Ann. Inst. Statist. Math.
Vol. 44, No. 4, 673–686 (1992)

JACKKNIFING IN GENERALIZED LINEAR MODELS*

JUN SHAO

Department of Mathematics, University of Ottawa,
585 King Edward, Ottawa, Ontario, Canada K1N 6N5

(Received September 28, 1990; revised August 28, 1991)

Abstract. In a generalized linear model, the jackknife estimator of the asymptotic covariance matrix of the maximum likelihood estimator is shown to be consistent. The corresponding jackknife studentized statistic is asymptotically normal. In addition, these results remain true even if there exist unequal dispersion parameters in the model. On the other hand, the variance estimator and the studentized statistic based on the standard method (substitution and linearization) do not enjoy this robustness property against the presence of unequal dispersion parameters.

Key words and phrases: Asymptotic covariance matrix, consistency, jackknife, robustness.

1. Introduction

The jackknife method (Quenouille (1956), Tukey (1958)) is widely used for estimating the variance of a point estimator. If the jackknife variance estimator is consistent, the studentized statistic based on the point estimator and the jackknife variance estimator is asymptotically normal, which provides fundamentals for large sample statistical inferences. If an alternative consistent variance estimator is available (e.g., the variance estimator obtained by using standard methods), the jackknife does not have any apparent superiority in terms of asymptotic performance. However, in many cases the standard method rests upon some model assumptions, whereas the jackknife is not logically based on the same assumptions and therefore its performance is less susceptible to violation of the model assumptions. This robustness property of the jackknife was recognized by Tukey and subsequent workers. For example, Hinkley (1977) and Wu (1986) found that the jackknife variance estimator is robust against the presence of unequal error variances in linear models. In this paper we study the jackknife in a much broader model: the generalized linear model (GLM), and find that the jackknife variance estimator is robust against the presence of unequal dispersion parameters.

* This research was supported by an Operating Grant from the Natural Science and Engineering Research Council of Canada.

Examples of GLM, including useful models such as logit models, log-linear models, gamma-distributed data models and survival data models, can be found in Nelder and Wedderburn (1972) and McCullagh and Nelder (1983). A GLM has the following structure: The responses $\{y_i\}_{i=1}^n$ are independent with densities (with respect to a measure ν),

$$c(y_i, \phi_i) \exp\{[\eta_i y_i - b(\eta_i)]/\phi_i\}, \tag{1.1}$$

where ϕ_i and η_i are unknown, $\phi_i > 0$, $\eta_i \in \boldsymbol{H} = \{\eta : 0 < \int c(y, \phi) \exp\{\eta y/\phi\} d\nu < \infty\}$ for all i, and $b''(\eta) > 0$ is assumed for all $\eta \in \boldsymbol{H}^0$ (the interior of $\boldsymbol{H}$). As a consequence, $\mu_i = E(y_i) = b'(\eta_i)$ and $\sigma_i^2 = \text{Var}(y_i) = \phi_i b''(\eta_i)$. There is a known injective *link function* g such that $g(\mu_i) = x_i^\tau \beta$, where x_i is a known p-vector, x_i^τ is its transpose and β is a p-vector of unknown parameters. The function g is third-order continuously differentiable on $b'(\boldsymbol{H}^0)$. Let $\mu(\eta) = b'(\eta)$, $\xi(t) = (g \circ \mu)^{-1}(t)$, $v(t) = b''[\xi(t)]$, $h(t) = \xi'(t)$ and $\zeta(t) = [h(t)]^2 v(t)$. It is assumed that $h(t) \neq 0$.

In a GLM, the parameter of interest is β. The nuisance parameters ϕ_i are called the *dispersion* parameters and are often assumed to be of the form $\phi_i = \tilde{\phi}_i/a_i$ with unknown $\tilde{\phi}_i$ and known weights a_i (see McCullagh and Nelder ((1983), p. 21)). Without loss of generality, we assume in the sequel that $a_i = 1$ for all i, since by replacing y_i by $a_i y_i$ and $b(\eta)$ by $a_i b(\eta)$, the results in this paper can be extended to the unequal a_i situation.

Under the assumption

$$\phi_i = \phi \quad \text{for all } i, \tag{A1}$$

β can be estimated by $\hat{\beta}_n$, the maximum likelihood estimator (MLE) of β based on observations $\{y_i\}_{i=1}^n$. Fahrmeir and Kaufmann (1985) showed that the distribution of $\hat{\beta}_n - \beta$ is asymptotically normal, i.e., for any fixed p-vector $l \neq 0$,

$$l^\tau(\hat{\beta}_n - \beta)/(l^\tau V_n l)^{1/2} \xrightarrow[d]{} N(0,1) \tag{1.2}$$

with $V_n = \phi M_n^{-1}(\beta)$, the asymptotic covariance matrix of $\hat{\beta}_n$, where

$$M_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \zeta(x_i^\tau \beta).$$

If V_n has a consistent estimator $\hat{V}_n$ satisfying

$$(l^\tau \hat{V}_n l)/(l^\tau V_n l) \to 1 \tag{1.3}$$

in probability or almost surely, then the studentized statistic

$$l^\tau(\hat{\beta}_n - \beta)/(l^\tau \hat{V}_n l)^{1/2} \xrightarrow[d]{} N(0,1). \tag{1.4}$$

Any $\hat{V}_n$ satisfying (1.3) can be used to access the point estimator $\hat{\beta}_n$ and the corresponding studentized statistic in (1.4) can be used to make statistical inferences.

The jackknife estimator of V_n is

$$\hat{V}_n^J = \frac{n-p}{n}\sum_{j=1}^{n}(\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^\tau, \tag{1.5}$$

where $\hat{\beta}_{nj}$ is the MLE of β based on the data with y_j removed, $j = 1, \ldots, n$. It is shown in this paper that $\hat{V}_n^J$ is consistent in the sense of (1.3) and therefore (1.4) holds with $\hat{V}_n = \hat{V}_n^J$. Since $V_n = \phi M_n^{-1}(\beta)$, a standard substitution method for the estimation of V_n is to estimate the matrix $M_n^{-1}(\beta)$ by $M_n^{-1}(\hat{\beta}_n)$ and ϕ by a consistent estimator $\hat{\phi}$. The resulting estimator $\hat{V}_n^S = \hat{\phi} M_n^{-1}(\hat{\beta}_n)$ also satisfies (1.3) and (1.4), but it rests upon the specific form of V_n and the model assumption (A1).

Assumption (A1) is questionable in many practical situations. For example, in the classical linear models where $b''(\eta) \equiv 1$, $\phi_i = \sigma_i^2 = \mathrm{Var}(y_i)$. Hence (A1) is the same as the homoscedasticity assumption on the errors. In a GLM, (A1) is equivalent to the assumption that the variances σ_i^2 vary with the means μ_i through the functions b', b'' and g. There are, of course, other sources of variations that lead to the unequality of the variances σ_i^2. For example, if the data are collected on several days (or by several persons), then there may be differences among the variances of the data collected on different days (or by different persons). Although in practice these variations are often slight or moderate, they are difficult to control. Residual analysis can sometimes be used to detect the unequality of ϕ_i.

Knowing that the dispersion parameters are unequal, one may consider the possibility of improving the estimator $\hat{\beta}_n$ which is obtained under assumption (A1). However, if the unequality of ϕ_i is caused by day-to-day, person-to-person and batch-to-batch variations which are hard to measure, it is difficult to improve $\hat{\beta}_n$ due to lack of information. In addition, in many cases one cannot ascertain the equality or the unequality of the dispersion parameters. Consequently, it is of interest to study the robustness (against the presence of unequal ϕ_i) of the variance estimators and studentized statistics.

The jackknife does not require assumption (A1) and therefore is more likely to be robust than the standard method. This is justified in Section 2. That is, regardless of equality or lack of equality of the dispersion parameters, the jackknife variance estimator and studentized statistic satisfy (1.3) and (1.4). The corresponding estimators based on the standard method, however, do not enjoy this robustness property.

Even if assumption (A1) holds (hence $\hat{V}_n^J$ and $\hat{V}_n^S$ are asymptotically equivalent), the use of the jackknife method has the following advantages: (1) The standard method uses $M_n^{-1}(\hat{\beta}_n)$ as an estimate of $M_n^{-1}(\beta)$. Even if $M_n(\hat{\beta}_n)$ is close to $M_n(\beta)$, $M_n^{-1}(\hat{\beta}_n)$ may not be close to $M_n^{-1}(\beta)$, i.e., the speed of the convergence of $M_n^{-1}(\hat{\beta}_n)$ may be slower than that of $M_n(\hat{\beta}_n)$, especially when the method used for computing the inverse of $M_n(\hat{\beta}_n)$ is not efficient and/or $M_n(\hat{\beta}_n)$ is nearly singular. (2) Unlike the standard method, the jackknife does not require a theoretical derivation of the formula of the asymptotic variance. See also the discussion in Shao (1989).

In the presence of person-to-person variations, in addition to the unequality of ϕ_i, there is often a positive correlation among the observations. The jackknife is generally not robust against data dependence (Ghosh (1986)), but is robust when the dependence is of a special structure. For example, the covariance matrix of $(y_1, \ldots, y_n)$ is block diagonal with small block sizes (Liang and Zeger (1986)). This is further discussed in Section 2.

The jackknife method usually requires more computations than the standard method. But this is not a serious problem nowadays with a modern computer. Furthermore, making use of the result in Lemma 2.1 of Section 2, we can accelerate the computation of the jackknife estimators. This is discussed in Section 2.

2. The main results

In this section we state some asymptotic results for the jackknife estimators. All the proofs are given in the next section. To start with, we introduce some more notations. The transpose of a matrix or vector A is denoted by A^τ. For a square matrix A, its minimum eigenvalue, maximum eigenvalue, trace and inverse are denoted by $\lambda_{\min}(A)$, $\lambda_{\max}(A)$, $\mathrm{tr}(A)$ and A^{-1}, respectively. The Euclidean norm of a vector or matrix A is define to be $[\mathrm{tr}(A^\tau A)]^{1/2}$ and is denoted by $\|A\|$. For a positive definite matrix A (i.e., $A^\tau = A$ and $\lambda_{\min}(A) > 0$), let $A^{L/2}$ denote a left square root of A, i.e., $A = A^{L/2}A^{R/2}$ with $A^{R/2} = (A^{L/2})^\tau$, $A^{-L/2} = (A^{L/2})^{-1}$ and $A^{-R/2} = (A^{R/2})^{-1}$. The $p \times p$ identity matrix is denoted by I.

Let $\boldsymbol{B}$ be the admissible set for β. Throughout the paper, $\boldsymbol{N}(\epsilon)$ denote the set $\{\gamma \in \boldsymbol{R}^p : \|\gamma - \beta\| \leq \epsilon\}$ and it is assumed that $\boldsymbol{N}(\epsilon_0) \subset \boldsymbol{B}$ for a positive ϵ_0.

Under assumption (A1), the MLE $\hat\beta_n$ is a solution of $L_n(\hat\beta_n) = \max_{\gamma\in\boldsymbol{B}} L_n(\gamma)$, where $L_n(\gamma) = \sum_i\{\xi(x_i^\tau\gamma)y_i - b[\xi(x_i^\tau\gamma)]\}$ is a normalized log-likelihood function under (A1). Usually $\hat\beta_n$ is obtained by solving $s_n(\gamma) = \partial L_n(\gamma)/\partial\gamma = 0$, where $s_n(\gamma) = \sum_i x_i h(x_i^\tau\gamma)[y_i - \mu_i(\gamma)]$ is the score function and $\mu_i(\gamma) = \mu[\xi(x_i^\tau\gamma)]$.

The following assumptions are needed for the proofs of our results:

(A2) The admissible set for x_i is a compact subset of $\boldsymbol{R}^p$.

(A3) $\lambda_{\min}(D_n) \to \infty$, where $D_n = \sum_i x_i x_i^\tau$, and there exists a constant $\delta \in (0,1]$ such that $\limsup_{n\to\infty}[\lambda_{\max}(D_n)]^{(1+\delta)/2}/\lambda_{\min}(D_n) < \infty$.

Assumption (A2) is satisfied in most practical situations. A still weaker assumption can be used to replace (A2), but we omitted the discussion for simplicity (see Fahrmeir and Kaufmann (1986)). Assumption (A3) was discussed in Wu (1981) and $\lambda_{\min}(D_n) \to \infty$ was shown to be necessary and sufficient for the consistency of $\hat\beta_n$ in the classical linear model (Drygas (1976), Lai *et al.* (1979)).

Under (A1)-(A3), (1.2) holds with $V_n = \phi M_n^{-1}(\beta)$ and

$$\|V_n^{-L/2}\hat V_n^J V_n^{-R/2} - I\| \xrightarrow[\text{a.s.}]{} 0, \tag{2.1}$$

which implies $(l^\tau \hat V_n^J l)/(l^\tau V_n l) \xrightarrow[\text{a.s.}]{} 1$ and the asymptotic normality of the jackknife studentized statistic.

Under assumption (A1), a standard method for estimating V_n is to estimate $M_n^{-1}(\beta)$ by $M_n^{-1}(\hat\beta_n)$, where $M_n(\gamma) = \sum_i x_i x_i^\tau \zeta(x_i^\tau\gamma)$, and ϕ by $\hat\phi = (1 - p/n)\sum_i r_i^2$, where $r_i = [y_i - \mu_i(\hat\beta_n)]/[v(x_i^\tau\hat\beta_n)]^{1/2}$ is Pearson's residual. Under assumptions (A1)–(A3), $\hat\beta_n \underset{\text{a.s.}}{\longrightarrow} \beta$ and

$$(2.2)\qquad \begin{aligned}&\|M_n^{-L/2}(\beta)M_n(\hat\beta_n)M_n^{-R/2}(\beta) - I\|\\ &\quad\le p^{1/2}\max_{i\le n}|\zeta(x_i^\tau\hat\beta_n)/\zeta(x_i^\tau\beta) - 1| \underset{\text{a.s.}}{\longrightarrow} 0.\end{aligned}$$

In the general case where ϕ_i are not necessarily equal, $\hat\phi - (1/n)\sum_{i=1}^n \phi_i \underset{\text{a.s.}}{\longrightarrow} 0$ (see Shao (1992)), which implies $\hat\phi \underset{\text{a.s.}}{\longrightarrow} \phi$ if (A1) holds. Therefore, under assumptions (A1)–(A3), $\hat V_n^S = \hat\phi M_n^{-1}(\hat\beta_n)$ is consistent, i.e., (2.1) holds with $\hat V_n^J$ replaced by $\hat V_n^S$.

As we discussed in Section 1, in practice assumption (A1) is often violated. A more reasonable assumption is

$$(\text{A1}')\qquad 0 < \inf_i \phi_i \le \sup_i \phi_i < \infty.$$

The jackknife estimator $\hat V_n^J$ is still consistent in this case under no extra condition.

Lemma 2.1. *Under assumptions* (A1′) *and* (A2)–(A3), *there exist* $\hat\beta_{nj}$, $j = 1,\ldots,n$, $n = 1,2,\ldots$, *such that*

$$(2.3)\qquad P\{s_{nj}(\hat\beta_{nj}) = 0, j = 1,\ldots,n, \textit{ for all sufficiently large } n\} = 1,$$

where $s_{nj}(\gamma) = \partial L_{nj}(\gamma)/\partial\gamma$ *and* $L_{nj}(\gamma) = \sum_{i\ne j}\{\xi(x_i^\tau\gamma)y_i - b[\xi(x_i^\tau\gamma)]\}$, *and*

$$(2.4)\qquad \max_{j\le n}\|\hat\beta_{nj} - \beta\| \underset{\text{a.s.}}{\longrightarrow} 0.$$

Theorem 2.1. *Under assumptions* (A1′) *and* (A2)–(A3), (1.2) *and* (2.1) *hold with*

$$(2.5)\qquad V_n = M_n^{-1}(\beta)\sum_{i=1}^n x_i x_i^\tau \zeta(x_i^\tau\beta)\phi_i M_n^{-1}(\beta).$$

From Theorem 2.1, the asymptotic covariance matrix is given by (2.5), which reduces to $\phi M_n^{-1}(\beta)$ if $\phi_i \equiv \phi$. Hence in view of (2.2), the variance estimator $\hat V_n^S$ obtained by the standard method is inconsistent if assumption (A1) is violated.

Results (2.3) and (2.4) can be used in computing $\hat\beta_{nj}$, $j = 1,\ldots,n$. Note that to obtain a solution of the likelihood equation we usually need an iteration method starting with an initial point in $\boldsymbol{B}$. Since (2.4) implies $\max_{j\le n}\|\hat\beta_{nj} - \hat\beta_n\| \underset{\text{a.s.}}{\longrightarrow} 0$, $\hat\beta_n$ can be used as an initial point for obtaining $\hat\beta_{nj}$, $j = 1,\ldots,n$. This may accelerate the computation of each $\hat\beta_{nj}$ and therefore the computation of the jackknife estimators.

In some cases there is a cluster structure among the data. For example, from the i-th subject ($i = 1,\ldots,n$), n_i repeated measures, $y_{i1},\ldots,y_{in_i}$, are made. For each i, y_{it} has density (1.1), but y_{it} and y_{is} are not necessarily independent, although observations from different clusters are independent (see, e.g., Liang and Zeger (1986)). In such a case the covariance matrix of the whole data vector is a block diagonal matrix with n_i as the size of the i-th block. Suppose that we still use $\hat\beta_n$, the solution of

$$\sum_{i=1}^{n}\sum_{t=1}^{n_i} x_{it}h(x_{it}^{\tau}\gamma)[y_{it}-\mu_{it}(\gamma)] = 0,$$

to estimate β. In the case of $\phi_i \equiv \phi$, Liang and Zeger (1986) showed that $\hat\beta_n$ is asymptotically normal. Asymptotic normality of $\hat\beta_n$ in the general case of unequal ϕ_i can be established using the same argument in proving Theorem 2.1. For the jackknife variance estimation, we can still use formula (1.5) with $\hat\beta_{nj}$ defined to be a solution of

$$\sum_{i\neq j}\sum_{t=1}^{n_i} x_{it}h(x_{it}^{\tau}\gamma)[y_{it}-\mu_{it}(\gamma)] = 0.$$

Using the same argument in the proof of Theorem 2.1 (see Section 3), we can show that (2.1) holds in this case (with V_n replaced by the asymptotic covariance matrix under the assumed cluster structure).

The result in Theorem 2.1 can be extended to the case where we need to estimate $\theta = f(\beta)$, where f is a known function from $\boldsymbol{R}^p$ to $\boldsymbol{R}^q$, $q \le p$. Let $\hat\theta_n = f(\hat\beta_n)$ and $\hat\theta_{nj} = f(\hat\beta_{nj})$, $j = 1,\ldots,n$. The jackknife estimator of the asymptotic covariance matrix of $\hat\theta_n$ is still defined by (1.5) with $\hat\beta_n$ and $\hat\beta_{nj}$ replaced by $\hat\theta_n$ and $\hat\theta_{nj}$, respectively.

We assume that the gradient $\nabla f(\gamma)$ exists for $\gamma \in \boldsymbol{N}(\epsilon)$ with an $\epsilon > 0$ and is continuous at β, and that $\nabla f(\beta)$ is of full rank. Furthermore, without loss of generality we assume that f is from $\boldsymbol{R}^p$ to $\boldsymbol{R}^p$ (therefore the inverse of $\nabla f(\beta)$ exists). This is because if $f : \boldsymbol{R}^p \to \boldsymbol{R}^q$ with $q < p$, then we can find a differentiable function $f_1 : \boldsymbol{R}^p \to \boldsymbol{R}^{p-q}$ such that the inverse of $\nabla\tilde f(\beta)$ exists, where $\tilde f = (f, f_1)^{\tau}$. Let $\tilde V_n$ be the asymptotic covariance matrix of $\tilde\theta_n = \tilde f(\hat\beta_n)$. Then the upper left $q\times q$ submatrix of $\tilde V_n$ is the same as the asymptotic covariance matrix of $\hat\theta_n$. Similarly, let $\tilde\theta_{nj} = \tilde f(\hat\beta_{nj})$ and $\tilde V_n^J$ be defined as in (1.5) with $\hat\theta_n$ replaced by $\tilde\theta_n$ and $\hat\theta_{nj}$ replaced by $\tilde\theta_{nj}$. Then the upper left $q\times q$ submatrix of $\tilde V_n^J$ is $\hat V_n^J$. Thus, the consistency of $\hat V_n^J$ follows from the consistency of $\tilde V_n^J$.

Denote $\nabla f(\beta)$ by ∇f. To establish the result we need one more condition:

$$\lim_{\gamma\to\beta, n\to\infty} \|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2}\|^2 = p. \tag{2.6}$$

Note that (2.6) is satisfied if f is linear in the sense of $f(\gamma) = C\gamma$ with a fixed matrix C. Another sufficient condition for (2.6) is that $\limsup_{n\to\infty}\lambda_{\max}(D_n)/$

$\lambda_{\min}(D_n) < \infty$ (i.e., δ in (A3) equals one). This is because $\|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1} \cdot D_n^{L/2} - I\| = \|D_n^{-L/2}[\nabla f(\gamma)\nabla f^{-1} - I]D_n^{L/2}\| \leq [\lambda_{\max}(D_n)/\lambda_{\min}(D_n)]^{1/2}\|\nabla f(\gamma) \cdot \nabla f^{-1} - I\|$ and (2.6) is equivalent to

$$\lim_{\gamma\to\beta, n\to\infty} \|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2} - I\| = 0, \tag{2.7}$$

since $\|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2} - I\|^2 = \|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2}\|^2 + \|I\|^2 - 2\,\mathrm{tr}[D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2}] = \|D_n^{-L/2}\nabla f(\gamma)\nabla f^{-1}D_n^{L/2}\|^2 + p - 2\,\mathrm{tr}[\nabla f(\gamma)\nabla f^{-1}]$.

THEOREM 2.2. *Suppose that assumptions* (A1′), (A2) *and* (A3) *hold and that*

(A4) *the function f is from $\boldsymbol{R}^p$ to $\boldsymbol{R}^p$, $\nabla f(\gamma)$ exists on $\boldsymbol{N}(\epsilon)$ and is continuous at β, ∇f^{-1} exists and* (2.6) *holds.*

Then

$$V_n^{-L/2}(\hat{\theta}_n - \theta) \underset{d}{\longrightarrow} N(0, I) \tag{2.8}$$

and (2.1) *hold with*

$$V_n = \nabla f^\tau(\beta)M_n^{-1}(\beta)\sum_{i=1}^n x_i x_i^\tau \zeta(x_i^\tau\beta)\phi_i M_n^{-1}(\beta)\nabla f(\beta). \tag{2.9}$$

3. Proofs

PROOF OF LEMMA 2.1. It suffices to show that, for any $\epsilon > 0$,

$$P\{L_{nj}(\gamma) - L_{nj}(\beta) < 0, \|\gamma - \beta\| = \epsilon, j = 1, \ldots, n, \text{ for all } n \geq n_y\} = 1, \tag{3.1}$$

where n_y is an integer depending on $y_1, y_2, \ldots$ and $L_{nj}(\gamma)$ is defined in Section 2. Let $s_{nj}(\gamma) = \partial L_{nj}(\gamma)/\partial\gamma$, $H_{nj}(\gamma) = -\partial s_{nj}(\gamma)/\partial\gamma$, $s_{nj} = s_{nj}(\beta)$ and $s_n = s_n(\beta)$. From Taylor expansion,

$$L_{nj}(\gamma) - L_{nj}(\beta) = (\gamma - \beta)^\tau s_{nj} - \frac{1}{2}(\gamma - \beta)^\tau H_{nj}(\gamma_{nj})(\gamma - \beta),$$

where γ_{nj} lies between γ and β. Then (3.1) is the same as

$$P\Big\{(\gamma - \beta)^\tau s_{nj} < \frac{1}{2}(\gamma - \beta)^\tau H_{nj}(\gamma_{nj})(\gamma - \beta), \quad \|\gamma - \beta\| = \epsilon, j = 1, \ldots, n, \text{ for all } n \geq n_y\Big\} = 1. \tag{3.2}$$

Since $\|s_{nj}\| \le \|s_n\| + \|s_{nj} - s_n\| = \|s_n\| + \|x_j e_j h(x_j^\tau \beta)\| \le 2\|s_n\|$ and $\|s_n\|/[\lambda_{\max}(M_n)]^{(1+\delta)/2}$ converges to zero almost surely (Wu ((1981), Lemma 2)), where $e_i = y_i - \mu_i(\beta)$ and $M_n = M_n(\beta)$,

$$\sup_{\|\gamma-\beta\|=\epsilon} \max_{j\le n} |(\gamma-\beta)^\tau s_{nj}|/[\lambda_{\max}(M_n)]^{(1+\delta)/2} \underset{\text{a.s.}}{\longrightarrow} 0. \tag{3.3}$$

Let $M_{nj}(\gamma) = \sum_{i\ne j} x_i x_i^\tau \zeta(x_i^\tau \gamma)$ and $w_{in}(\gamma) = \zeta(x_i^\tau\gamma) x_i^\tau M_n^{-1}(\gamma) x_i$. Under (A2) and (A3), $\sup_{\gamma\in \boldsymbol{N}(\epsilon)} \max_{i\le n} w_{in}(\gamma) \to 0$. Then $1 - \sup_{\gamma\in \boldsymbol{N}(\epsilon)} \max_{i\le n} w_{in}(\gamma) \ge 1/2$ for large n and therefore $M_{nj}(\gamma) \ge M_n(\gamma)/2$ for all $j = 1,\ldots,n$ and $\gamma \in \boldsymbol{N}(\epsilon)$. Let $c_1 = \inf_{\tau\in \boldsymbol{N}(\epsilon)} \inf_i \zeta(x_i^\tau\gamma)/\zeta(x_i^\tau\beta)$. Then $c_1 > 0$ and

$$M_n(\gamma_{nj}) \ge c_1 M_n, \quad j = 1,\ldots,n, \quad \text{for all sufficiently large } n.$$

Hence $(\gamma-\beta)^\tau M_{nj}(\gamma_{nj})(\gamma-\beta) \ge \epsilon^2 c_1 \lambda_{\min}(M_n)/2$ and by (A3), for all sufficiently large n,

$$(\gamma-\beta)^\tau M_{nj}(\gamma_{nj})(\gamma-\beta) \ge c_2[\lambda_{\max}(M_n)]^{(1+\delta)/2}, \quad \|\gamma-\beta\| = \epsilon, \quad j = 1,\ldots,n, \tag{3.4}$$

where c_2 is a positive constant. Let $Z_{nj}(\gamma) = \sum_{i\ne j} x_i x_i^\tau \psi(x_i^\tau\gamma)[\mu_i(\beta) - \mu_i(\gamma)]$, where $\psi(t) = h'(t)$. Then

$$\max_{j\le n} \|Z_{nj}(\gamma)\|^2 \le \sum_{i=1}^n \|x_i\|^4 \psi^2(x_i^\tau\gamma)[\mu_i(\beta) - \mu_i(\gamma)]^2,$$

which is bounded by $c_3\lambda_{\max}(M_n)$ for all $\gamma \in \boldsymbol{N}(\epsilon)$ and a constant c_3, since (A2) holds and ψ, μ and ξ are continuous functions. Hence

$$\sup_{\tau\in \boldsymbol{N}(\epsilon)} \max_{j\le n} \|Z_{nj}(\gamma)\|^2/[\lambda_{\max}(M_n)]^{1+\delta} \to 0. \tag{3.5}$$

Let $W_{nj}(\gamma) = \sum_{i\ne j} x_i x_i^\tau \psi(x_i^\tau\gamma) e_i$ and $W_n(\gamma) = \sum_i x_i x_i^\tau \psi(x_i^\tau \gamma) e_i$. Under (A1′) and (A2)–(A3),

$$\sup_{\gamma\in \boldsymbol{N}(\epsilon)} \|W_n(\gamma)\|/[\lambda_{\max}(M_n)]^{(1+\delta)/2} \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.6}$$

and

$$\sup_{\gamma\in \boldsymbol{N}(\epsilon)} \left| \sum_{i=1}^n \|x_i\|^4 \psi^2(x_i^\tau\gamma)(e_i^2 - \sigma_i^2) \right| /[\lambda_{\max}(M_n)]^{1+\delta} \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.7}$$

by Corollary A of Wu (1981). Since $\sup_{\gamma\in \boldsymbol{N}(\epsilon)} \sum_i \|x_i\|^4 \psi^2(x_i^\tau\gamma)\sigma_i^2/[\lambda_{\max}(M_n)]^{1+\delta} \to 0$ and $\|W_{nj}(\gamma) - W_n(\gamma)\|^2 = \|x_j\|^4 \psi^2(x_i^\tau\gamma) e_j^2$,

$$\sup_{\tau\in \boldsymbol{N}(\epsilon)} \max_{j\le n} \|W_{nj}(\gamma) - W_n(\gamma)\|^2/[\lambda_{\max}(M_n)]^{1+\delta} \underset{\text{a.s.}}{\longrightarrow} 0$$

by (3.7). This together with (3.5) and (3.6) imply

$$\sup_{\gamma\in \boldsymbol{N}(\epsilon)} \max_{j\le n} \|R_{nj}(\gamma)\|^2/[\lambda_{\max}(M_n)]^{1+\delta} \to 0, \tag{3.8}$$

where $R_{nj}(\gamma) = Z_{nj}(\gamma) + W_{nj}(\gamma)$. Note that $H_{nj}(\gamma_{nj}) = M_{nj}(\gamma_{nj}) - R_{nj}(\gamma_{nj})$. From (3.4) and (3.8), there is a constant $c_4 > 0$ such that

$$P\{(\gamma-\beta)^\tau H_{nj}(\gamma_{nj})(\gamma-\beta) \ge c_4[\lambda_{\max}(M_n)]^{(1+\delta)/2},\ \|\gamma-\beta\| = \epsilon, j=1,\ldots,n, \text{ for all } n \ge n_y\} = 1,$$

which and (3.3) imply (3.2). This completes the proof.

PROOF OF THEOREM 2.1. From Shao (1992),

$$U_n^{-L/2}(\hat\beta_n - \beta) \xrightarrow[d]{} N(0,I), \tag{3.9}$$

where $U_n^{-L/2}$ is a left square root of the matrix in (2.5). Then (1.2) follows from (3.9).

From the mean value theorem and $s_{nj}(\hat\beta_{nj}) = 0$ (see (2.3)), $s_{nj}(\hat\beta_n) = \tilde H_{nj}(\hat\beta_{nj} - \hat\beta_n)$, where $\tilde H_{nj} = \int_0^1 H_{nj}(t_{nj})dt$ and $t_{nj} = \hat\beta_n + t(\hat\beta_{nj} - \hat\beta_n)$. By (2.4),

$$\sup_{t\in[0,1]} \max_{j\le n} \|t_{nj} - \beta\| \le \|\hat\beta_n - \beta\| + \max_{j\le n} \|\hat\beta_{nj} - \hat\beta_n\| \xrightarrow[\text{a.s.}]{} 0. \tag{3.10}$$

From (A2)–(A3), (3.8) and (3.10),

$$\begin{aligned}\max_{j\le n}\left\|\int_0^1 M_n^{-L/2} R_{nj}(t_{nj}) M_n^{-R/2}dt\right\| &\le \sup_{t\in[0,1]}\max_{j\le n}\|M_n^{-L/2}R_{nj}(t_{nj})M_n^{-R/2}\| \\ &\le \sup_{t\in[0,1]}\max_{j\le n}\|R_{nj}(t_{nj})\|/\lambda_{\min}(M_n) \xrightarrow[\text{a.s.}]{} 0.\end{aligned}$$

From (3.10) and (A2), there is a constant c_5 such that $\max_{j\le n}\|x_j\|^2\zeta(x_j^\tau t_{nj}) \le c_5$. Then

$$\max_{j\le n}\left\|\int_0^1 M_n^{-L/2} x_j x_j^\tau \zeta(x_j^\tau t_{nj}) M_n^{-R/2}dt\right\| \le c_5/\lambda_{\min}(M_n) \xrightarrow[\text{a.s.}]{} 0.$$

Similarly,

$$\begin{aligned}&\max_{j\le n}\left\|\int_0^1 M_n^{-L/2} M_n(t_{nj}) M_n^{-R/2}dt - I\right\| \\ &= \max_{j\le n}\left\|\int_0^1 M_n^{-L/2}[M_n(t_{nj}) - M_n] M_n^{-R/2}dt\right\| \\ &\le p^{1/2}\max_{j\le n}\sup_{t\in[0,1]}\sup_i |\zeta(x_i^\tau t_{nj})/\zeta(x_i^\tau\beta) - 1| \xrightarrow[\text{a.s.}]{} 0.\end{aligned}$$

Thus, from $H_{nj}(\gamma) = M_n(\gamma) - x_j x_j^\tau \zeta(x_j^\tau \gamma) - R_{nj}(\gamma)$, we have

$$\|M_n^{-L/2} \tilde{H}_{nj} M_n^{-R/2} - I\| \underset{\text{a.s.}}{\longrightarrow} 0. \tag{3.11}$$

Let $T_{nj} = U_n^{-L/2} M_n^{-R/2}[(M_n^{-L/2} \tilde{H}_{nj} M_n^{-R/2})^{-1} - I]$. Then

$$\max_{j \le n} \|T_{nj}\| \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.12}$$

by (3.11) and

$$\|U_n^{-L/2} M_n^{-R/2}\|^2 = \operatorname{tr}(M_n^{-L/2} U_n^{-1} M_n^{-R/2}) \le p / \inf_i \phi_i, \tag{3.13}$$

where $U_n = U_n^{L/2} U_n^{R/2}$. Let $e_i(\gamma) = y_i - \mu_i(\gamma)$. Then $s_{nj}(\hat{\beta}_n) = -x_j h(x_j^\tau \hat{\beta}_n) e_j(\hat{\beta}_n)$ since $s_n(\hat{\beta}_n) = 0$ (Lemma 2.1). Let $\alpha_{nj} = -U_n^{-L/2} M_n^{-1} x_j h(x_j^\tau \hat{\beta}_n) e_j(\hat{\beta}_n)$ and $\Delta_{nj} = -T_{nj} M_n^{-L/2} x_j h(x_j^\tau \hat{\beta}_n) e_j(\hat{\beta}_n)$. Then

$$\begin{aligned} &U_n^{-L/2}(\hat{\beta}_{nj} - \hat{\beta}_n) \\ &\quad = U_n^{-L/2} M_n^{-R/2} (M_n^{-L/2} \tilde{H}_{nj} M_n^{-R/2})^{-1} M_n^{-L/2} s_{nj}(\hat{\beta}_n) \\ &\quad = \alpha_{nj} + \Delta_{nj} \end{aligned}$$

and

$$\begin{aligned} &U_n^{-L/2} \sum_{j=1}^n (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^\tau U_n^{-R/2} \\ &\quad = \sum_{j=1}^n (\alpha_{nj}\alpha_{nj}^\tau + \Delta_{nj}\Delta_{nj}^\tau + \alpha_{nj}\Delta_{nj}^\tau + \Delta_{nj}\alpha_{nj}^\tau). \end{aligned} \tag{3.14}$$

Let $u_i = \mu_i(\beta) - \mu_i(\hat{\beta})$. Then $e_i(\hat{\beta}_n) = e_i + u_i$ and

$$\sum_{j=1}^n \alpha_{nj}\alpha_{nj}^\tau = \sum_{j=1}^n (a_{nj}a_{nj}^\tau + d_{nj}d_{nj}^\tau + a_{nj}d_{nj}^\tau + d_{nj}a_{nj}^\tau), \tag{3.15}$$

where $a_{nj} = U_n^{-L/2} M_n^{-1} x_j h(x_j^\tau \hat{\beta}) e_i$ and $d_{nj} = U_n^{-L/2} M_n^{-1} x_i h(x_j^\tau \hat{\beta}_n) u_i$. From (3.13) and the continuity of the functions μ, ξ, v and h, there is a constant c_6 such that

$$\left\| \sum_{j=1}^n d_{nj} d_{nj}^\tau \right\| \le c_6 \max_{j \le n} u_j^2 \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.16}$$

and

$$\begin{aligned} &\left\| U_n^{-L/2} M_n^{-1} \sum_{j=1}^n x_j x_j^\tau [h^2(x_j^\tau \hat{\beta}_n) - h^2(x_j^\tau \beta)] \sigma_j^2 M_n^{-1} U_n^{-R/2} \right\| \\ &\quad \le c_6 \max_{j \le n} |h^2(x_j^\tau \hat{\beta}_n)/h^2(x_j^\tau \beta) - 1| \underset{\text{a.s.}}{\longrightarrow} 0. \end{aligned}$$

From Corollary A of Wu (1981), (A2)-(A3) and (3.13),

$$\left\| U_n^{-L/2} M_n^{-1} \sum_{j=1}^{n} x_j x_j^\tau h^2(x_j^\tau \hat{\beta}_n)(e_j^2 - \sigma_j^2) M_n^{-1} U_n^{-R/2} \right\| \underset{\text{a.s.}}{\longrightarrow} 0.$$

Since $\sum_j a_{nj} a_{nj}^\tau = U_n^{-L/2} M_n^{-1} \sum_j x_j x_j^\tau h^2(x_j^\tau \hat{\beta}_n) e_j^2 M_n^{-1} U_n^{-R/2}$,

$$\left\| \sum_{j=1}^{n} a_{nj} a_{nj}^\tau - I \right\| \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.17}$$

follows from $U_n^{-L/2} M_n^{-1} \sum_j x_j x_j^\tau h^2(x_j^\tau \beta) \sigma_j^2 M_n^{-1} U_n^{-R/2} = I$. Let l and $\tilde{l}$ be fixed p-vectors. By Cauchy-Schwarz inequality and (3.16)-(3.17),

$$\begin{aligned} \left(\sum_{j=1}^{n} l^\tau a_{nj} d_{nj}^\tau \tilde{l} \right)^2 &\le \sum_{j=1}^{n} (l^\tau a_{nj})^2 \sum_{j=1}^{n} (d_{nj}^\tau \tilde{l})^2 \\ &= \left(l^\tau \sum_{j=1}^{n} a_{nj} a_{nj}^\tau l \right) \left(\tilde{l}^\tau \sum_{j=1}^{n} d_{nj} d_{nj}^\tau \tilde{l} \right) \underset{\text{a.s.}}{\longrightarrow} 0. \end{aligned} \tag{3.18}$$

From (3.15)-(3.18),

$$\left\| \sum_{j=1}^{n} \alpha_{nj} \alpha_{jn}^\tau - I \right\| \underset{\text{a.s.}}{\longrightarrow} 0. \tag{3.19}$$

From (3.12) and (3.19),

$$\left\| \sum_{j=1}^{n} \Delta_{nj} \Delta_{nj}^\tau \right\| \le \max_{j \le n} \|T_{nj}\|^2 \sup_i \phi_i \sum_{j=1}^{n} \alpha_{nj}^\tau \alpha_{nj} \underset{\text{a.s.}}{\longrightarrow} 0.$$

Then (3.18) holds with a_{nj} and d_{nj} replaced by α_{nj} and Δ_{nj}, respectively. Thus, by (3.14) and $(n-p)/n \to 1$, (2.1) holds with V_n given by (2.5).

PROOF OF THEOREM 2.2. Since any left square root of the matrix V_n in (2.9) is equal to $\nabla f^\tau U_n^{L/2} P_n$ with an orthogonal matrix P_n, where $U_n^{-L/2}$ is given in the proof of Theorem 2.1, (2.8) follows from

$$U_n^{-L/2} (\nabla f^\tau)^{-1} (\hat{\theta}_n - \theta) \underset{d}{\longrightarrow} N(0, I). \tag{3.20}$$

From the mean value theorem for vector valued functions, $\hat{\theta}_n - \theta = F_n^\tau (\hat{\beta}_n - \beta)$ with $F_n = \int_0^1 \nabla f(\beta + t(\hat{\beta}_n - \beta)) dt$. Then $\Gamma_n^\tau U_n^{-L/2} (\hat{\beta}_n - \beta) = U_n^{-L/2} (\nabla f^\tau)^{-1} (\hat{\theta}_n - \theta)$, where $\Gamma_n = U_n^{R/2} F_n \nabla f^{-1} U_n^{-R/2}$. By Theorem 2.1, (3.20) holds if $\|\Gamma_n - I\| \underset{\text{a.s.}}{\longrightarrow} 0$. From (A1′) and (A2), there is a constant c_7 such that

$$\|U_n^{R/2} D_n^{L/2}\|^2 = \text{tr}(D_n^{L/2} U_n D_n^{R/2}) \leq p \sup_i \phi_i / \inf_i \zeta(x_i^\tau \beta) \leq c_7$$

and

$$\|D_n^{-L/2} U_n^{-R/2}\|^2 = \text{tr}(D_n^{-L/2} U_n^{-1} D_n^{-R/2}) \leq p \sup_i \zeta(x_i^\tau \beta) / \inf_i \phi_i \leq c_7.$$

Since $\|\beta + t(\hat{\beta}_n - \beta) - \beta\| = t\|\hat{\beta}_n - \beta\| \leq \|\hat{\beta}_n - \beta\| \underset{\text{a.s.}}{\longrightarrow} 0$ by Lemma 2.1,

$$\|D_n^{-L/2} \nabla f(\beta + t(\hat{\beta}_n - \beta)) \nabla f^{-1} D_n^{L/2} - I\| \underset{\text{a.s.}}{\longrightarrow} 0 \quad \text{uniformly in} \quad t \in [0,1]$$

under assumption (A4) (see (2.7)). Then

$$\begin{aligned} \|\Gamma_n - I\| &= \|U_n^{R/2} D_n^{L/2} D_n^{-L/2} (F_n \nabla f^{-1} - I) D_n^{L/2} D_n^{-L/2} U_n^{-R/2}\| \\ &\leq c_7 \|D_n^{-L/2} (F_n \nabla f^{-1} - I) D_n^{L/2}\| \\ &\leq c_7 \int_0^1 \|D_n^{-L/2} \nabla f(\beta + t(\hat{\beta}_n - \beta)) \nabla f^{-1} D_n^{L/2} - I\| dt \underset{\text{a.s.}}{\longrightarrow} 0. \end{aligned}$$

This proves (2.8).

From the mean value theorem, $\hat{\theta}_{nj} - \hat{\theta}_n = G_{nj}^\tau (\hat{\beta}_{nj} - \hat{\beta}_n)$, where $G_{nj} = \int_0^1 \nabla f(t_{nj}) dt$ and $t_{nj} = \hat{\beta}_n + t(\hat{\beta}_{nj} - \hat{\beta}_n)$. From (A4) and (3.10),

$$\max_{j \leq n} \|\Lambda_{nj}\| \leq c_4 \max_{j \leq n} \|D_n^{-L/2} (G_{nj} \nabla f^{-1} - I) D_n^{L/2}\| \underset{\text{a.s.}}{\longrightarrow} 0,$$

where $\Lambda_{nj} = U_n^{R/2} G_{nj} \nabla f^{-1} U_n^{-R/2} - I$. Then by Theorem 2.1,

$$\left\| \sum_{j=1}^n \Lambda_{nj}^\tau U_n^{-L/2} (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^\tau U_n^{-R/2} \right\| \underset{\text{a.s.}}{\longrightarrow} 0 \tag{3.21}$$

and therefore

$$U_n^{-L/2} (\nabla f^\tau)^{-1} \sum_{j=1}^n G_{nj}^\tau (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^\tau U_n^{-R/2} - I \underset{\text{a.s.}}{\longrightarrow} 0. \tag{3.22}$$

From the definition of the jackknife estimator $\hat{V}_n^J$,

$$\begin{aligned} &\frac{n}{n-p} U_n^{-L/2} (\nabla f^\tau)^{-1} \hat{V}_n^J \nabla f^{-1} U_n^{-R/2} \\ &\quad = \sum_{j=1}^n U_n^{-L/2} (\nabla f^\tau)^{-1} G_{nj}^\tau (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^\tau G_{nj} \nabla f^{-1} U_n^{-R/2} \end{aligned}$$

$$= \sum_{j=1}^{n} \Lambda_{nj}^{\tau} U_n^{-L/2} (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^{\tau} U_n^{-R/2} \Lambda_{nj}$$

$$- \sum_{j=1}^{n} U_n^{-L/2} (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^{\tau} U_n^{-R/2}$$

$$+ \sum_{j=1}^{n} U_n^{-L/2} (\nabla f^{\tau})^{-1} G_{nj}^{\tau} (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^{\tau} U_n^{-R/2}$$

$$+ \sum_{j=1}^{n} U_n^{-L/2} (\hat{\beta}_{nj} - \hat{\beta}_n)(\hat{\beta}_{nj} - \hat{\beta}_n)^{\tau} G_{nj} \nabla f^{-1} U_n^{-R/2}.$$

Then from Theorem 2.1 and (3.21)-(3.22),

$$\|U_n^{-L/2} (\nabla f^{\tau})^{-1} \hat{V}_n^J \nabla f^{-1} U_n^{-R/2} - I\| \underset{\text{a.s.}}{\longrightarrow} 0. \tag{3.23}$$

This proves (2.1) if $V_n^{L/2} = \nabla f^{\tau} U_n^{L/2}$, where V_n is given by (2.9). For arbitrary left square root $V_n^{L/2}$, there is an orthogonal matrix P_n such that $V_n^{L/2} = \nabla f^{\tau} U_n^{L/2} P_n$. Then

$$V_n^{-L/2} \hat{V}_n^J V_n^{-R/2} - I = P_n^{\tau} [U_n^{-L/2} (\nabla f^{\tau})^{-1} \hat{V}_n^J \nabla f^{-1} U_n^{-R/2} - I] P_n.$$

Hence the result follows from (3.23) and $\|P_n\|^2 = p$. This completes the proof.

Acknowledgements

I wish to thank the referees for their helpful suggestions and comments.

References

Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **34**, 119–127.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Ann. Statist.*, **13**, 342–368.

Fahrmeir, L. and Kaufmann, H. (1986). Asymptotic inferences in discrete response models, *Statistical Papers*, **27**, 179–205.

Ghosh, M. (1986). Discussion of C. F. J. Wu's paper, *Ann. Statist.*, **14**, 1308–1310.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations, *Technometrics*, **19**, 285–292.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Strong consistency of least squares estimators in multiple regression, *J. Multivariate Anal.*, **9**, 343–361.

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

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

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *J. Roy. Statist. Soc. Ser. A*, **135**, 370–384.

Quenouille, M. (1956). Notes on bias in estimation, *Biometrika*, **43**, 353–360.

Shao, J. (1989). Jackknifing weighted least squares estimators, *J. Roy. Statist. Soc. Ser. B*, **51**, 139–156.

Shao, J. (1992). Asymptotic theory in generalized linear models with nuisance scale parameters, *Probab. Theory Related Fields*, **91**, 25–41.

Tukey, J. (1958). Bias and confidence in not quite large samples, *Ann. Math. Statist.*, **29**, 614.

Wu, C. F. J. (1981). Asymptotic theory of nonlinear least squares estimation, *Ann. Statist.*, **9**, 501–513.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion), *Ann. Statist.*, **14**, 1261–1350.