Ann. Inst. Statist. Math.
Vol. 43, No. 4, 647–656 (1991)

DIAGNOSTICS AND SCORE STATISTICS IN REGRESSION*

P. C. WANG

Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan 32054, Republic of China

(Received June 27, 1989; revised June 21, 1990)

Abstract. In a generalized linear model, we have a linear predictor. We extend to a nonlinear one and propose a unified method to establish diagnostic procedures for such models with nonlinear links. Applications of the procedures to various useful models are given with examples.

Key words and phrases: Generalized linear model, nonlinear, diagnostic plot, case influence.

1. Introduction

Many investigations have been centred on the generalized linear model since it was proposed by Nelder and Wedderburn (1972). McCullagh and Nelder (1983) provided an excellent review and some future research problems. Further research topics were suggested by Pregibon (1984). Diagnostic procedures for checking assumptions in such models have been proposed and investigated. Landwehr *et al.* (1984) and Wang (1985, 1987) proposed several diagnostic plots for selecting explanatory variables in an appropriate form and detecting the effect of observations on the selections. Pierce and Schafer (1986) discussed useful residuals for model checking. Williams (1987) investigated the case influence on the test statistics for selecting appropriate models.

The score statistic has been found useful to derive diagnostic procedures for various purposes. In the normal linear model, Atkinson (1983) used it to establish constructed variable plots for diagnosing the need for the transformation of the responses, and Lawrance (1987) suggested a standardized score statistic to improve Atkinson's statistic (1982) for the same purpose; Cook and Weisberg (1983) used it and its graphical version for detecting heteroscedasticity; Pregibon (1982) and Chen (1983) used it to test the need for additional explanatory variables in generalized linear models (GLM). Wang (1985, 1987) used it to set up added-variable and constructed-variable plots for the GLM. Chen and Wang (1991) used it for diagnostics on Cox's regression model. More discussions on the usefulness of the score statistics can be found in Chen (1983, 1985).

* Research supported by National Science Council of the Republic of China (NSC-25009).

In this article, we explore the score statistic for more general problems and obtain its graphical versions for diagnostic purposes. Some of the procedures mentioned above become special cases in the results. In the next section, the GLM is reviewed and extended to generalized nonlinear models (GNLM). In some cases, the latter model can be treated as an alternate for the GLM. This idea will be clear in Section 4. Under the framework of generalized nonlinear models, the score statistic and its graphical version are obtained to include certain parameters into the model in Section 3. The explorations of these for several useful situations are given with some illustrative examples in Section 4. The emphasis is on those GNLM whose null models become GLM such as GLM with parametrized links and dispersions. Section 5 gives concluding remarks.

2. Background

There are two components and one link function involved in a GLM. The random component is that in which the response y with mean $E(y) = \mu$ has a probability density function

$$h(y, \theta, \phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y, \phi)\} \tag{2.1}$$

with smooth functions $a(\ )$, $b(\ )$ and $c(\ )$ and parameters θ and ϕ. A known ϕ is assumed for our discussions. When it is unknown, it is replaced by its maximum likelihood estimate. The systematic component involves explanatory variables in a linear predictor

$$\eta = x^T \beta \tag{2.2}$$

where x is a vector of corresponding observed values of explanatory variables and β is a vector of parameters. The link function g is then used to connect these two components by $\eta = g(\mu)$. Note that the normal linear model is the simplest example with the identity link in this setup. Logistic regression, Poisson regression and exponential regression, the three common useful models other than the normal, are also within the framework. The exact relation between θ and the components for these cases can be found in McCullagh and Nelder (1983).

As nonlinear regressions in the normal case, we can extend the linear predictor (2.2) to a nonlinear predictor

$$\eta = f(x, \beta, \gamma) \tag{2.3}$$

where f can be any known smooth function and γ is another vector parameter. This possibility has also been discussed in Section 10.4 of McCullagh and Nelder (1983) and Chapter 6 of Cox *et al.* (1987). We call the model with the nonlinear predictor the generalized nonlinear model, because it can be viewed as the generalization of the nonlinear normal model. A GLM is also a special case of generalized nonlinear models. With some special functions f, the model becomes a GLM under certain conditions, for example, $f(x, \beta, \gamma) = (x^T\beta)^{1-\gamma}$ when $\gamma = 0$. In the following section, we explore the test statistic for testing if the parameters γ in the generalized nonlinear model are significant, and its graphical version for detecting the need of a single parameter and influential observations on parameter γ.

3. The diagnostic plot

Assume that $y_1, y_2, \ldots, y_n$ are a random sample from a specified generalized nonlinear model with the corresponding values of p explanatory variables, $x_1, \ldots, x_n$. The log-likelihood function for (β^T, γ) is

$$l(\beta, \gamma) = \sum_{i=1}^{n} \{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\}. \tag{3.1}$$

Following Pregibon's (1982) derivation of the score statistic in GLM we can obtain that for the testing of $\gamma = 0$ in GNLM. For convenience, k_α is used to denote the partial derivative of function k having more than two parameters with respect to parameter α and "$\cdot$" is the differentiation sign when the function has one parameter. Now let g^* be the inverse function of $\eta = g(\dot{b}(\theta))$, then $\theta_i = g^*(\eta_i) = g^*(f(x_i, \beta, \gamma))$. Under the null hypothesis that $\gamma = 0$, the maximum likelihood estimate (MLE) $\hat{\beta}$ of β can be obtained by an iterative method, and then $\hat{\eta}_i$ and $\hat{\theta}_i$ of η_i and θ_i by $f(x_i, \hat{\beta}, 0)$ and $g^*(f(x_i, \hat{\beta}, 0)$ respectively. In order to find the score statistic for testing $\gamma = 0$, we obtain the first derivative of (3.1) with respect to γ and evaluate it at $(\beta, \gamma) = (\hat{\beta}, 0)$,

$$U(\hat{\beta}) = \sum_{i=1}^{n} [y_i - \dot{b}(\hat{\theta}_i)]\dot{\theta}(\hat{\eta}_i) f_\gamma(x_i, \hat{\beta}, 0)^T/a(\phi) = Z^T S \tag{3.2}$$

where S is a vector of $(y_i - \dot{b}(\hat{\theta}_i))/a(\phi)$ and Z a matrix of $\dot{\theta}(\hat{\eta}_i) f_\gamma(x_i, \hat{\beta}, 0)$. The information matrix for (β, γ) evaluated at $(\beta, \gamma) = (\hat{\beta}, 0)$ is

$$I(\hat{\beta}, 0) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\gamma} \\ I_{\gamma\beta} & I_{\gamma\gamma} \end{bmatrix} \tag{3.3}$$

with

$$I_{\beta\beta} = \sum_{i=1}^{n} \ddot{b}(\hat{\theta}_i)\dot{\theta}^2(\hat{\eta}_i) f_\beta(x_i, \hat{\beta}, 0) f_\beta(x_i, \hat{\beta}, 0)^T/a(\phi),$$

$$I_{\beta\gamma} = I_{\gamma\beta}^T = \sum_{i=1}^{n} \ddot{b}(\hat{\theta}_i)\dot{\theta}^2(\hat{\eta}_i) f_\beta(x_i, \hat{\beta}, 0) f_\gamma(x_i, \hat{\beta}, 0)^T/a(\phi) \qquad \text{and}$$

$$I_{\gamma\gamma} = \sum_{i=1}^{n} \ddot{b}(\hat{\theta}_i)\dot{\theta}^2(\hat{\eta}_i) f_\gamma(x_i, \hat{\beta}, 0) f_\gamma(x_i, \hat{\beta}, 0)^T/a(\phi).$$

Let $V = \mathrm{diag}\{\ddot{b}(\hat{\theta}_i)/a(\phi)\}$ and $\mathcal{X}$ be a matrix of $\dot{\theta}(\hat{\eta}_i) f_\beta(x_i, \hat{\beta}, 0)^T$, (3.3) can be simplified to be

$$\begin{pmatrix} \mathcal{X}^T V \mathcal{X} & \mathcal{X}^T V Z \\ Z^T V \mathcal{X} & Z^T V Z \end{pmatrix}. \tag{3.4}$$

Consequently, the score statistic for testing $\gamma = 0$, $U(\hat{\beta})^T[I_{\gamma\gamma} - I_{\gamma\beta}I_{\beta\beta}^{-1}I_{\beta\gamma}]^{-1}U(\hat{\beta})$, becomes

$$S^T Z[(V^{1/2}Z)^T(I - H)V^{1/2}Z]^{-1}Z^T S \tag{3.5}$$

where $H = V^{1/2}\mathcal{X}(\mathcal{X}^T V\mathcal{X})^{-1}\mathcal{X}^T V^{1/2}$. When the dimension of γ is one, we can establish a graphical version of (3.5) for diagnostic purposes.

Let $W = V^{1/2}\mathcal{X}\hat{\beta} + V^{-1/2}S$, we contruct an approximate model

$$W = V^{1/2}\mathcal{X}\beta + V^{1/2}Z\gamma + E, \tag{3.6}$$

with $E \sim N(0, \sigma^2 I)$. The F-stastic for $\gamma = 0$ under model (3.6) is computationally equivalent to the score statistic in (3.5). Thus the added-variable plot for variable $V^{1/2}Z$, which is a plot of $R = V^{-1/2}S$ versus $(I - H)V^{1/2}Z$ (Cook and Weisberg (1982)), can be treated as a graphical version of statistic (3.5). The significant slope of the regression line in the plot corresponds to the significance of the parameter γ in the generalized nonlinear model. The influential observations in the plot would be influential on the inclusion of the parameter into the model. For later convenience, we call R and $(I - H)V^{1/2}Z$ the residuals and constructed residuals respectively. We used "constructed" since $V^{1/2}Z$ is never observed as a variable. Also, the parameter γ does not necessarily correspond to a variable. We named the plot "parameter plot."

4. Special Cases

The generalized nonlinear model can be viewed as a generalization of the GLM. Thus the tests or diagnostic plots in GLM such as the added-variable plot in Wang (1985) and score tests in Pregibon (1980) and Chen (1983) automatically become special cases of the results in the previous section. The cases we discuss in the following have not appeared, explicitly at least, in the literature.

4.1 *Constructed-variable plots in GLM*

For any x, we denote $x^{(\lambda)}$ as the family of power transformations proposed by Box and Cox (1964),

$$x^{(\lambda)} = \begin{cases} \dfrac{(x+1)^\lambda - 1}{\lambda} & \text{if} \quad \lambda \neq 0 \\ \log(x+1) & \text{if} \quad \lambda = 0. \end{cases}$$

Under the assumptions of a generalized nonlinear model on the response, we define

$$f(x_i, \beta, \gamma) = x_{i1}^T\beta_1 + (x_{i2}^{(\lambda)})^T\beta_2$$

where $\lambda = \gamma + 1$ and, $x_i^T = (x_{i1}^T, x_{i2}^T)$ and $\beta^T = (\beta_1^T, \beta_2^T)$ are both partitioned into two components: one with q elements, the other with $p - q$ elements. Under the null hypothesis $\gamma = 0$, the predictor $f(x_i, \beta, \gamma)$ becomes linear so that the model switches back to a GLM. In this case, the matrix $\mathcal{X} = [\dot{\theta}(\hat{\eta}_i) f_\beta(x_i, \hat{\beta}, 0)^T]$ is manipulated to be equal to $\mathcal{X} = DX$, with $D = \text{diag}\{\dot{\theta}(\hat{\eta}_i)\}$ and $X^T = (x_1, \ldots, x_n)$. When X is partitioned as (X_1, X_2), the same as β, we simplify Z to be $DU^{(1)T}\hat{\beta}_2$, where $U^{(\lambda)}$ is the derivative of $X_2^{(\lambda)}$ with respect to λ. Consequently, the score statistic (3.5) becomes

$$(S^T Z)^2/(V^{1/2}Z)^T(I - H)^{-1}V^{(1/2)}Z \tag{4.1}$$

with $H = V^{1/2}DX(X^T DVDX)^{-1}X^T DV^{1/2}$, which is the same as the statistic for the constructed variable plot derived by Wang (1987), but using different notations.

Notice that the constructed variable Z is different from that of Wang (1987) but the residuals and constructed residuals are the same hence, the parameter plot here is the same as the constructed-variable plot defines. The example for the application of parameter plot in this case can thus be found in Wang (1987).

4.2 *Parameter plots for links in GLM*

Pregibon (1980) suggested a family of power transformations $g(\mu; \alpha, \lambda) = \{(\mu + \alpha)^\lambda - 1\}/\lambda$ as an alternative to the identity link for normal responses, and he derived a chi-square distributed score statistic to test whether $\alpha = \lambda = 1$ which would mean that the identity link is appropriate. Here, using the identiby link with $f(x, \beta, \alpha, \lambda) = \{(x^T\beta + \alpha)^\lambda - 1\}/\lambda$ would result in the same statistic for the same purpose. When α is known to be one, we have a parameter plot for diagnosing the need of the transformed link. A similar equivalence is valid for all models in the GLM as long as the $f(x, \beta, \gamma)$ is selected correctly. For example, when the responses are binomial, we take the canonical link with $f(x, \beta, \lambda) = (\eta_i^\lambda - (1-\eta_i)^\lambda)/\lambda$ and $\eta_i = \exp(x_i^T\beta)/[1 + \exp(x_i^T\beta)]$. This is equivalent to the GLM with binomial errors and the link of

$$g(\mu; \lambda) = \frac{(\mu/n)^\lambda - 1}{\lambda} - \frac{(1 - \mu/n)^\lambda - 1}{\lambda}, \tag{4.2}$$

a subfamily of transformations given for binomial responses in Pregibon (1980). The above establishment would provide $Z = (z_i)$, a vector of $\{\log^2 \hat{\eta}_i - \log^2(1 - \hat{\eta}_i)\}/2$, $V = \text{diag}\{n_i\hat{\eta}_i(1 - \hat{\eta}_i)\}$ and $S = (s_i)$, a vector of $y_i - n_i\hat{\eta}_i$, for the statistic (3.5) and the corresponding parameter plot.

To illustrate the usefulness of the procedure for binomial response, an example from Prentice (1976) using a beetle sample is considered. The data can also be found in Pregibon (1980). The response is the number of beetles killed among those exposed to gaseous carbon disulphide with a log dosage as an explanatory variable. After fitting the GLM to (4.2) for binomial responses, we find that the statistic for $\lambda = 1$ is 7.55 with parameter plot given in Fig. 1. The indication of linearity in the figure suggests the need for a transformation, i.e. an alternative link, and no influential observation on this indication. This confirms the main result in Pregibon (1980).

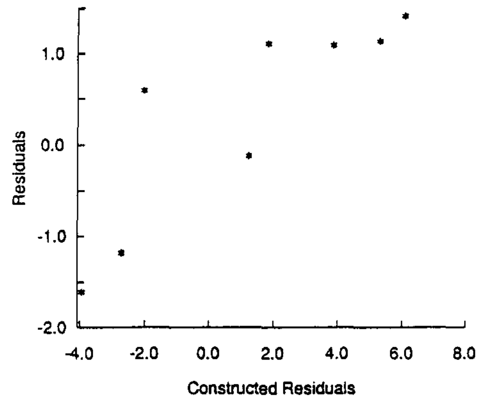

Fig. 1. Parameter plot for beetle data.

4.3 *Parameter plots for dispersions in GLM*

McCullagh (1986) suggested a method for modelling ordered categorical data with location and dispersion effects. Binomial data are a special case of categorical data and thus we use the same idea to model a GLM with dispersion effects. Use the link as usual and $f(x,z,\beta,\gamma) = x^T\beta/\exp(z^T\gamma)$ with z, a vector of observed values of some explanatory variables. For practical considerations, we extend to the case that some components of x and z might come from the same explanatory variables. β is referred to as the location parameter and γ as the dispersion parameter.

To illustrate this procedure, we used the data with binomial responses. 6 mice in a cage with an approximately fixed environment were administered at different dosage levels of chloral hydrate and ethanol. The number of animals with loss of righting reflex 30 min. after drug administration within each cage was recorded as a response. For our illustration, just data of 25 cages chosen from Carter *et al.* (1987) are used and listed in Table 1. More detailed information about the experiment and data are given in Carter *et al.* (1987).

We considered the levels of two drugs as location and dispersion explanatory variables to see if there are any dispersion effects. That is, $f(x,z,\beta,\gamma) = (\beta_0 + x_1\beta_1 + x_2\beta_2)/\exp(x_1\gamma_1 + x_2\gamma_2)$ was considered, where x_1 and x_2 are the dosage levels of two drugs injected to the animals. The score statistic for $\gamma_1 = \gamma_2 = 0$ is 6.25, which suggests the need to include these two dispersion parameters. However, the parameter plots of two parameters, given in Figs. 2(a) and (b), indicate two influential observations, 5 and 21, that might affect our conclusions. In fact, when these two points were deleted, the score statistic reduced to 3.5. Also, the impact of these two observations is expected on the physical grounds, which might be detected by careful data examination. With a fixed 200 (mg/kg) of ethanol, the dosage level of chloral hydrate obtained in cage 5 was the lowest among those cages having mice with loss of righting reflex. A similar interpretation can be given to cage 21 when compared to the cages of mice injected with 100 (mg/kg) chloral hydrate.

Table 1. Mice data.

Cage number	Ethanol (mg/kg)	Chloral hydrate (mg/kg)	Response
1	200	100	0
2	200	150	0
3	200	200	0
4	200	250	0
5	200	300	3
6	900	100	0
7	900	150	0
8	900	200	0
9	900	350	6
10	900	300	4
11	1600	100	0
12	1600	150	0
13	1600	200	5
14	1600	250	5
15	1600	300	6
16	2300	100	0
17	2300	150	4
18	2300	200	6
19	2300	250	6
20	2300	300	6
21	3000	100	6
22	3000	150	6
23	3000	200	6
24	3000	250	6
25	3000	300	6

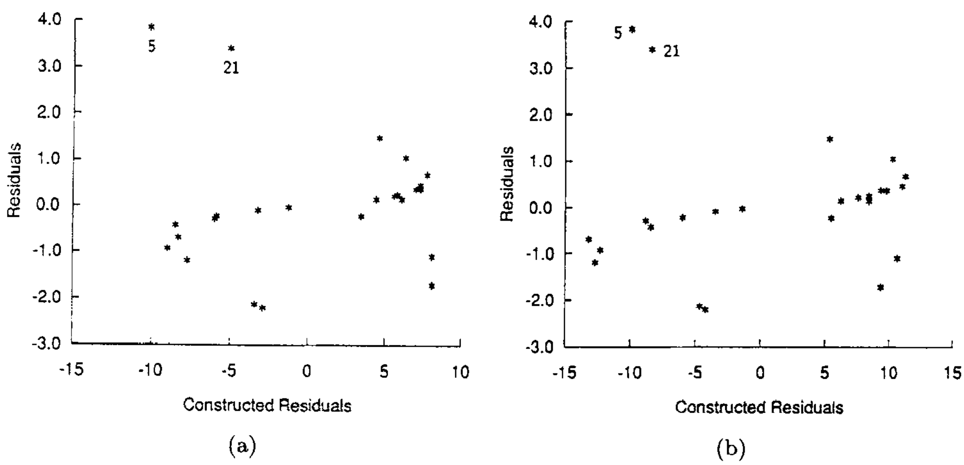

Fig. 2. (a) Parameter plot for chloral hydrate in Mice Data. (b) Parameter plot for ethanol in mice data.

4.4 *Parameter plots for nonlinear normal regression*

We use the case where the response is normal to illustrate the usefulness of the parameter plot for an arbitrary function f. In this case, we have the identity link and the null model is not a GLM, which is different from previous examples. However, the proposed procedure still works. When this simple plot is not satisfactory, Cook's more complicate procedure (1987) would give us more diagnostic information. The residuals R become the standardized residuals in the nonlinear normal regression while the constructed residuals are $(I-H)V^{1/2}Z$ with Z, a vector of $f_\gamma(x_i, \hat{\beta}, 0)$.

To illustrate parameter plots and corresponding statistics, we consider the example of Box and Hill (1974) with the model

$$y_i = \frac{\beta_0\beta_2(x_{i2} - z_i/1.632)}{1 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma z_i} + \epsilon_i \tag{4.3}$$

for $i = 1, 2, \ldots, 24$, where ϵ_i's are independent $N(0, \sigma^2)$.

Apart from using β instead of θ, γ instead of β_3, and the last variable Z, the rest of variables and case indices are the same as those given by Box and Hill who described a weighted analysis based on the linear version of (4.3) obtained using y_i^{-1} as the response. In this example, the statistic (3.5) for $\gamma = 0$ has a value 8.725 and the parameter plot given in Fig. 3. The information on γ seems to be spread throughout the data except when case 23 stands apart from the overall trend. Removing case 23 reduced $\hat{\gamma}$ by a factor of 3, but its standard error was reduced by a factor of 8. The qualitative nature of this change is, of course, consistent with the indication in Fig. 3.

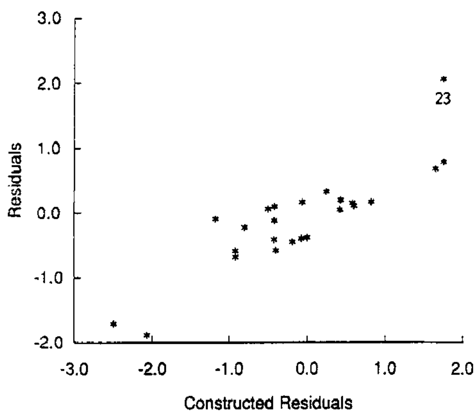

Fig. 3. Parameter plot for Box and Hill data.

From the above discussions of parameter plots on various special cases, we can see that these plots can be used in much the same way as added-variable plots in normal linear regression. Many of the useful characteristics of added-variable plots, however, carry over as only approximations in parameter plots: The slope of the regression through the origin of R on $(I - H)V^{1/2}Z$ is $\hat{\gamma}^1$, where $\hat{\gamma}^1$ is the one-step estimator of γ obtained by applying Newton-Raphson iteration with starting value $(\hat{\beta}, 0)$. In fact, the full set of one-step estimator $(\hat{\beta}^1, \hat{\gamma}^1)$ is just the vector of starting values plus the vector of the least squares estimators from the constructed model (3.6). The detailed interpretations of $\hat{\gamma}^1$ for some special cases can be obtained from cited references such as Pregibon (1980) and Wang (1987).

5. Discussion and summary

Nonlinear model problems are usually more difficult to understand than linear model problems of similar size. We have found the diagnostic procedures described in the previous sections to be useful data analytic tools for determining the importance of parameters in the generalized nonlinear model, although in particular problems, the depth of understanding provided by these methods may be substantially less than that in normal linear regression. Part of the difficulty is that the methods discussed in this paper all rely, to one degree or another, on "large" samples and the adequacy of the approximation. In particular, the slope of the regression line in an added-variable plot or parameter plot under the normal linear model is the MLE of the coefficient of the variable considered. However, under the generalized nonlinear model, even a GLM or nonlinear normal regression, this is not true due to the need of an interative procedure for gaining MLE, although the slope might be a good starting estimate.

The fitting of nonlinear normal regression can be achieved, maybe with some difficulties, by various algorithms. For those of generalized nonlinear models it might be even harder, which needs more research. Before the fitting, easy diagnostic procedures such as those proposed previously to reduce them to a GLM are worthwhile.

When a single variable or parameter is considered, the simultaneous use of the statistic and corresponding plot has been advocated by several authors, for example, Cook and Weisberg (1982) and Wang (1985). Here the same recommendation should be applied. The score statistic determines the need of the parameter considered and the significance of the slope in the parameter plot. The plot gives the overall impression of the scatter of the data and detects any influential observations on the importance of the parameter.

Acknowledgements

I would like to thank the referees and editors for their careful readings and invaluable comments that led to improvements in the presentation of this article.

References

Atkinson, A. C. (1982). Regression diagnostics, transformations and constructed variables (with discussion), *J. Roy. Statist. Soc. Ser. B*, **44**, 1–36.

Atkinson, A. C. (1983). Diagnostics regression analysis and shifted power transformations, *Technometrics*, **25**, 23–34.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation (with discussion), *J. Roy. Statist. Soc. Ser. B*, **26**, 211–246.

Box, G. E. P. and Hill, W. P. (1974). Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, **16**, 385–389.

Carter, W. H., Gennings, C., Staniswalis, J. G., Cambell, E. D. and White, K. L. (1987). A statistical approach to the construction and analysis of isobolograms (unpblished).

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

Chen, C. (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**, No. 3 (in press).

Cook, R. D. (1987). Parameter plots in nonlinear regression, *Biometrika*, **74**, 669–677.

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

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression, *Biometrica*, **70**, 1–10.

Cox, D. R., Davison, A. C., Firth, D., Nelder, J. A. and O'Brien, C. M. (1987). Generalized linear models and beyond, Skeleton Notes, Department of Mathematics, Imperial College, London.

Landwehr, J. M., Pregibon, D. and Shoemaker, A. C. (1984). Graphical methods for assessing logistic regression models (with discussion), *J. Amer. Statist. Assoc.*, **79**, 61–83.

Lawrance, A. J. (1987). The score statistic for regression transformation, *Biometrika*, **74**, 275–280.

McCullagh, P. (1986). Discussion of "Testing in industrial experiments with ordered categorical data", *Technometrics*, **28**, 307

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*, Chapman and Hall, London.

Nelder, J. A. and Wedderburn, R. W. M. (1982). Generalized linear models, *J. Roy. Statist. Soc. Ser. A*, **135**, 370–384.

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 GLIM with applications, GLIM.82: Proceedings International Conference on Generalized Linear Models, *Lecture Notes in Statistics*, **14**, Springer, New York.

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

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

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.

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