Statistica Sinica **8**(1998), 477-487

AN OPTIMAL TEST FOR THE MEAN FUNCTION HYPOTHESIS

K. F. Cheng and J. W. Wu

National Central University and Tamkang University

Abstract: The conditional mean of the response variable Y given the covariates $X = x \in R^p$ is usually modelled by a parametric function $g(\beta x)$, where $g(\cdot)$ is a known function and β is a row vector of p unknown parameters. In this paper, a new method for testing the goodness of fit of the model $g(\beta x)$ for the mean function is presented. The new test depends on the selection of weight functions. An expression for the efficacy of the proposed test under a sequence of local alternatives will be given. With the application of this result one can direct the choice of the optimal weight functions in order to maximize the efficacy. The new test is simple in computation and consistent against a broad class of alternatives. Asymptotically, the null distribution is independent of the underlying distribution of Y given $X = x$. Two pratical examples are given to illustrate the method. Further, simulation studies are given to show the advantages of the proposed test.

Key words and phrases: Consistency, estimation equation, goodness of fit, mean function, pitman efficiency.

1. Introduction

The regression analysis of independent observations with a quasi-likelihood model has been studied extensively (see McCullagh and Nelder (1989)). The quasi-likelihood approach has a number of attractive features that can be summarized as follows. First, one only needs to specify the models for the conditional mean and variance of the response variable Y given covariate values $X = (x_1, \ldots, x_p)$. Second, the approach provides consistent estimators for the regression parameters that only require that the model for the mean function is correctly specified. Thus, regardless of whether the "working" variance function is correctly specified, consistent estimators of the regression parameters are obtained. In addition, robust variance estimators that are consistent even when the working variance function is misspecified can easily be obtained (see Liang and Zeger (1986)). Extension of the quasi-likelihood approach for longitudinal responses was also considered by Liang and Zeger (1986). They proposed the generalized estimating equations approach.

From the above it is seen that a fundamental step in parametric regression analysis is to select a correct model for the mean function. Usually a functional form for the mean function is postulated that depends upon a $1 \times p$ parameter vector. Such a parametric approach can generally display great advantages in terms of interpretability and precision, and can be found in most introductory courses in regression analysis. This is because often the main interest of the regression analysis is to draw inferences about covariate effects on the marginal means of individual observations. This paper proposes a test for the mean function models. The test can be used to detect the adequacy of the parametric models, such as a linear model or a logit model, for the mean function.

Su and Wei (1991) also proposed a test procedure for testing the misspecification of the mean function model. They defined their test statistic using the supremum of the partial sum of residuals. Their test is designed to test the validity of the assumed model against very general alternatives. In contrast, our new procedure is designed to test the null hypothesis against particular alternatives. This can be done because the test statistic depends on the selection of weight functions. With proper choice of weight functions, one can achieve this objective. In fact, we shall derive the efficacies of the proposed test under sequences of local alternatives. One can apply this result to guide the choice of optimal weight functions in order to maximize the efficacy.

To test the adequacy of a particular hypothesized model for the means, one can also apply nonparametric regression method. Recent papers on this general topic include Cox, Koh, Wahba, and Yandell (1988), Staniswalis and Severini (1991), and Firth, Glosup and Hinkley (1991), etc.

Suppose one has additional assumptions about the conditional distributuion of the response variable given covariate values; then the statistics based on the deviance residuals can also be used to detect certain departure from a hypothesized model for the means (see McCullagh and Nelder (1989)). This likelihood-based approach, however, can no longer be applied without assumptions about the error distribution. The approach considered in the present paper is not a likelihood-based method. In section 2, the basic structure of the test procedure is described. Section 3 gives the efficacy of the test under local atternatives and derives optimal weight functions. Practical examples and simulation studies are given in Section 4. Conclusions are given in Section 5.

2. Construction of Test Statistics

Let $\{(Y_i, X_i), i = 1, \ldots, n\}$ denote a random sample, where the Y_i's are real-valued independent response variables and $X_i = (X_{i1}, \ldots, X_{ip})^T$ is a p-dimensional column vector of covariates. The expectation of $Y|X = x$ is denoted by $\mu(x)$. We assume the parametric model $g(\beta x)$ for $\mu(x)$ and thus the null hypothesis is $H_0 : \mu(x) = g(\beta x)$, for some p-dimensional row vector of regression coefficients β. To satisfy the usual indentifiability constraint under H_0, we also assume that there exists a unique set of parameter values β_0 such that the mean function $\mu(x) = g(\beta_0 x)$ for all x in the covariate space.

Under the hypothesized link function $g^{-1}(\cdot)$, the estimate $\hat{\beta}_c$ of β can be obtained by considering the solution of the estimating equations

$$\sum_{i=1}^{n}\{y_i - g(\beta x_i)\}\dot{C}\{g(\beta x_i)\}\dot{g}(\beta x_i)x_i = 0, \tag{1}$$

where $C(\cdot)$ is considered as a weight function. Given some regularity conditions, there exists a sequence of solutions $\hat{\beta}_c$ to (1) such that $\hat{\beta}_c$ converges in probability to a constant vector β_c satisfying

$$E\{\mu(X) - g(\beta_c X)\}\dot{C}\{g(\beta_c X)\}\dot{g}(\beta_c X)X = 0. \tag{2}$$

Suppose $\dot{C}(\cdot) = $ constant; then the system of equations (1) gives the least squares extimates. Further if $[\dot{C}\{g(\beta x)\}]^{-1} = \text{Var}\,(Y \mid X = x)$, then the system of equations (1) is exactly the system of the quasi-likelihood estimating equations (see Wedderburn (1974)). Wedderburn also showed that the quasi-likelihood estimates are the same as the maximum likelihood estimates based on the natural exponential family with variance function $C^{-1}(\mu)$, when such a family exists.

Under H_0 and provided that the solution to (2) is unique, we have $\beta_c = \beta_0$. Thus, for any different weight functions $C(\cdot)$ and $C_*(\cdot)$ such that $C(t)$ is not proportional to $C_*(t)$ the result implies that, for large sample size, the difference between $\hat{\beta}_c$ and $\hat{\beta}_{c*}$ should be small. Also, the random vector $H_n(\hat{\beta}_c) = n^{-1}\sum_{i=1}^{n}\{y_i - g(\hat{\beta}_c x_i)\}\dot{C}_*\{g(\hat{\beta}_c x_i)\}\dot{g}(\hat{\beta}_c x_i)x_i$ should be close to zero. On the other hand, under the alternative hypothesis, $H_n(\hat{\beta}_c)$ in general will converge in probability to a nonzero vector $H(\beta_c) = E\{\mu(X) - g(\beta_c X)\}\dot{C}_*\{g(\beta_c X)\}\dot{g}(\beta_c X)X$. Thus the test based on the statistic $H_n(\hat{\beta}_c)$ seems to have the potential to detect a broad range of model misspecifications.

The exact sampling distribution of $\sqrt{n}H_n(\hat{\beta}_c)$ is difficult to derive. However, large sample results can be used to approximate the sampling distribution of $\sqrt{n}H_n(\hat{\beta}_c)$. Under H_0 and assuming that certain regularity conditions are satisfiesed, one can show that $\sqrt{n}H_n(\hat{\beta}_c) \xrightarrow{d} \text{MVN}(0, \sum_c)$ as $n \to \infty$. Denote a consistent estimator of the covariance matrix by $\hat{\sum}_c$. Then the Wald chi-squared Statistic $W_n = nH_n^T(\hat{\beta}_c)\hat{\sum}_c^{-1}H_n(\hat{\beta}_c)$ can be used to test H_0. The null distribution of W_n converges weakly to a central chi-squared variable with p degrees of freedom, and a large value of W_n means that one should reject H_0. More precisely, to carry out the test, one simply computes W_n and compares it to the critical value of the χ^2_p distribution for a given size of the test.

Asymptotically, the goodness-of-fit test based on W_n has good properties. Particularly, the test W_n is consistent against any misspecification of the mean function model under which $H(\beta_c)$ is nonzero. However, for small sample sizes, the performance of W_n depends on the dimension p of X. In some special cases such as $p \leq 3$, the test gives quite satisfactory performance. But for $p > 3$, it may simply be impossible to test all indicators in $H_n(\hat{\beta}_c)$ jointly. Sometimes the test tends to reject the null hypothesis more often than it should when H_0 is true. An explanation of this undesirable behaviour of W_n is that most of the $\hat{\sum}_c$ matrices are nearly singular and lead to large simulated values of W_n. In such a situation, the χ^2_p distribution is not an appropriate approximation of the null distribution of W_n.

Based on $H_n(\hat{\beta}_c)$, another approach to test H_0 is to consider linear combinations of the components in $H_n(\hat{\beta}_c)$. In this paper we consider the test statistic $S_n(\hat{\beta}_c) = \hat{\beta}_c H_n(\hat{\beta}_c)$. The reasons are simple. First, the test based on $S_n(\hat{\beta}_c)$ is seen to be efficient under various situations discussed in Section 4. Further, it is easier to establish a rule for selecting optimal weight functions $C(\cdot)$ and $C_*(\cdot)$. The latter result will be derived in Section 3.

The large sample null distribution of $\sqrt{n}S_n(\hat{\beta}_c)$ is stated in the following. The proof is simple and hence will be omitted. We first assume the regularity conditions being satisfied so that for $n \to \infty$, one can show $\sqrt{n}(\hat{\beta}_c - \beta_0) \xrightarrow{d} \text{MVN}(0, \sum)$, where the asymptotic covariance matrix can be consistently estimated by $\hat{\sum} = A_n^{-1}(\frac{1}{n}\sum_{i=1}^n C_{ni}^{\otimes 2})A_n^{-1}$, where $A_n = n^{-1}\sum_{i=1}^n \dot{g}^2(z_i)\dot{C}\{g(z_i)\}x_i^{\otimes 2}$, $C_{ni} = \{y_i - g(z_i)\}\dot{C}\{g(z_i)\}\dot{g}(z_i)x_i$, $z_i = \hat{\beta}_c x_i$, and for any column vector a, $a^{\otimes 2}$ denotes the matrix aa^T. Further, by applying the delta method, one can establish the asymptotic result: $\sqrt{n}S_n(\hat{\beta}_c) \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$. Here, the variance can be consistently estimated by using the estimator

$$\hat{\sigma}^2 = B_n \hat{\sum} B_n^T + (n^{-1}\sum_{i=1}^n D_{ni}^2) - 2\{n^{-1}\sum_{i=1}^n (B_n A_n^{-1} C_{ni}) D_{ni}\},$$

where $B_n = n^{-1}\sum_{i=1}^n \dot{g}^2(z_i)\dot{C}_*\{g(z_i)\}z_i x_i^T$ and $D_{ni} = \{Y_i - g(z_i)\}\dot{C}_*\{g(z_i)\}\dot{g}(z_i)z_i$. Thus to carry out the test, one computes $Z^2 = nS_n^2(\hat{\beta}_c)/\hat{\sigma}^2$ and compares it to the critical value of the χ^2_1 distribution for a given size of the test.

Before closing this section, we exhibit a simple simulation study to compare the performance of W_n and Z^2. We consider the problem of testing a probit model $H_0 : g(\beta X) = \Phi(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$, where $\Phi(t)$ is the cdf of the $N(0, 1)$ distribution. Let the true model be $Y_i = Y_i^*/50$, and for each i, Y_i^* is from a binomial $(50, p(X_i))$ distribution, where $p(X_i) = \Phi(X_{i1} + X_{i2} + X_{i3} + X_{i4} + \gamma X_{i5})$, $(X_{i1}, X_{i2}, X_{i3} + X_{i4} + X_{i5})$ is from MVN$((0.5, 0.5, -0.1, -0.5, -0.5), .01\ I)$, and I is the usual identity matrix. In the study we choose $\dot{C}(t) = 1$ and $\dot{C}_*(t) = \{t(1 - t)\}^{-1}$ so that $\hat{\beta}_c$ is the usual least squares estimate and under H_0, $\dot{C}_*^{-1}\{g(\beta X)\} = 50\ \text{Var}\,(Y|X) = g(\beta X)\cdot\{1 - g(\beta X)\}$. Further, we note that a consistent estimator $\hat{\sum}_c$ of $\sum$ needed for computing W_n is given by $\hat{\sum}_c = \hat{\sum}_{11} + 2\hat{\sum}_{12} + \hat{\sum}_{22}$, where $\hat{\sum}_{11} = n^{-1}\sum_{i=1}^n \{A(\hat{\beta}_c, X_i, Y_i) - \bar{A}\}\{A(\hat{\beta}_c, X_i, Y_i) - \bar{A}\}^T$,

$$\hat{\sum}_{12} = n^{-1}\sum_{i=1}^n \Big[\{A(\hat{\beta}_c, X_i, Y_i) - \bar{A}\}B^T(\hat{\beta}_c, X_i, Y_i)\{-n^{-1}\sum_{i=1}^n \frac{\partial B(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\}^{-1}$$

$$\cdot\{n^{-1}\sum_{i=1}^{n}\frac{\partial A(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\}^T\Big],$$

$$\hat{\sum}_{22} = \Big\{n^{-1}\sum_{i=1}^{n}\frac{\partial A(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\Big\}^T\Big\{n^{-1}\sum_{i=1}^{n}\frac{\partial B(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\Big\}^{-1}$$
$$\cdot\Big\{n^{-1}\sum_{i=1}^{n}B(\hat{\beta}_c, X_i, Y_i)B^T(\hat{\beta}_c, X_i, Y_i)\Big\}\Big\{n^{-1}\sum_{i=1}^{n}\frac{\partial B(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\Big\}^{-1}$$
$$\cdot\Big\{n^{-1}\sum_{i=1}^{n}\frac{\partial A(\hat{\beta}_c, X_i, Y_i)}{\partial \beta}\Big\}^T,$$

$$A(\beta, X, Y) = \{Y - g(\beta X)\}\dot{g}(\beta X)\dot{C}_*\{g(\beta X)\}X,$$
$$B(\beta, X, Y) = \{Y - g(\beta X)\}\dot{g}(\beta X)\dot{C}\{g(\beta X)\}X,$$

and $\bar{A} = n^{-1}\sum_{i=1}^{n} A(\hat{\beta}_c, X_i, Y_i)$.

The estimated powers of the tests Z^2 and W_n at level $\alpha = .05$ are given in Table 1. All the empirical powers reported in the table were based on 500 replications of $\{(X_i, Y_i),\ i = 1, 2, \ldots, 50\}$. These results clearly show that the test based on Z^2 is definitely better than that based on W_n.

Table 1. Empirical powers of Z^2 and W_n tests

γ	Z^2	W_n
.00	.045	.064
.25	.230	.100
.50	.240	.130
.75	.250	.150
1.00	.250	.150
1.25	.260	.180

3. Efficacy and Optimal Weights

For the purpose of simplicity, we assume in this section that the conditional distribution of Y depends on x only throught βx. Thus under the local alternatives considered below, Var $(Y \mid x)$ depends on n and βx. Assume this variance function converges to $V(\beta x)$ as $n \to \infty$.

Consider a sequence of local alternatives with link functions $g_{0,n}^{-1}(t)$ such that, uniformly on t,

$$\sqrt{n}\{g_{0,n}(t) - g(t)\} \to h(t), n \to \infty, \tag{3}$$

where $h(t)$ is some function. Under such a sequence of local alternatives, and assuming that the proper regularity conditions are satisfied, one can show $\sqrt{n}S_n(\hat{\beta}_c)$

$\xrightarrow{d} N(\mu, \sigma^2)$ as $n \to \infty$. Here, taking $W = \beta_0 X$, we have

$$\mu = E[h(W)\dot{C}\{g(W)\}\dot{g}(W)W\sigma_*^2/\sigma_i^2] - E[h(W)\dot{C}_*\{g(W)\}\dot{g}(W)W],$$
$$\sigma^2 = E[\dot{C}_*\{g(W)\} - \dot{C}\{g(W)\}\sigma_*^2/\sigma_c^2]^2\dot{g}^2(W)W^2V(W),$$
$$\sigma_c^2 = E[\dot{C}\{g(W)\}\dot{g}^2(W)], \text{ and } \sigma_*^2 = E[\dot{C}_*\{g(W)\}\dot{g}^2(W)].$$

Appealing to the notion of Pitman efficiency, we find that the efficacy $e(S_n)$ of the test against the local alternatives satisfying expression (3) is given by

$$e(S_n) = \frac{[E[\dot{C}_*\{g(W)\} - \dot{C}\{g(W)\}\sigma_*^2/\sigma_c^2]h(W)\dot{g}(W)W]^2}{E[\dot{C}_*\{g(W)\} - \dot{C}\{g(W)\}\sigma_*^2/\sigma_c^2]^2\dot{g}^2(W)W^2V(W)}.$$

Note that the efficacy is invariant under a sclar multiplication of the weight function $C(\cdot)(C_*(\cdot))$. By the Cauchy-Schwartz inequality, $e(S_n) \leq E\{h^2(W)/V(W)\}$, which is independent of $C(\cdot)$ and $C_*(\cdot)$. The inequality is an equality if and only if

$$\dot{C}_*\{g(t)\}\dot{g}(t)t = \dot{C}\{g(t)\}\dot{g}(t)t\ \sigma_*^2/\sigma_c^2 + d \cdot h(t)/V(t). \tag{4}$$

for some nonzero constant d and all t in the space of $\beta_0 X$. Since the weight functions can be multiplied by some proper scalar constants so that $\sigma_c^2 = \sigma_*^2$; thus, for such weight functions, application of result (4) shows that S_n is the optimal test of H_0 against the local alternatives (3) with $h(t)$ satisfying $h(t) = d \cdot V(t)[\dot{C}_*\{g(t)\} - \dot{C}\{g(t)\}]\dot{g}(t)t$.

4. Examples and Simulations

4.1. Examples

We use two examples to illustrate the proposed method. Let $\hat{\beta}_c$ be defined as the least squares estimate (that is choose $\dot{C}(t) = 1$) and $\dot{C}_*\{g(t)\} = V^{-1}(t)$. Thus, the test is optimal for testing H_0 against local alternatives (3.1) with $h(t) = \{d - (d\sigma_*^2/\sigma_c^2)V(t)\}\dot{g}(t)t$ for some constant d. The first example is taken from Bissell (1972). There are 32 independent counts of the number y of flaws in rolls of fabric of length x. The data are given in Table 1 of Firth et al. (1991). They assumed that there is a constant flaw rate and thus the mean count $E(Y \mid X = x) = \beta x$. The least squares estimate is $\hat{\beta}_c = .0151$. Following Firth et al. (1991), we also assume Y to be a Poisson variate; thus, we chose $\dot{C}_*(\beta x) = V^{-1}(\beta x) = (\beta x)^{-1}$. The corresponding value of $nS_n^2(\hat{\beta}_c)/\hat{\sigma}^2$ is .0593. The approximated p value is .808, showing a strong support for the constant flaw rate assumption. This agrees with the conclusion stated in Firth et al. (1991).

The second example considers the data given by Finney (1947); see also Table 2 of Su and Wei (1991). The purpose of the analysis is to study the effect of the rate and volume of air inspired on a transient vasoconstriction in the skin of the digits. Su and Wei (1991) suggested that the logistic regression model should be proper for this data set since their test has an approximated p value equal to .34. We also assume the logistic regression model to determine $\dot{C}_*\{g(t)\} = V^{-1}(t)$ and apply the new method to this data set. The approximated p value of the new test is .55. This also shows that the logit of $E(Y \mid X = x)$ should be approximately linear.

4.2. Testing linearity

In this subsection we give one simulation to study the power performance of the new test. Consider the problem of testing linearity, $H_0 : E(Y \mid x) = \beta_1 x$, against $H_1 : E(Y \mid x) = \beta_1 x + \beta_2 x^2, \beta_2 \neq 0$, in the analysis of regression functions of one variable. Let the true model be $Y_i = X_i + \gamma X_i^2 + \varepsilon_i$, where the ε_i's are independent $N(0, 1)$ random variables, and the design points X_i's were also generated from a $N(0, 1)$ distribution independent of the ε_i's. In this context, the most popular test for testing H_0 is the likelihood ratio test using the F statistic. Here we shall compare the empirical powers of the test Z^2, Su and Wei's test G_n(Su and Wei (1991)), and the F test. In computing Z^2, we again define $\hat{\beta}_c$ to be the least squares estimate (i.e. $\dot{C}\{g(t)\} = 1$) but select $\dot{C}_*\{g(t)\} = 1 + t$, since with this choice of weight functions, $\sigma_c^2 = \sigma_*^2 = 1$. Also note that in this case, $g(t) = t$ and thus the quantities $\dot{C}_*\{g(\hat{\beta}_c x_i)\}\dot{g}(\hat{\beta}_c x_i)(\hat{\beta}_c x_i)$ used in defining $S_n(\hat{\beta}_c)$ (and Z^2) are $(1 + \hat{\beta}_c x_i)(\hat{\beta}_c x_i)$, $i = 1, \ldots, n$, respectively. The test Z^2 is optimal, in the sense described in section 3, for testing H_0 (against the local alternatives (3) with $h(t) = dt^2$, where $d \neq 0$ is any constant. That is to say, Z^2 is optimal, in the sense described in Section 3, for testing H_0) against H_1 for small values of β_2. All the empirical powers reported in this section were based on 1000 random samples of $\{(x_i, y_i), i = 1, \ldots, 50\}$. Further, for each observed sample 500 random samples of $\{z_i, i = 1, 2, \ldots, 50\}$ from the $N(0, 1)$ distribution were also generated to calculate the approximated p value of the test G_n (see Su and Wei (1991)). The estimated power curves for the tests Z^2, G_n and F at level $\alpha = .05$ are presented in Figure 1. The new test Z^2 is uniformly more powerful than the test G_n. The difference between the two power curves is significant. The likelihood-based F test is better than the Z^2 test. However, the difference between the two power curves seems less significant. Note that the Z^2 test is optimal with respect to a rather specific alternative, and could possibly limit the practical usefulness of the approach. The same can be said for the F test. To study the robustness of the discussed tests, we consider the true model as $Y_i = X_i + \gamma X_i^4 + \varepsilon_i$. Under the same simulation set-up as before, the estimated power curves for the tests Z^2, G_n and F at level $\alpha = .05$ are given in Figure 2. The figure shows that Z^2 has reasonable performance and is still uniformly more powerful than G_n. Thus, at least in this case, Z^2 is seen to be robust.

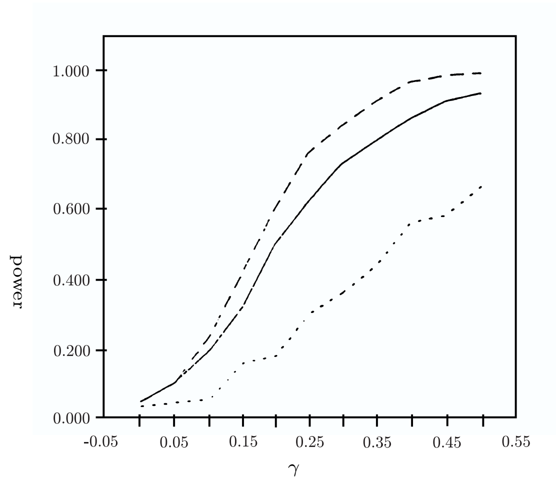

Figure 1. Power comparisons among the new test Z^2 the test G_n and the F test. The solid line represents the power curve of the test Z^2. The long dashed line represents the power curve of the F test, and the short dashed line represents the power curve of the test G_n.

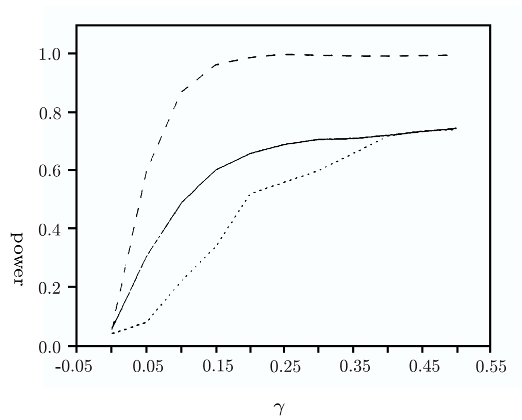

Figure 2. Power comparisions among the new test Z^2, the test G_n and the F test. The solid line represents the power curve of the test Z^2. The long dashed line represents the power curve of the F test, and the short dashed line represents the power curve of the test G_n

4.3. Testing for parameter variation in regression models

Testing for random coefficient variation in regression models has become an established part of modern applied statistical analysis (see McCabe and Leybourne (1933) and papers cited therein). This is because relationships between variables may change across units or through time. In this subsection we consider a regression model of the general form

$$Y_i = g(\gamma_i X_i) + \varepsilon_i$$

and it is assumed the parameter vector γ_i varies according to

$$\gamma_i = \beta + \eta_i, \ i = 1, \ldots, n,$$

where β is fixed, ε_i and η_i are independent random error terms with zero mean, $E(\varepsilon^2) = \sigma^2$, $E(\eta^T \eta) = \Omega(\omega)$, and $g(\cdot)$ is a known function. We also suppose that Ω is positive semidefinite for all $\omega \neq 0$, with $\Omega(0) = 0$, and that β, ω, and σ^2 are all unknown. Our aim is to test $H_0 : \omega = 0$.

For simplicity of presentation, we assume the dimensionality p of β is 1 and $\Omega(\omega) = \omega^2$ and follow McCable and Leybourne (1993) to assume that $g(\cdot)$ has continuous derivatives up to the second order and that, for local alternative values of ω, higher order terms can be ignored. Then a second-order Taylor series expansion about β gives

$$E(Y \mid X = x) = g(\beta x) + \ddot{g}(\beta x)\omega^2 x^2/2,$$

and

$$\text{Var}\,(Y \mid X = x) = \{\dot{g}(\beta x)\omega x\}^2 + \sigma^2.$$

As a consequence, from our previous disscussions it is seen that if $\ddot{g}(t) \neq 0$, then testing $\omega = 0$ against local alternative values of ω is equivalent to testing $H_0 : \mu(x) = g(\beta x)$ against local alternatives with $h(t) = t^2 \ddot{g}(t)$ as mentioned in Section 3. Also, for local alternative values of ω, we may assume the variance function $V(\cdot)$ is approximately constant. Under this situation, we see, that if one uses the weight function $C(\cdot)$ to define the estimate $\hat{\beta}_c$, then an optimal weight function $C_*(\cdot)$ to define the test statistic Z^2 should satisfy

$$\dot{C}_*(g(t))\dot{g}(t)t = \dot{C}(g(t))\dot{g}(t)t + dt^2\ddot{g}(t), \tag{5}$$

were d is any constant.

In this subsection, we compare the score-based test LM_g for the general case by McCable and Leybourne (1993) and the test based on the statistic Z^2 satisfying (5) with $\dot{C}(t) = 1$ under the following non-linear regression model :

$$Y_i = \exp(\gamma_i X_i) + \varepsilon_i.$$

Note that in this case, $g(t) = \exp(t)$, and according to (5) the quantities $\dot{C}_*\{g(\hat{\beta}_c x_i)\}\dot{g}(\hat{\beta}_c x_i)(\hat{\beta}_c x_i)$ used in defining $S_n(\hat{\beta}_c)$ (and Z^2) are $\{\exp(\hat{\beta}_c x_i)(\hat{\beta}_c x_i) + d(\hat{\beta}_c x_i)^2 \exp(\hat{\beta}_c x_i)\}$, $i = 1, \ldots, n$, respectively. Here, $\hat{\beta}_c$ is the least squares estimate. The null hypothesis is $H_0 : \omega = 0$, and the model was simulated with sample size $n = 50$ and over 1000 replications. The variable X_i was generated as $U(0, 3)$, and ε_i was generated as $N(0, 1)$. Random parameter variation was generated as $\gamma_i = 0.2 + \eta_i$, where $\eta_i \sim N(0, \omega^2)$. The estimated power curves for the tests Z^2 and LM_g at level $\alpha = .05$ are presented in Figure 3. It is clearly evident that for small or moderately large values of ω, the new test Z^2 is uniformly more powerful than the test LM_g. For large values of ω, the difference between the two power curves is not significant.

All the caculations in this section were programmed in FORTRAN 77 on a VAX 8650 machine. Random variables X_i, η_i and ε_i were generated using IMSL subroutines RNNOR and RNUN.

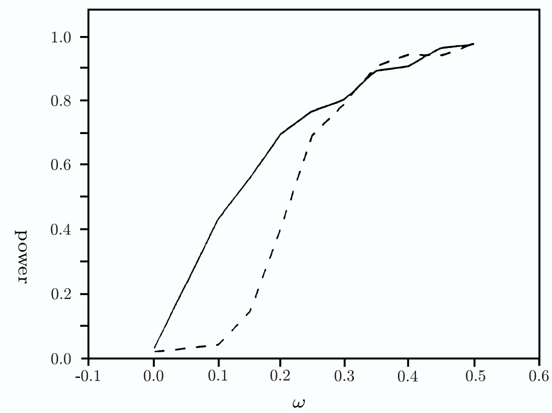

Figure 3. Power comparisons between the new test Z^2 and the test LM_g. The solid line represents the power curve of the test Z^2. The dashed line represents the power curve of the test LM_g.

5. Conclusions

In this article, we propose some conceptually simple methods to test the appropriateness of the mean function models. The approach discussed here is not a likelihood-based method, since no additional assumptions about the conditional distribution of the response variable given covariate values are assumed. The new test depends on the selection of weight functions. We have derived the efficacies of the test under sequences of alternatives. One can apply this result to guide the optimal choice of the weight functions in order to maximize the efficacy. Note that there are more than one pair of $C(\cdot)$ and $C_*(\cdot)$ satisfying (4), though their asymptotic relative efficiencies are the same. In practice, for the sake of simplicity in computation, we suggest that one takes $\dot{C}(t) = 1$ to derive the least squares extimate $\hat{\beta}_c$ and then applies (4) to select an optimal $C_*(t)$.

Asymptotically, the test procedure is distribution free, because the asymptotic null distribution of the new test statistic is independent of the distribution of Y given $X = x$. Further, it is also easy to see that the test is consistent for testing H_0 against the alternative hypothesis with $\beta_c H(\beta_c) \neq 0$. The new test is quite attractive because it is simple in computation and seems to have reasonable performance when the sample size is finite. One can also easily extend the method to test the null hypothesis using highly stratified data.

Acknowledgements

This research is supported by the National Science Council, R.O.C. The authors thank a referee for careful reading and useful suggestions.

References

Bissell, A. F. (1972). A negative binomial model with varying element sizes. *Biometrika* **59,** 435-441.

Cox, D. D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16**, 113-119.

Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34**, 320-334.

Firth, D., Glosup, J. and Hinkley, D. V. (1991). Model checking with nonparametric curves. *Biometrika* **78**, 245-252.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

McCabe, B. P. M. and Leybourne, S. J. (1993). Testing for parameter variation in non-linear regression models. *J. Roy. Statist. Soc. Ser. B* **55**, 133-144.

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

Staniswalis, J. G. and Severini, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.* **86**, 684-692.

Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86**, 420-426.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.

Graduate Institute of Statistics, National Central University, Chungli, Taiwan.
E-mail: p421272@twncu.865.ncu.edu.tw

Department of Statistics, Tamkang University, Taipei.
E-mail: jwwu@stat.tku.edu.tw

(Received October 1995; accepted June 1997)