance Functions†

Sin-Ho Jung and Kee-Won Lee [1]

ABSTRACT

Using regression methods based on quasi-likelihood equation, one only needs to specify the conditional mean and variance functions for the response variable in the analysis. In this paper, an omnibus lack-of-fit test is proposed to test the validity of these two functions. Our test is consistent against the alternative under which either the mean or the variance is not the one specified in the null hypothesis. The large-sample null distribution of our test statistic can be approximated through simulations. Extensive numerical studies are performed to demonstrate that the new test preserves the prescribed type I error probability. Power comparisons are conducted to show the advantage of the new proposal.

Key Words : Mean function; Pseudo-likelihood; Quasi-likelihood; Variance function.

---

†Sin-Ho Jung's study was supported by 1995 research grant from Hallym University.
[1]Department of Statistics, Hallym University, Chunchon, Kangwon 200-702, Korea.

1. INTRODUCTION

Because of its flexibility, the method of quasi-likelihood (see McCullagh and Nelder, 1989) for regression problems has become increasingly popular among data analysts. Using this approach, one only needs to specify the mean and variance functions for the response variable in the analysis. If the mean function is correctly specified, valid inferences about the regression coefficients can be obtained. However, efficient estimation for these parameters requires knowledge of the structure of the variance. Furthermore, oftentimes estimation of the variance function is of independent interest (see Carroll and Ruppert, 1988).

For generalized linear models, tests for the validity of the mean function of the response variables are available in the literature (see, for example, Pregibon, 1980, 1985; Tsiatis, 1980; Christensen, 1989; Cox, Koh, Wahba, and Yandell, 1988; Su and Wei, 1991). However, as far as we know, there is no general procedure available for testing the appropriateness of the variance function without specifying particular alternative hypotheses.

In this paper, we propose an omnibus test for testing the adequacy of the quasi-likelihood. Our test is consistent against the alternative under which either the mean or the variance function is not the one specified in the null hypothesis. The new procedure is asymptotically distribution-free in the sense that its large-sample properties do not depend on a specific probabilistic mechanism for the response variables. Moreover, if the global null hypothesis, hypothesis specifying mean and variance functions, is rejected, the proposed lack-of-fit test would provide further information about the appropriateness of the assumed mean function, which is the most crucial part in the quasi-likelihood. If the mean function seems adequate, one may need to modify the variance function.

The large-sample null distribution of our test statistic can be approximated through simulations. For actual sample sizes, the appropriateness of this approximation is carefully examined in this paper. Power comparisons with likelihood ratio type tests are also performed to show the advantage of the new test. For future studies, similar procedures with censored data will be considered.

2. INFERENCES ABOUT THE MEAN AND VARIANCE FUNCTIONS

Let Y be the response variable and X, a $p \times 1$ vector, be the corresponding covariate vector. Given $X = x$, the conditional mean of Y is denoted by $f(x, \beta)$, where β is some fixed unknown $p \times 1$ vector. The heteroscedasticity of the regression model can be expressed by the variance function $\sigma^2 g^2(x, \beta, \theta)$, where σ is an unknown scale parameter and θ is an unknown $r \times 1$ parameter. For example, the variance may be modeled as proportional to a power of the mean: $g(x, \beta, \theta) = (f(x, \beta))^\theta$ (see Davidian and Carroll, 1987).

Now, let $(Y_i, X_i), i = 1, \ldots, n$, be n independent copies of (Y, X). Conditional on the sequence of the observed covariate vectors $x_1, \ldots, x_n$, estimation of β, σ, and θ can be obtained based on the following quasi- and pseudo-likelihood equations $S_1(\beta, \theta) = 0$ and $S_2(\beta, \sigma, \theta) = 0$, respectively, where

$$S_1(\beta, \theta) = n^{-1/2} \sum_{i=1}^{n} \left(\frac{Y_i - f(x_i, \beta)}{g(x_i, \beta, \theta)}\right) h_i(\beta, \theta),$$

$$S_2(\beta, \sigma, \theta) = n^{-1/2} \sum_{i=1}^{n} \left\{\left(\frac{Y_i - f(x_i, \beta)}{g(x_i, \beta, \theta)}\right)^2 - \sigma^2\right\} \begin{pmatrix} 1 \\ r_i(\beta, \theta) \end{pmatrix},$$

$$h_i(\beta, \theta) = \frac{f_\beta(x_i, \beta)}{g(x_i, \beta, \theta)}, \; f_\beta(x_i, \beta) = \frac{\partial f(x_i, \beta)}{\partial \beta}, \; \text{and } r_i(\beta, \theta) = \frac{\partial}{\partial \theta} \log g(x_i, \beta, \theta).$$

Let the resulting estimators be denoted by $\hat{\beta}, \hat{\theta}$, and $\hat{\sigma}$. The asymptotic properties of these estimators are well-documented in McCullagh and Nelder (1989) and Carroll and Ruppert (1988). In particular, the following asymptotic expansions are quite useful for developing our lack-of-fit test in the next section:

$$\sqrt{n}(\hat{\beta} - \beta) \approx -A_{\beta\beta}^{-1} S_1(\beta, \theta), \tag{2.1}$$

and

$$\sqrt{n}\left(\begin{pmatrix} \hat{\beta} \\ \hat{\sigma} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \beta \\ \sigma \\ \theta \end{pmatrix}\right) \approx -A^{-1} S(\beta, \sigma, \theta), \tag{2.2}$$

where $S(\beta, \sigma, \theta) = \{S_1^T(\beta, \theta), S_2^T(\beta, \sigma, \theta)\}^T$,

$$A = \begin{pmatrix} A_{\beta\beta} & A_{\beta\sigma} & A_{\beta\theta} \\ A_{\sigma\beta} & A_{\sigma\sigma} & A_{\sigma\theta} \\ A_{\theta\beta} & A_{\theta\sigma} & A_{\theta\theta} \end{pmatrix}, \quad A_{\beta\beta} = -n^{-1} \sum_{i=1}^{n} h_i(\beta, \theta) h_i^T(\beta, \theta), \quad A_{\beta\sigma} = 0,$$

$$A_{\beta\theta}=0,\ \ A_{\sigma\beta}=-2n^{-1}\sigma^2\sum_{i=1}^{n}\{\frac{\partial}{\partial\beta}\log(g(x_i,\beta,\theta))\}^T,\ \ A_{\sigma\sigma}=-2\sigma,$$

$$A_{\sigma\theta}=-2n^{-1}\sigma^2\sum_{i=1}^{n}r_i^T(\beta,\theta),\ \ A_{\theta\beta}=-2n^{-1}\sigma^2\sum_{i=1}^{n}r_i(\beta,\theta)\{\frac{\partial}{\partial\beta}\log(g(x_i,\beta,\theta))\}^T,$$

$$A_{\theta\sigma}=\sigma^{-1}A_{\sigma\theta}^T,\ \text{and}\ A_{\theta\theta}=-2n^{-1}\sigma^2\sum_{i=1}^{n}r_i(\beta,\theta)r_i^T(\beta,\theta).$$

Note that for (2.1), if the variance function is misspecified, θ should be replaced by θ^* which is the limit of $\hat{\theta}$, as $n\to\infty$.

To show that a more efficient inference procedure can be obtained with correctly specified variance function in the quasi-likelihood, extensive simulations are performed. From (2.1), the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$ can be approximated by

$$A_{\hat{\beta}\hat{\beta}}^{-1}\{\sum S_i(\hat{\beta},\hat{\theta})S_i^T(\hat{\beta},\hat{\theta})\}A_{\hat{\beta}\hat{\beta}}^{-1},$$

where

$$A_{\hat{\beta}\hat{\beta}}=A_{\beta\beta}|_{\beta=\hat{\beta}}\ \text{and}\ S_i(\hat{\beta},\hat{\theta})=n^{-1/2}\{Y_i-f(x_i,\hat{\beta})\}\frac{h_i(\hat{\beta},\hat{\theta})}{g(x_i,\hat{\beta},\hat{\theta})}.$$

On the other hand, without any knowledge of the structure of the variance, one still can make inference about β in the following manner. Let

$$S_1^*(\beta)=n^{-1/2}\sum_{i=1}^{n}\{Y_i-f(x_i,\beta)\}f_\beta(x_i,\beta)$$

be the estimating equation for β. Let $\tilde{\beta}$ be the solution of $S_1^*(\beta)=0$. Then $\sqrt{n}(\tilde{\beta}-\beta)$ is asymptotically normally distributed with mean 0 and variance

$$\{f_\beta(x_i,\tilde{\beta})f_\beta(x_i,\tilde{\beta})^T/n\}^{-1}\{\sum S_i^*(\tilde{\beta})S_i^{*T}(\tilde{\beta})\}\{f_\beta(x_i,\tilde{\beta})f_\beta(x_i,\tilde{\beta})^T/n\}^{-1},$$

where $S_i^*(\tilde{\beta})=n^{-1/2}(Y_i-f(x_i,\tilde{\beta}))f_\beta(x_i,\tilde{\beta})$. In one of simulation studies, we generate 100 samples $\{Y_i,x_i;i=1,\ldots,100\}$, where for each sample, the covariates $x_i,i=1,\ldots,100$ are fixed and are taken from the first 100 concentrations of esterase in Esterase Count Data(Carroll and Ruppert, 1988). Each Y is generated from a normal distribution with mean x and variance $x^{2\theta}$ with $\theta=0.5$. Figure 1 is the histograms of standard deviations of estimators of β_1 from the above two different estimating procedures. We observe obvious improvement in efficiency in estimating β_1 by using correct variance function.

3. A TEST FOR LACK OF FIT

Let H_{10} be the hypothesis that the conditional mean of Y is $f(x,\beta)$, and let H_{20} be the hypothesis that the conditional variance of Y is $\sigma^2 g^2(x,\beta,\theta)$. Furthermore, let H_0 be the hypothesis that both H_{10} and H_{20} are true.

If one is only interested in testing H_{10}, i.e., testing the appropriateness of the mean function, then the lack-of-fit test proposed by Su and Wei (1991) based on the partial sums of residuals works well. Specifically, in our present setting, their test statistic is $G_1 = \sup|W_1(t)|$, where

$$W_1(t) = n^{-1/2}\sum_{i=1}^{n}\{Y_i - f(x_i,\hat{\beta})\}I(x_i \le t),$$

$I(\cdot)$ is the indicator function, the event $[x \le t]$ indicates that all the components of x are less than or equal to those of t, and the supremum is taken with respect to t in R^p. Under H_{10}, one would expect that the partial sum process $W_1(\cdot)$ fluctuates about 0. Thus, a large value of G_1 leads to a conclusion of the misspecification of the mean function. It is important to note that asymptotically the test based on G_1 is still valid even if the variance function is wrongly specified.

Now, we use the above idea to test the global null hypothesis H_0 concerning both mean and variance functions. Consider another multi-parameter stochastic process:

$$W_2(t) = n^{-1/2}\sum_{i=1}^{n}[\{Y_i - f(x_i,\hat{\beta})\}^2 - \hat{\sigma}^2 g^2(x_i,\hat{\beta},\hat{\theta})]I(x_i \le t).$$

Let G_2 be the supremum of $|W_2(t)|$ with respect to t. If H_0 is true, then both W_1 and W_2 are centered around 0. Therefore, a reasonable test statistic for testing H_0 is $G = \max\{G_1, G_2\}$. A large value of G suggests a rejection of H_0. Using a similar argument given in Su and Wei (1991), one can easily see that the test based on G is consistent. That is, if either mean or variance function is not the one given in H_0, the power of the test G goes to 1, as $n \to \infty$.

It seems difficult, if not impossible, to derive the large-sample null distribution of G analytically. However, this large sample null distribution can be easily obtained through simulations. Under H_0, $W_1(t)$ is asymptotically equivalent to $V_1(t) + \hat{\eta}_1^T(t;\beta)\sqrt{n}(\hat{\beta}-\beta)$, where $V_1(t) = n^{-1/2}\sum_{i=1}^{n} e_i(\beta)I(x_i \le t)$, $e_i(\beta) = Y_i - f(x_i,\beta)$, and $\hat{\eta}_1(t;\beta) = -n^{-1}\sum_{i=1}^{n} f_\beta(x_i,\beta)I(x_i \le t)$. Approximating $n^{1/2}(\hat{\beta}-\beta)$ by (2.1),

$$W_1(t) \approx V_1(t) - \hat{\eta}_1^T(t;\beta)A_{\beta\beta}^{-1}S_1(\beta,\theta). \tag{3.1}$$

Now, let $\{y_i; i \geq 1\}$ be the observed values of $\{Y_i; i \geq 1\}$. Also, let $\{Z_i; i \geq 1\}$ be a sequence of independent and identically distributed random variables from, say, $N(0,1)$. Replacing $e_i(\beta)$ in (3.1) with $(y_i - f(x_i,\beta))Z_i$, let the resulting process be denoted by $\hat{W}_1(t;\beta,\theta)$, which is

$$n^{-1/2}\sum_{i=1}^{n}\{y_i - f(x_i,\beta)\}Z_i\{I(x_i \leq t) - \hat{\eta}_1^T(t;\beta)A_{\beta\beta}^{-1}\frac{h_i(\beta,\theta)}{g(x_i,\beta,\theta)}\}.$$

Note that only Z's are random quantities in the above expression.

Using Taylor's series expansion and (2.2), $W_2(t)$ can be approximated by

$$V_2(t) - \hat{\eta}_2^T(t;\beta,\sigma,\theta)A^{-1}S(\beta,\sigma,\theta), \tag{3.2}$$

where

$$V_2(t) = n^{-1/2}\sum_{i=1}^{n}[\{Y_i - f(x_i,\beta)\}^2 - \sigma^2 g^2(x_i,\beta,\theta)]I(x_i \leq t),$$

$$\hat{\eta}_2(t;\beta,\sigma,\theta) = -2n^{-1}\begin{pmatrix} \sigma^2\sum_{i=1}^{n} g(x_i,\beta,\theta)g_\beta(x_i,\beta,\theta) \\ \sigma\sum_{i=1}^{n} g^2(x_i,\beta,\theta) \\ \sigma^2\sum_{i=1}^{n} g(x_i,\beta,\theta)g_\theta(x_i,\beta,\theta)\end{pmatrix} I(x_i \leq t),$$

$g_\beta = \frac{\partial}{\partial\beta}g$,and $g_\theta = \frac{\partial}{\partial\theta}g$. Let $\hat{W}_2(t;\beta,\sigma,\theta)$ be the process obtained by replacing $\{(Y_i - f(x_i,\beta))^2 - \sigma^2 g^2(x_i,\beta,\theta)\}$ and $\{Y_i - f(x_i,\beta)\}$ with $\{(y_i - f(x_i,\beta))^2 - \sigma^2 g^2(x_i,\beta,\theta)\}Z_i$ and $\{y_i - f(x_i,\beta)\}Z_i$ in (3.2), respectively. Let $W(t) = \{W_1(t), W_2(t)\}$ and

$$\hat{W}(t;\beta,\sigma,\theta) = \{\hat{W}_1(t;\beta,\theta), \hat{W}_2(t;\beta,\sigma,\theta)\}.$$

Let $\hat{G}_1 = \sup_{t\in R^p} \mid \hat{W}_1(t;\hat{\beta},\hat{\theta}) \mid$, $\hat{G}_2 = \sup_{t\in R^p} \mid \hat{W}_2(t;\hat{\beta},\hat{\sigma},\hat{\theta}) \mid$, and $\hat{G} = \max\{\hat{G}_1,\hat{G}_2\}$. We will show that G and $\hat{G}$ have the same asymptotic distribution. Without loss of any generality, let $\sigma = 1$ and let all the components of X be between 0 and 1. For some technical reasons, we split the covariate vector X into two parts X_1 and X_2, say, where X_1 consists of all the discrete covariates and X_2, a $q \times 1$ vector, consists of all the continuous components of X. Let $t^T = (t_1^T, t_2^T)$. Then $V(t_1,t_2) = \sum V(k,t_2)$, where the $\sum$ denotes summation over all possible outcomes k, generated from X_1, which are less than t_1 componentwise, $V = \{V_1, V_2\}$,

$$V_1(k,t_2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(\beta)I(x_{1i} = k, x_{2i} \leq t_2), \text{ and}$$

$$V_2(k,t_2)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{e_i^2(\beta)-g^2(x_i,\beta,\theta)\}I(x_{1i}=k,x_{2i}\le t_2).$$

Note that these $V(k,t_2)$'s are independent to each other and are in the space $[0,1]^q$. It follows that the process $W(t)=W(t_1,t_2)$ can be approximated by

$$\sum_{k\le t_1}\begin{pmatrix}V_1(k,t_2)\\V_2(k,t_2)\end{pmatrix}-\begin{pmatrix}\hat\eta_1^T(t;\beta)A_{\beta\beta}^{-1}S_1(\beta,\theta)\\\hat\eta_2^T(t;\beta,\theta)A^{-1}S(\beta,\theta)\end{pmatrix}.$$

Now, define $\hat V=\{\hat V_1,\hat V_2\}$ and $\hat S=\{\hat S_1^T,\hat S_2^T\}^T$, where

$$\hat V_1(k,t_2;\beta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_ig(x_i,\beta,\theta)I(x_{1i}=k,x_{2i}\le t_2),$$

$$\hat V_2(k,t_2;\beta,\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i^2-1)g^2(x_i,\beta,\theta)I(x_{1i}=k,x_{2i}\le t_2),$$

$$\hat S_1(\beta,\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_ih_i(\beta,\theta),$$

$$\hat S_2(\beta,\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i^2-1)\begin{pmatrix}1\\r_i(\beta,\theta)\end{pmatrix},$$

and Z's are independent of $\{Y_i,x_i;i=1,\ldots,n\}$. The idea of proving that W and $\hat W$ have the same asymptotic distribution is to show, for any finite collection of points $t_2^{(1)},\ldots,t_2^{(m)}$ in $[0,1]^q$, $\{S(\beta,\theta),V(k,t_2^{(j)});j=1,\ldots,m,k\in\mathcal{C}\}$ and $\{\hat S(\beta,\theta),\hat V(k,t_2^{(j)});\ j=1,\ldots,m,k\in\mathcal{C}\}$ have the same limiting distribution, where $\mathcal{C}$ is the collection of all possible outcomes generated from X_1. The next stage is to show both $V(k,\cdot)$ and $\hat V(k,\cdot)$ are tight for each k. This follows from Theorem 3 and the Remark given at the bottom of page 1665 in Bickel and Wichura (1971). Hence W and $\hat W$ have the same limiting distribution. Applying the continuous mapping theorem, we can show that G and $\hat G$ have the same asymptotic distribution.

Now, let D_k be the set which consists of all the kth components from the observed covariate vectors $x_1,\ldots,x_n$ and let $D=\prod_{k=1}^{p}D_k$. Then, $G_j=\sup_{t\in D}\mid W_j(t)\mid,j=1,2$. Therefore, there is no need of a complicated numerical method to compute the supremum. Suppose that g is the observed value of G. The p-value, $\Pr(G\ge g)$, of our test can be estimated based on $\hat G$ through simulations. To estimate $\Pr(\hat G\ge g)$, we generate random samples $\{Z_1,\ldots,Z_n\}$ from $N(0,1)$ and compute $\hat G$ repeatedly to obtain the empirical proportion that $\hat G\ge g$.

Naturally, if H_0 is rejected, one would like to know which one, the mean or the variance function, is misspecified. Because of the structure of G, we are able to perform a two-stage test. Let g_1 be the observed value for G_1 and let α be the nominal total Type I error probability. If the p-value based on G for testing H_0 is less than α, then we estimate the p-value $\Pr(G_1 \geq g_1)$ through simulations using $\Pr(\hat{G}_1 \geq g_1)$. If the second p-value is less than α, then we claim that the mean function is misspecified; otherwise, the assumption of the variance function is not adequate. It follows from the argument in Marcus, Peritz and Gabriel (1976) that this sequential multiple test procedure has the prescribed level α of significance (asymptotically) for any combination of true hypotheses.

We apply the above lack-of-fit test to the Esterase count data (see Carroll and Ruppert, 1988, p.48). Here, the response variable Y is the observed number of bindings and the covariate x is the concentration of esterase. Carroll and Ruppert (1988) analyze this data set with a mean function f being $\beta_0 + \beta_1 x$ and a variance function g being $\sigma^2(\beta_0 + \beta_1 x)^{2\theta}$. The point estimates for β_0, β_1, σ, and θ based on quasi- and pseudo-likelihood functions are $-38.02, 18.18, 0.25$, and 1.0, respectively.

Using our test G, the approximated p-value is 0.636 based on 1,000 random samples $\{Z_1, \ldots, Z_{108}\}$ generated from $N(0,1)$ in estimating the distribution of $\hat{G}$. This indicates that there is no gross inadequacies with the model. The graphic methods which exhibit heterogeneity of variance presented in Carroll and Ruppert (1988, p.48-p.50) also support this conclusion.

4. SIMULATION STUDIES

For actual sample sizes, it is important to know if it is adequate to use the estimated distribution of $\hat{G}$ obtained from simulations to approximate the null distribution of G. Extensive numerical studies are conducted to examine if the new test preserves the nominal Type I error probability. The results indicate that the above approximation is satisfactory for actual sample sizes $(50 - 100)$. For small prescribed Type I error probabilities, say, ≤ 0.05, our test tends to be conservative. For example, in one of the simulation studies, we generate 500 samples $\{Y_i, x_i; i = 1, \ldots, n\}$, where, for each sample, the covariates $x_i, i = 1, \ldots, n$, are fixed and are taken from the first n concentrations of esterase in Esterase Count Data (Carroll and Ruppert, 1988). Each response Y is generated from a normal distribution with mean x and variance $x^{2\theta}$. For each realized sample $\{y_i, x_i\}$, we estimate the p-value through the distribution of $\hat{G}$ by generating 500 samples $\{Z_i, i = 1, \ldots, n\}$ from $N(0, 1)$. In Table 1, empirical sizes of our test are reported with $n = 50$ and 100. For $n = 50$, the new proposal appears to be conservative, especially for small α.

The proposed test is consistent against a broad class of alternatives. However, it is important to know if it is powerful enough for practical use. If there is a particular alternative hypothesis one is interested in testing against, a likelihood ratio type test can be constructed. Naturally, if the true model is in the class of the models specified by the alternative hypothesis, the likelihood ratio test should be better than the proposed omnibus test. On the other hand, if the alternative hypothesis does not include the true model, the new test performs much better than the likelihood ratio one. We present two sets of results from our extensive numerical comparisons in Table 2a and 2b. For Table 2a, we assume that the response Y is from a normal distribution with mean x and variance $x^{2\theta}$, where x's are from the first n concentrations of esterase in Esterase Count Data. Here, the null hypothesis H_0 assumes that Y has mean $\beta_0 + \beta_1 x$ and variance σ^2, which is free of x. The alternative hypothesis is that Y has mean $\beta_0 + \beta_1 x$ and variance $\sigma^2(\beta_0 + \beta_1 x)^{2\theta}$. For cases with $n = 100$, our test performs reasonably well. For Table 2b, we assume that Y is from normal with mean x and variance $x^{2\theta}$ and the null hypothesis is still H_0. However, the alternative assumes that the mean is $\beta_0 + \beta_1 x + \beta_2 x^2$ and the variance is σ^2. For this case, the omnibus test is much better than the likelihood ratio tests.

5. REMARKS

The numerical method proposed here is constructed based on the partial sums of "mean and variance residuals" in a very natural way. From our extensive numerical studies, we find that the new proposal is sensitive to detect a misspecified mean or variance function if the assumed function and the true one do not intersect with each other too frequently in the domain of covariates. The likelihood ratio type tests are helpful during the process of model selection. On the other hand, the proposed omnibus lack-of-fit test is quite useful especially at the final stage of model building process.

Table 1. Empirical levels of the lack-of-fit test G

Nominal level	$n = 50$, θ = .0	.2	.4	.6	$n = 100$, θ = .0	.2	.4	.6
.01	.004	.000	.002	.002	.004	.004	.002	.004
.02	.010	.008	.010	.008	.012	.014	.014	.012
.03	.016	.014	.014	.010	.020	.020	.026	.020
.04	.022	.020	.018	.016	.026	.030	.030	.028
.05	.034	.030	.028	.022	.036	.042	.038	.040
.06	.046	.040	.038	.030	.044	.048	.048	.046
.07	.056	.048	.048	.042	.062	.066	.062	.060
.08	.072	.060	.062	.050	.074	.074	.076	.074
.09	.078	.060	.072	.056	.082	.086	.082	.084
.10	.086	.074	.076	.062	.090	.094	.094	.092

Table 2. Power comparison between the new test and likelihood ratio tests

2a. against an alternative which includes the true model

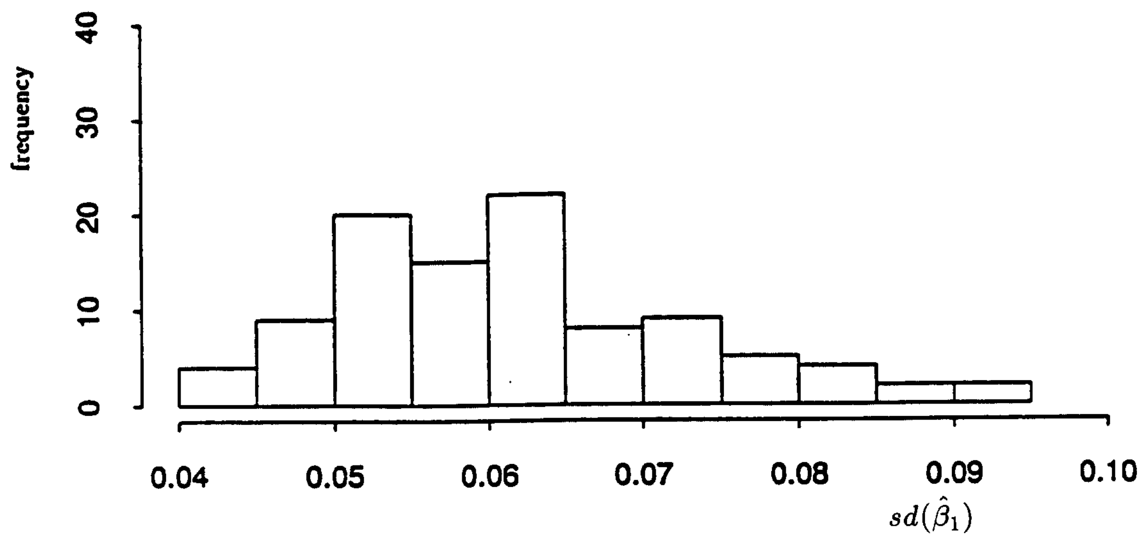
	$n = 50$, θ = .2	.4	.6	$n = 100$, θ = .2	.4	.6
New test	.046	.116	.150	.162	.418	.688
Likelihood ratio test	.112	.216	.388	.236	.754	.960

2b. against an alternative which does not include the true model ($n = 100$)

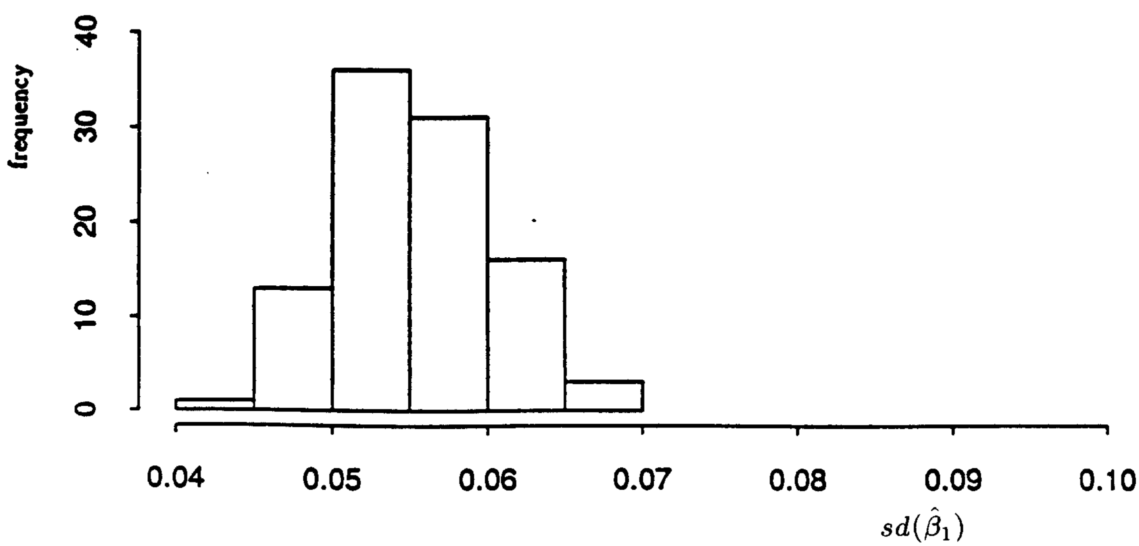
	θ = .2	.25	.3	.4	.5
New test	.162	.192	.290	.418	.588
Likelihood ratio test	.088	.064	.100	.110	.068

Figure 1. Histograms of standard deviations of $\hat{\beta}_1$

(a) Without considering the variance functions



(b) With considering the variance functions



REFERENCES

(1) Bickel, P.J. and Wichura, M.J. (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*, **42**, 1656-1670.

(2) Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* New York: Chapman & Hall.

(3) Christensen, R. (1989). Lack-of-fit Tests Based on Near or Exact Replication. *The Annals of Statistics.* **17**, 673-683.

(4) Cox, D. and Koh, E. (1986). A smoothing spline based test of model adequacy in polynomial regression, *Technical Report* 787, Department of Biostatistics, University of Wisconsin.

(5) Cox, D., Koh, E., Wahba, G., and Yandell, B.S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *The Annals of Statistics*, **16**, 113-119.

(6) Davidian, M. and Carroll, R.J. (1987). Variance Function Estimation. *Journal of the American Statistical Association*, **82**, 1079-91.

(7) Marcus, R., Peritz, E. and Gabriel, K.R. (1976). On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance. *Biometrika*, **63**, 655-60.

(8) McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* (2nd ed.), London: Chapman & Hall.

(9) Pregibon, D. (1980). Goodness of Link Test for Generalized Linear Models. *Applied Statistics.* **29**, 15-23.

(10) Pregibon, D. (1985). Link tests. in *Encyclopedia of Statistical Sciences* Vol.5, 82-85, ed. Kotz, S., Johnson, N.L. and Read, C.B., New York: Wiley.

(11) Su, J.Q. and Wei, L.J. (1991). A Lack of Fit Test for the Generalized Linear Model. *Journal of the American Statistical Association*, **86**, 420-26.

(12) Tsiatis, A.A. (1980). A Note on a Goodness-of-Fit Test for the Logistic Regression Without Replication. *Biometrika*, **67**, 250-251.