The Quasi-score statistic in quasi-likelihood model

Article in Statistics: A Journal of Theoretical and Applied Statistics · January 2001
DOI: 10.1080/02331880108802752

2 authors:

Jong-Wuu Wu
National Chiayi University

Wen-Chuan Lee
Chang Jung Christian University

Statistics 35 (2001) pp. 523–535
Reprints available directly from the publisher
Photocopying permitted by license only

THE QUASI-SCORE STATISTIC IN QUASI-LIKELIHOOD MODEL

JONG-WUU WU[a,*] and WEN-CHUAN LEE[b]

[a]*Tamkang University, ROC;*
[b]*Ming Chuan University, ROC*

(Received 26 March 1997; In final form 13 February 2001)

In quasi-likelihood model, we propose the quasi-score statistic to establish test procedure for testing the hypothesis that whether the link function is correct or not. In addition, four practical examples are given to show the advantage of the proposed test.

Keywords: Quasi-likelihood model; Link function; Quasi-score statistic; Dimension reduction

1. INTRODUCTION

Let $\{(Y_i, X_i), i = 1, 2, \ldots, n\}$ denote a random sample, where Y_i is a real-valued response variable and $(X_{i1}, \ldots, X_{ip})^T$ is a p-dimension column vector of covariates. The distribution of $Y|X = x$ is assumed to be a member of the family of distributions depending on the parameter β and θ and x. Here β is a p-dimension row vector of regression coefficients and θ is a q-dimension row vector parameters. Under this set-up, the conditional mean $\mu(x)$ of Y given $X = x$ can be written as $\mu(x) = g(\beta, x; \theta)$ for some known function g. For example, for a particular value $\theta_0, g(\beta, x; \theta_0) = g(x^T\beta; \theta_0)$ can be referred to as a link function (also see McCullagh and Nelder, 1989) and $g(x^T\beta; \theta)$ can be considered as a parametric family of link functions. Therefore, $g(\beta, x; \theta_0)$ and $g(\beta, x; \theta)$ of this paper are also called by a link function

*Corresponding author.

and a parametric family of link functions, respectively. This paper provides a method for testing the adequacy of a particular hypothesized family of link functions (that is, testing the family link function $g(\cdot, \cdot; \theta) = g(\cdot, \cdot; \theta_0)$, if $g(\cdot, \cdot; \theta_0)$ is correct link function).

In the literature, several parametric families of link functions were considered by Prentice (1976); Aranda-Ordaz (1981); Pregibon (1980, 1985) and Wang (1991) *etc.* However, so far there exists no formal test for assessing the appropriateness of these hypothesized parametric families. As to the tests for the adequacy of a particular assumed link function, the general approach is to use the tests based on the deviance statistic or Pearson chi-squared statistic. Pierce and Schafer (1986) argued that the deviance residuals should provide a better basis for goodness-of-fit tests than the Pearson residuals. Williams (1987) investigated the case influence on the test statistics for selecting appropriate models.

In addition, the score statistic has been found useful for deriving diagnostic procedures for various purposes. In generalized linear model, Pregibon (1982) and Chen (1983, 1985) use it to test the need for additional explanatory variables, Wang (1985, 1987) use it to set up added-variable and constructed-variable plots. Chen and Wang (1991) use it for diagnostics on Cox's regression model. Finally, Wang (1991) use it to set up parameter plot in generalized nonlinear model.

In this article, the distribution of $Y|X=x$ is unknown in quasi-likelihood model, therefore we can't use the score statistic as the diagnostic tool. So, we propose the quasi-score statistic for detecting global lack-of-fit of a quasi-likelihood model for the link function with specifying a particular family of alternatives. Moreover, the diagnostic as Pregibon (1980); Chen (1983) and Wang (1985, 1991) *etc.*, are special cases of our test procedure. In the next section, we give the basic structure of the quasi-likelihood model. Section 3 develops the quasi-score statistic, set up parameter plot, and gives sufficient conditions for the asymptotic distribution of the quasi-score statistic. Four practical examples and conclusions are given in Section 4.

2. QUASI-LIKELIHOOD MODEL

Assume Y to be the response variable with the conditional mean, given $X=x$,

$$E(Y|X = x) = g(x^T\beta) \tag{2.1}$$

and the conditional covariance

$$\mathrm{Var}(Y|X=x)=\phi V[g(x^T\beta)]. \tag{2.2}$$

Here $g(\cdot)$ and $V(\cdot)$ are given functions and ϕ and β are unknown parameters to be estimated. Models of the general form given by (2.1) and (2.2) are discussed by McCullagh and Nelder (1989) using the name quasi-likelihood models. Generally, the main concern is the estimation of the regression parameter β. Since a fully specified likelihood is not assumed, Wedderburn (1974) has shown that the ordinary least-squares equations may be generalized to the quasi-likelihood equations

$$\sum_{i=1}^{n}\frac{\partial K(\beta;X_i,Y_i)}{\partial\beta}\equiv\sum_{i=1}^{n}\frac{Y_i-g(X_i^T\beta)}{V[g(X_i^T\beta)]}\cdot\frac{\partial g(X_i^T\beta)}{\partial\beta}=0 \tag{2.3}$$

to be solved for β giving a solution vector $\hat{\beta}$, say. Here $K(\beta;x,y)$ is called the quasi-likelihood function. McCullagh (1983) has shown that under certain conditions, $\hat{\beta}$ is consistent and asymptotically normal provided the link function is correctly specified. Suppose the link function is incorrectly specified, the estimator β can not be consistent, though one can show that $\hat{\beta}$ still converge to some constant vector, say β_*. On the other hand, if model (2.1) is correct but the variance function $V(\cdot)$ is incorrectly defined, the estimator $\hat{\beta}$ can still be consistent, except that now the estimator becomes less efficient. Therefore, in the following section, we will propose test statistic to test whether the link function is correct or not, and its graphical version for detecting the need of a single parameter and influential observations on the need.

3. QUASI-SCORE STATISTIC AND PARAMETER PLOT

3.1. Basic Framework for the Test Procedures

We extend the model given by (2.1) and (2.2) for the response variable Y and a parametric family of link function for modelling the mean function of Y. Therefore, in the present, we consider the following model:

$$E(Y|X=x)=g(\beta,x;\theta) \tag{3.1}$$

and

$$\mathrm{Var}(Y|X = x) = \phi V_0[g(\beta, x; \theta)]. \tag{3.2}$$

According to our set-up, let the link function $g(\cdot)$ and the variance function $V_0(\cdot)$ be known, and we assume the null hypothesis $H_0 : \theta = \theta_0$ (That is, whether the link function $g(\beta, x; \theta)$ is $g(\beta, x; \theta_0)$) or not, where $g(\beta, x; \theta_0)$ is defined as true link function, θ_0 is known value, β in the parameter space $\Pi \in R^p$ and θ in the parameter space $\Theta \in R^q$. Model (3.1)–(3.2) includes the class of generalized nonlinear models, see Wang (1991).

Under the hypothesized link function $g(\cdot, \cdot; \theta_0)$, the quasi-likelihood estimate $\hat{\beta}_n$ is a solution of the quasi-likelihood equation

$$\sum_{i=1}^{n} \frac{Y_i - g(\beta, X_i; \theta_0)}{V_0[g(\beta, X_i; \theta_0)]} \cdot \frac{\partial g(\beta, X_i; \theta_0)}{\partial \beta} = 0. \tag{3.3}$$

We know that $\hat{\beta}_n$ will converge in probability to the true regression parameter value, say β_0 under H_0 and the regularity conditions of the Appendices.

In some applications ϕ may be known: often $\phi = 1$. More often ϕ must be estimated in order to construct tests for testing H_0. In the absence of information beyond second moments of Y given $X = x$, there is little alternative to using

$$\hat{\phi}_n = \frac{1}{n-p} \sum_{i=1}^{n} \frac{[Y_i - g(\hat{\beta}_n, X_i; \theta_0)]^2}{V_0[g(\hat{\beta}_n, X_i; \theta_0)]}.$$

For large n, $\hat{\phi}_n$ will converge in probability to the true dispersion parameter value, say ϕ_0 under H_0 and the conditions (R6) and (R7) in the Appendices. In order to test $H_0 : \theta = \theta_0$, we define a quasi-score statistic at $(\beta, \theta) = (\hat{\beta}_n; \theta_0)$ given by

$$U(\hat{\beta}_n) = \sum_{i=1}^{n} \frac{Y_i - g(\hat{\beta}_n, X_i; \theta_0)}{\hat{\phi}_n V_0[g(\hat{\beta}_n, X_i; \theta_0)]} \cdot \frac{\partial g(\hat{\beta}_n, X_i; \theta_0)}{\partial \theta}.$$

General asymptotic theory for $U(\hat{\beta}_n)$ under $H_0 : \theta = \theta_0$ will be given in Theorem 3.1. Before stating the basic conditions and asymptotic results, we first denote $\alpha = (\beta, \theta)$, $\alpha_0 = (\beta_0, \theta_0)$, $\partial l(\alpha; x, y)/\partial \theta = ((y - g(\beta, x; \theta))/V_0[g(\beta, x; \theta)]) \cdot (\partial g(\beta, x; \theta)/\partial \theta)$, $\partial l(\alpha; x, y)/\partial \beta = ((y - g(\beta, x; \theta))/$

$V_0[g(\beta, x; \theta)]) \cdot (\partial g(\beta, x; \theta)/\partial\beta)$, and $g(\alpha, x) = g(\beta, x; \theta)$ for the ease of presentation. We note that direct application of the "δ-method;" delivers the asymptotically normal distribution of $U(\hat{\beta}_n)$ and a consistent estimator of its covariance matrix. To simplify our presentation of the asymptotic result, we define

$$I_{ab} = \frac{1}{\phi_0} \cdot E\left[\frac{1}{V_0[g(\alpha_0; X)]} \cdot \left(\frac{\partial g(\alpha_0; X)}{\partial a}\right) \cdot \left(\frac{\partial g(\alpha_0; X)}{\partial b}\right)^T\right],$$

where a, $b = \theta$ or β. Therefore, under the above regularity conditions and $H_0 : \theta = \theta_0$, we obtain

$$\hat{\beta}_n - \beta_0 = \phi_0^{-1} I_{\beta\beta}^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\alpha_0; X_i, Y_i)}{\partial \beta}\right] + o_p(n^{-1/2}),$$

and

$$n^{-1} U(\hat{\beta}_n) = \frac{1}{\phi_0} \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial l(\alpha_0; X_i, Y_i)}{\partial \theta} - \phi_0 I_{\theta\beta} (\hat{\beta}_n - \beta_0)\right] + o_p(n^{-1/2}).$$

So, the asymptotic distribution of $n^{-1/2} U(\hat{\beta}_n)$ is equal to that of

$$T_n^* = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{1}{\phi_0} \frac{\partial l(\alpha_0; X_i, Y_i)}{\partial \theta} - \frac{1}{\phi_0} I_{\theta\beta} I_{\beta\beta}^{-1} \frac{\partial l(\alpha_0; X_i, Y_i)}{\partial \beta} \right\}.$$

Since T_n^* is the average of i.i.d. random variables, therefore the central limit theorem establishes the assertion of Theorem 3.1. The asymptotic covariance matrix is also stated in Theorem 3.1 with detailed analysis omitted.

THEOREM 3.1 *Suppose assumptions Conditions (R1)–(R7) given in the Appendices are satisfied. Then we have*

$$n^{-1/2} U(\hat{\beta}_n) \xrightarrow{d} MVN(0, \Sigma_0), \text{ as } n \to \infty,$$

where the covariance matrix $\Sigma_0 = I_{\theta\theta} - I_{\theta\beta} I_{\beta\beta}^{-1} I_{\beta\theta}$.

For the estimate $U(\hat{\beta}_n)$ to be of practical use, one needs some estimate of its asymptotic covariance matrix. Therefore, under H_0, the asymptotic covariance matrix Σ_0 can be consistently estimated by

$\hat{\Sigma}_0 = (1/n)[\hat{I}_{\theta\theta} - \hat{I}_{\theta\beta}\hat{I}_{\beta\beta}^{-1}\hat{I}_{\beta\theta}]$. Here we define

$$\hat{I}_{ab} = \frac{1}{\hat{\phi}_n}\sum_{i=1}^{n}\frac{1}{V_0[g(\hat{\beta}_n, X_i; \theta_0)]}\cdot\left[\frac{\partial g(\hat{\beta}_n, X_i; \theta_0)}{\partial a}\right]\cdot\left[\frac{\partial g(\hat{\beta}_n, X_i; \theta_0)}{\partial b}\right]^T,$$

where $a, b = \theta$ or β.

Accordingly, the quasi-score test statistic $W_n = U(\hat{\beta}_n)^T[\hat{I}_{\theta\theta} - \hat{I}_{\theta\beta}\hat{I}_{\beta\beta}^{-1}\hat{I}_{\beta\theta}]^{-1}U(\hat{\beta}_n)$ can be used to test H_0. Under the null hypothesis, W_n converges weakly to a central chi-square variable with q degree of freedom, and large value of W_n means that one should reject H_0. More specifically, to carry out the test, one computes W_n and compares it to critical value of the χ_q^2 distribution of a given size of test. Suppose the dispersion parameter ϕ is given, say $\phi = 1$, then we replace $\hat{\phi}_n$ by 1 in the test statistic W_n.

In addition, let Z be a vector of $([Y_i - g(\hat{\beta}_n, X_i; \theta_0)]/\hat{\phi}_n V_0[g(\hat{\beta}_n, X_i; \theta_0)])$ and W be a matrix of $[\partial g(\hat{\beta}_n, X_i; \theta_0)/\partial\theta]^T$, then we can rewrite $U(\hat{\beta}_n)$ as

$$U(\hat{\beta}_n) = W^T Z.$$

Moreover, let $V = \text{diag}[1/\hat{\phi}_n V_0[g(\hat{\beta}_n, X_i; \theta_0)]]$ and Ψ be a matrix of $[\partial g(\hat{\beta}_n, X_i; \theta_0)/\partial\beta]^T$, we can simplify the quasi-score test statistic W_n to become

$$W_n = Z^T W[(V^{1/2}W)^T(I - H)V^{1/2}W]^{-1}W^T Z, \tag{3.4}$$

where $H = V^{1/2}\Psi[\Psi^T V\Psi]^{-1}\Psi^T V^{1/2}$. When the dimension of θ is one ($q = 1$), we can establish a graphical version of (3.4) for diagnostic purposes.

Let $S = V^{1/2}\Psi\hat{\beta}_n + V^{-1/2}Z$ and $\theta^* = \theta - \theta_0$, we construct an approximate model

$$S = V^{1/2}\Psi\beta + V^{1/2}W\theta^* + \varepsilon, \tag{3.5}$$

where $\varepsilon \sim N(0, \sigma^2 I)$. The F-statistic for $\theta^* = 0$ under model (3.5) is computationally equivalent to the quasi-score statistic in (3.4). Thus the add-variable plot for variable $V^{1/2}W$, which is a plot of $R = V^{-1/2}Z$ *versus* $(I - H)V^{1/2}W$ (see Cook and Weisberg, 1982), can be treated as a graphical version of statistic (3.4). The significant slope of the regression line in the plot corresponds to the significance of the parameter $\theta(=\theta_0)$ in the quasi-likelihood model. For later convenience, we call R and $(I-H)V^{1/2}W$ as the residuals and W-residuals respectively (see Wang, 1985). We name the plot as parameter plot. If $Y_1, \ldots, Y_n$ are a random sample from a specified generalized nonlinear model with the corresponding values of p explanatory variables $X_1, \ldots, X_n$ and $\theta_0 = 0$, then the parameter plot reduce to plot of Wang (1991).

4. EXAMPLES AND FINAL REMARKS

To illustrate the usefulness of test procedure in quasi-likelihood model, we now apply the proposed test W_n to some familiar data sets.

In the first example, we consider the data reported by Bissel (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 I of Firth *et al.* (1991). They assumed that there is a constant flaw rate and thus the mean count $E(Y|X=x) = x^T\beta$ and Y to be a Poisson variate. Therefore, we consider the family of power transformations proposed by Box and Cox (1964) given as

$$g(x^T\beta;\theta) = \begin{cases} ((x^T\beta+1)^\theta - 1)/\theta, & \text{if } \theta \neq \theta; \\ \ln(x^T\beta+1), & \text{otherwise.} \end{cases}$$

So, under $H_0: \theta = 1$, we obtain $\hat{\beta}_n = 0.0151$, and find that the quasi-score test statistic $W_n = 0.8733 \times 10^{-2}$ (P-value $= 0.9255$). Therefore, this concludes that we have stronger evidence not to reject H_0. This agrees with the conclusion stated in Firth *et al.* (1991).

The second example is taken from Draper and Smith (1981, p. 205). The problem is to analyze the relationship between the pounds of steam used monthly $= Y$, and $X = (X_1, X_2)$, where $X_1 =$ the operating days per month and $X_2 =$ the average atmospheric temperature. Draper and Smith used a simple additive model with normal errors to model the data. But Su and Wei (1991) calculated the approximated p value of their test G_n to obtain P-value $= 0.048$. In addition, Cheng and Wu (1994) calculated the value of their test M_n to obtain $M_n = 19.5561$ (P-value $= 2.103 \times 10^{-4}$). They leads to the rejection of the simple additive model. Here, we use the family of link function given as the first example. Using our test, we obtain $\hat{\beta}_n = (9.1269, 0.2028, -0.0724)$ and find that the quasi-score test statistic $W_n = 11.701$ (P-value $= 6.2477 \times 10^{-4}$) giving stronger evidence to reject the simple additive model. So, this agrees with the conclusion of Su and Wei (1991) and Cheng and Wu (1994).

Our third example considers the data given by Bliss (1935) where the mortality of adult flour beetles is related to exposure to gaseous carbon disulphide with log dosage; see also Table I of Prentice (1976) or Table IV of Aranda-Ordaz (1981). From Table I of Cheng and Wu (1994), we have been known that the P-values of (M_n, G_n) are $(1.999 \times 10^{-2}, 0.03)$, $(0.997, 0.990)$, and $(0.999, 0.990)$ for logistic model, Prentice's model, and complementary log–log model, respectively. Both their test M_n and the test G_n all suggest that the complementary log–log model and Prentice's model are almost equally satisfactory. Now, first, we consider the family of link function (see Prentice, 1976) given by

$$g(x^T\beta;\theta) = \left\{ \frac{\exp(\beta_0 + x_1\beta_1)}{1 + \exp(\beta_0 + x_1\beta_1)} \right\}^{\theta}. \tag{4.1}$$

For $\theta = 1$, (4.1) reduces to the logistic model, while for $\theta = 0.279$, the Prentice's model is obtained. So, under $H_0 : \theta = 1$, we obtain $\hat{\beta}_n = (-60.718, 34.271)$, and find that the quasi-score test statistic $W_n = 6.8095$ (P-value $= 9.0647 \times 10^{-3}$). Giving the sufficient evidence to reject the logistic model. This agrees with the conclusion of Cheng and Wu (1994). In addition, under $H_0 : \theta = 0.279$, we obtain $\hat{\beta}_n = (-116.301, 63.970)$ and find that the quasi-score test statistic $W_n = 0.5653 \times 10^{-4}$ (P-value $= 0.9940$). Giving stronger evidence not to reject the Prentice's model. This see that the Prentice's model does have better support from the data, under the family (4.1). Next, we consider the family of link function (see Aranda-Ordaz, 1981) giving by

$$g(x^T\beta;\theta) = 1 - [1 + \theta\exp(\beta_0 + x_1\beta_1)]^{-1/\theta}, \quad \theta \geq 0. \tag{4.2}$$

For $\theta = 1$, (4.2) reduces to the logistic model, while for $\theta = 0$, the complementary log–log model is obtained. So, under $H_0 : \theta = 1$, we find that the quasi-score test statistic $W_n = 6.4956$ (P-value $= 1.0814 \times 10^{-2}$), giving the sufficient evidence to reject the logistic model. In addition, under $H_0:\theta=0$, we obtain $\hat{\beta}_n = (-39.574, 22.042)$, and find that the quasi-score test statistic $W_n = 0.1541 \times 10^{-2}$ (P-value $= 0.9687$). Giving stronger evidence not to reject the complementary log−log model. This see that the complementary log−log model does have better support from the data, under the family (4.2). Therefore, according to the above analysis, we suggest that the Prentice's model and the complementary log−log model are more appropriate for the data. This agrees with the conclusion of Cheng and Wu (1994).

Our fourth example consider the data reported by Carr (1960). There are 24 independent counts of the reaction rate y, partial pressures of hydrogen x_1, n-pentane x_2, and isopentane x_3. The data are given in Table I of Box and Hill (1974). They assumed that $E(Y|X=x) = (\beta_0\beta_2(x_2 - x_3/1.632)/(1. + x_1\beta_1 + x_2\beta_2 + x_3\theta))$, and Y to be a normal variate. Therefore, we consider the family of link function (see Box and Hill, 1974) given by

$$g(\beta, x; \theta) = \frac{\beta_0\beta_2(x_2 - x_3/1.632)}{1 + x_1\beta_1 + x_2\beta_2 + x_3\theta}. \qquad (4.3)$$

Box and Hill described a weighted analysis based on the linear version of (4.3) obtained using y_i^{-1} as the response. In this example, we obtain $\hat{\beta}_n = (-3839.393, 6\text{E}-06, -1.1\text{E}-05)$ and the quasi-score test statistic $W_n = 4.8943$ (P-value $= 2.6946 \times 10^{-2}$) for $H_0:\theta=0$. Therefore, this concludes that we have evidence to reject H_0. This agrees with the conclusion stated in Box and Hill (1974) and Wang (1991).

In addition, the parameter plot of the first and second examples are given in Figures 1 and 3, respectively. The parameter plot of Figure 1 indicate two outlier that might affect our conclusions. In fact, when these two points were deleted, the parameter plot given in Figure 2. From Figures 1 and 2, we saw that they display no systematic tendencies to be positive and negative (that is, $\theta^* = 0$). Therefore, these two points were not influential observations. Moreover, the Wald statistic W_n become 0.01617 (P-value $= 0.8988$) and $\hat{\beta}_n$ reduces to 0.0138. These change does not affect the before conclusion. So, we don't think that these two points need to deleted. The parameter plot of Figure 3, it shows the need of a transformation, *i.e.*, an alternative link, and no influential observation on this indication. This confirms the result in Cheng and Wu (1994).

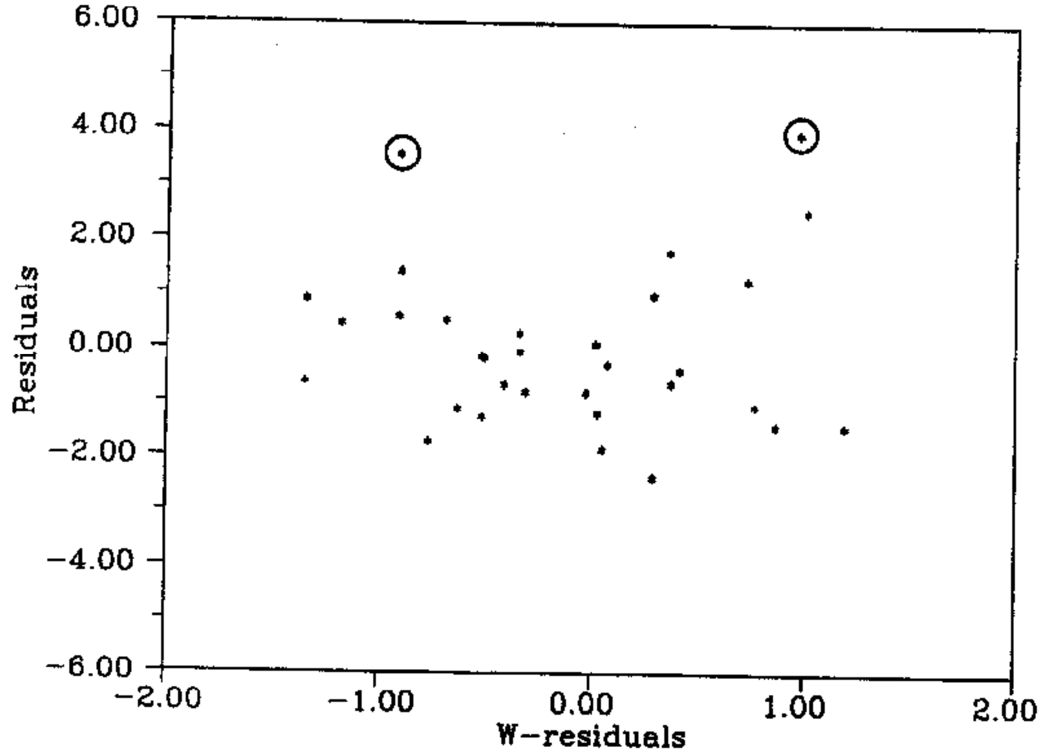

FIGURE 1 Parameter plot for Bissell data.

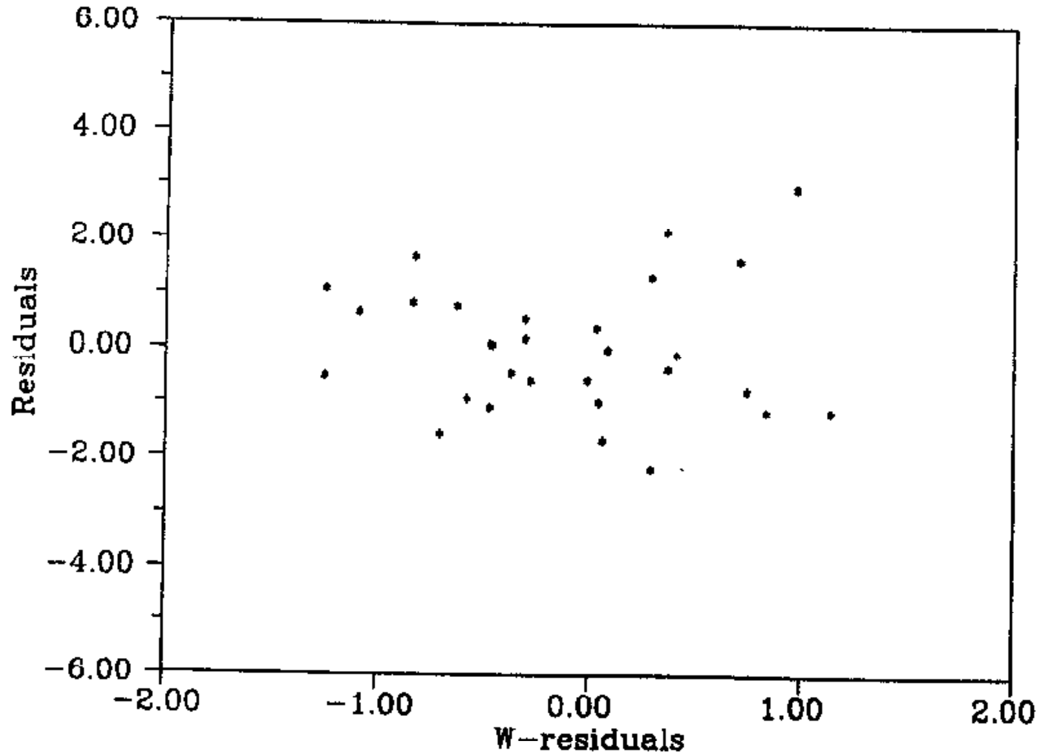

FIGURE 2 Parameter plot for Bissell data.

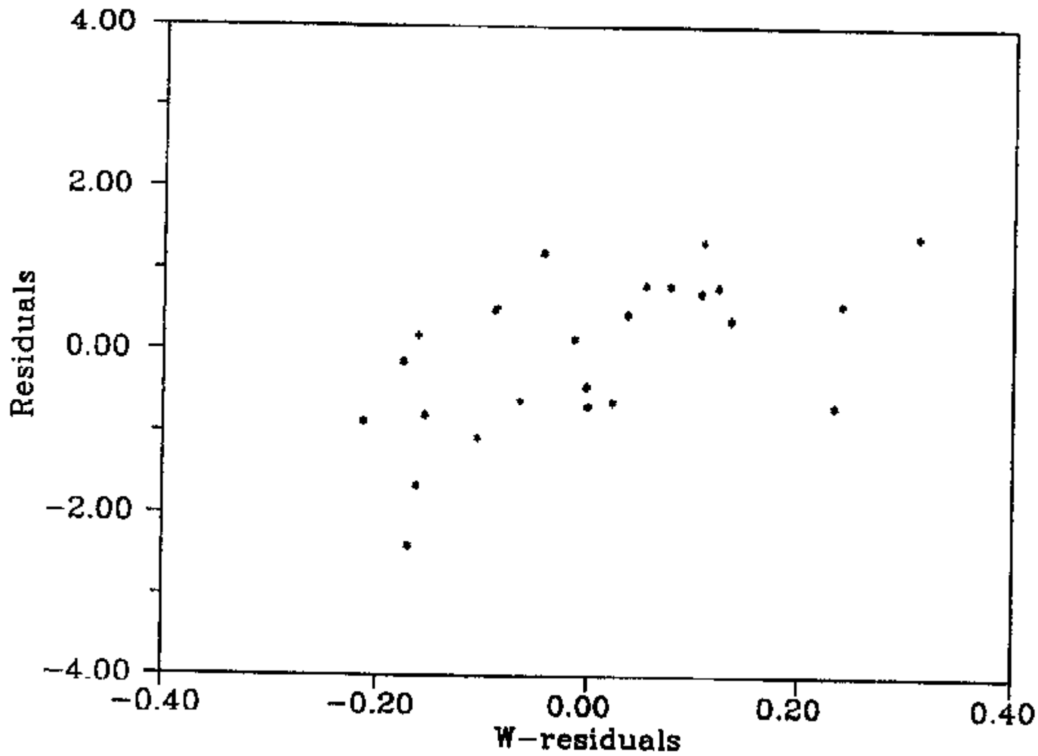

FIGURE 3 Parameter plot for Draper Smith data.

Generally speaking, in calculating the quasi-score statistic we only need to specify the first two moments of the response variable Y given $X=x$. Asymptotically, the test procedure is distribution free, since the asymptotic null distribution of the quasi-score statistic is independent of the underlying conditional distribution of Y given $X=x$. In addition, from the above example, we know that the quasi-score test W_n is also a good alternative in comparison with other tests.

Acknowledgements

This research partially supported by the National Science Council, R.O.C. (Plan No. NSC 83-0208-M-032-019). The authors thank a referee for careful reading and useful suggestions.

References

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

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

Bliss, C. I. (1935) The calculation of the dosage-mortality curve, *Annals of Applied Biology*, **22**, 134–167.

Box, G. E. P. and Cox, D. P. (1964) An analysis of transformations (with discussion), *Journal of the Royal Statistical Society, Ser. B*, **26**, 211–252.

Box, G. E. P. and Hill, W. J. (1974) Correcting Inhomogeneity of Variance with Power Transformation Weighting, *Technometrics*, **16**, 385–389.

Carr, N. L. (1960) Kinetics of catalytic isomerization of n-pentane, *Indus. Eng. Chem.*, **52**, 391–396.

Chen, C. F. (1983) Score tests for regression models, *J. Amer. Statist. Assoc.*, **78**, 158–161.

Chen, C. F. (1985) Robustness aspects of score tests for generalized linear and partially linear regression model, *Technometrics*, **27**, 277–283.

Chen, C. H. and Wang, P. C. (1991) Diagnostic plots in Cox's Regression Model, *Biometrics*, **47**, 841–850.

Cheng, K. F. and Wu, J. W. (1994) Testing Goodness of Fit for A Parametric Family of Link Functions, *J. Amer. Statist. Assoc.*, **89**, 657–664.

Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*, London: Chapman and Hall.

Draper, N. R. and Smith, H. (1981) *Applied Regression Analysis*, 2nd edn., New York: John Wiley.

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

McCullagh, P. (1983) Quasi-likelihood function, *Ann. Statist.*, **11**, 59–67.

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

Pierce, D. A. and Schafer, D. W. (1986) Residuals in generalized linear models, *J. Amer. Statist. Assoc.*, **81**, 977–986.

Pregibon, D. (1980) Goodness of link tests for generalized linear models, *Appl. Statist.*, **29**, 15–24.

Pregibon, D. (1982) Score tests in GIM with applications. In: *Lecture Notes in Statistics, no.* 14, *GLIM*.82: *Proceedings of the International Conference on Generalized Linear Models*. Gilchrist, R. (Ed.), Springer-Verlag, New York.

Pregibon, D. (1984) Review for generalized linear models, *Ann. Statist.*, **12**, 1589–1596.

Prentice, R. L. (1976) A generalization of the probit and logit methods for dose response curves, *Biometrics*, **32**, 761–768.

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.

Wang, P. C. (1985) Adding a variable in generalized linear models, *Technometrics*, **27**, 273–276.

Wang, P. C. (1987) Residual plots for detecting nonlinearity in generalized linear models, *Technometrics*, **29**, 435–438.

Wang, P. C. (1991) Diagnostics and score statistics in regression, *Ann. Inst. Statist. Math.*, **43**, 647–656.

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

Williams, D. A. (1987) Generalized linear model diagnostics using the deviance and single case deletions, *Appl. Statist.*, **36**, 181–191.

Jong-Wuu Wu
Department of Statistics
Tamkang University
Tamsui, Taipei
Taiwan
R.O.C.

Wen-Chuan Lee
Department of Business Administration
Ming Chuan University
Taipei
Taiwan
R.O.C.

APPENDICES

(R1) For all α in a neighborhood $N(\alpha_0)$ of α_0, the derivatives $(\partial^2 l(\alpha; x, y)/\partial\beta_i\partial\beta_j)$ and $(\partial^3 l(\alpha; x, y)/\partial\beta_i\partial\beta_j\partial\beta_k)$ exist for all (x, y) in the support of (X, Y), $i, j, k = 1, 2, \ldots, p$.

(R2) For all α in a neighborhood $N(\alpha_0)$ of α_0, the derivatives $(\partial^2 l(\alpha; x, y)/\partial\theta_i\partial\beta_j)$ and $(\partial^3 l(\alpha; x, y)/\partial\theta_i\partial\beta_j\partial\beta_k)$ exist for all (x, y) in the support of (X, Y), $i = 1, 2, \ldots, q$; $j, k = 1, 2, \ldots, p$.

(R3) There exist function $H_i(x, y)$, possibly depending on α_0, and $i = 1, 2, \ldots, p$, such that $E\{H_i(X, Y)\} < \infty$, and for all α in $N(\alpha_0)$, and (x, y) in the support of (X, Y)

$$\left|\frac{\partial^3 l(\alpha; x, y)}{\partial\beta_i\partial\beta_j\partial\beta_k}\right| \leq H_i(x, y), \quad j, k = 1, 2, \ldots, p.$$

(R4) There exist function $H_i^*(x, y)$, possibly depending on α_0, and $i = 1, 2, \ldots, q$, such that $E\{H_i^*(X, Y)\} < \infty$, and for all α in $N(\alpha_0)$, and (x, y) in the support of (X, Y)

$$\left|\frac{\partial^3 l(\alpha; x, y)}{\partial\theta_i\partial\beta_j\partial\beta_k}\right| \leq H_i^*(x, y), \quad j, k = 1, 2, \ldots, p.$$

(R5) $E\{(\partial l(\alpha_0; X, Y)/\partial\beta)\} \cdot \{(\partial l(\alpha_0; X, Y)/\partial\beta)\}^T$ and $E\{(\partial l(\alpha_0; X, Y)/\partial\theta)\} \cdot \{(\partial l(\alpha_0; X, Y)/\partial\theta)\}^T$ are positive definite matrices, $E\{(\partial^2 l(\alpha_0; X, Y)/\partial\beta_i\partial\beta_j)\}$ and $E\{(\partial^2 l(\alpha_0; X, Y)/\partial\beta_i\partial\beta_j)\}$ are nonsingular matrices, and $E\{(\partial^2 l(\alpha_0; X, Y)/\partial\theta_i\partial\theta_j)\} < \infty$.

(R6) The functions $g(\alpha; x)$ and $(\partial g(\alpha; x/\partial\theta)$ are uniformly continuous at α_0 for all x in the support of X.

(R7) $E\{[(Y - g(\alpha_0; x))^2/V_0[g(\alpha_0; X)]]^2 + (1/V_0[g(\alpha_0; X)])\} < \infty$.