A LARGE SAMPLE THEORY IN GENERALIZED LINEAR MODELS WITH NUISANCE SCALE PARAMETERS *

by

Jun Shao
Purdue University

Technical Report #88-11

Department of Statistics
Purdue University

March 1988

* This research was supported in part by the NSF–AFOSR grant ISSA–860068.

⋆

A LARGE SAMPLE THEORY IN GENERALIZED LINEAR MODELS WITH NUISANCE SCALE PARAMETERS [1]

Jun Shao

Purdue University

Abstract

This paper establishes the asymptotic normality and the consistency-robustness of the weighted least squares estimator (WLSE) in the generalized linear models with multiple nuisance scale parameters. In addition, noting that the asymptotic robust statistical inference in presence of nuisance scale parameters requires a consistency-robust estimator of the asymptotic covariance matrix of the WLSE, this paper derives a class of covariance estimators and proves their consistency-robustness.

AMS 1980 *subject classifications.* Primary 62J02, 62F12.

Key word and phrases: Weighted least squares, nuisance parameters, asymptotic normality, asymptotic covariance matrix, consistency-robustness.

[1] This research was supported in part by the NSF-AFOSR grant ISSA-860068.

1. Introduction

The generalized linear model (GLIM) is characterized by the following structure (see Nelder and Wedderburn (1972), McCullagh and Nelder (1983)):

(i) The responses $\{ y_i \}_{i=1}^{\infty}$ are independent with densities

$$c(y_i, \phi_i)\exp\{\phi_i^{-1}[\theta_i y_i - b(\theta_i)]\} \tag{1.1}$$

with respect to a σ-finite measure ν, where ϕ_i and θ_i are unknown parameters, $0< inf_i \phi_i \leq sup_i \phi_i < \infty$, $\theta_i \in \Theta$ for all i and $\Theta=\{ \theta: 0<\int c(y, \phi)\exp\{\phi^{-1}\theta y\}d\nu<\infty \}$. Consequently,

$$\mu_i = E(y_i) = b'(\theta_i) \quad \text{and} \quad \sigma_i^2 = Var(y_i) = \phi_i b''(\theta_i). \tag{1.2}$$

(ii) The mean $\mu_i=\mu(\theta_i)$ is related to the linear combination $x_i^\tau\beta$ by an injective *link function* g: $\mathbf{M}\to\mathbf{R}$, $x_i^\tau\beta=g(\mu_i)$, where the regressors $\{ x_i \}_{i=1}^{\infty}$ are known p-vectors, β is a p-vector of unknown parameters, x^τ is the transpose of the vector x, $\mathbf{M}=\mu(\Theta^0)$ and Θ^0 is the interior of Θ and is assumed to be nonempty.

Examples of GLIM can be found in McCullagh and Nelder (1983). In a GLIM, β is usually the parameter of interest and has to be estimated from a finite sample of n observations $y_1,..., y_n$. The unknown scale factors ϕ_i are *nuisance* parameters.

In the special case of $\phi_i=\phi$ for all i, the maximum likelihood estimator (MLE) of β, to be denoted by $\hat{\beta}_n$, is a solution of the log likelihood equation

$$H^\tau V^{-1}(y-\mu) = 0, \tag{1.3}$$

where $y=(y_1 \ldots y_n)^\tau$, $\mu=(\mu_1(\beta) \ldots \mu_n(\beta))^\tau$, $\mu_i(\beta)=g^{-1}(x_i^\tau\beta)$, $V=diag[\,b''(\theta_i)\,]_{n\times n}$ and $H=(\partial\mu_1/\partial\beta \ldots \partial\mu_n/\partial\beta)^\tau$. In this case, the MLE of β is equivalent to the weighted least squares estimator (WLSE) (Bradley (1973)). In addition, $\hat{\beta}_n$ possesses certain desirable properties which make large sample statistical inference possible. That is, under some regularity conditions, $\hat{\beta}_n$ is consistent and asymptotically normal (Fahrmeir and Kaufmann (1985)).

Often in practice either ϕ_i's are unequal or one can not ascertain their equality. In either case, due to lack of information, one can not obtain good estimates of ϕ_i's. Consequently, it is tempting to overlook the presence of unequal nuisance parameters and make statistical inferences based on a solution $\hat{\beta}_n$ of (1.3), which does not make use of any information about

ϕ_i's. It is desirable to study robustness of statistical inferences based on $\hat{\beta}_n$ in presence of unequal nuisance parameters. To do so, this paper presents an asymptotic theory of $\hat{\beta}_n$ in the GLIM with nuisance scale parameters.

When ϕ_i's are unequal, $\hat{\beta}_n$ is not necessarily the MLE but the WLSE or the generalized least squares estimator (GLSE), since the true log likelihood equation is

$$H^{\tau}\Phi^{-1}V^{-1}(y-\mu) = 0,$$

where $\Phi = diag[\ \phi_i\ ]_{n\times n}$.

The paper is organized as follows.

In Section 2, the notations and assumptions are discussed. In Section 3, it is shown that despite the nuisance parameters, the asymptotic distribution of $\hat{\beta}_n$ remains normal. This result rests upon the relative stability of the central limit theorem. The *consistency-robustness* of $\hat{\beta}_n$ is also established in Section 3. (We say that an estimator is consistency-robust iff the estimator is consistent no matter whether ϕ_i are equal or not.) These results may still hold even if the distribution of y_i are not completely specified as in (1.1). See the discussion in Section 3.

The results in Section 3 show that the mean of the asymptotic distribution of $\hat{\beta}_n$ is the same for both equal and unequal ϕ_i's. However, the covariance matrix of the asymptotic distribution of $\hat{\beta}_n$, which will be called the asymptotic covariance matrix in the sequel, is affected by the unequality of ϕ_i's. Consequently, for robust statistical inference based on $\hat{\beta}_n$, it is crucial to find a consistency-robust estimator of the asymptotic covariance matrix. The asymptotic covariance matrix is shown to be of the following form:

$$\Sigma_n = \sum_{i=1}^n l_i(\beta)\sigma_i^2,$$

where $l_i(\beta)$, $i=1,\ldots,n$, are matrix functions of β. In Section 4, we will propose and present justifications for estimators of Σ of the following form:

$$\hat{\Sigma}_n = \sum_{i=1}^n l_i(\hat{\beta}_n)w_i r_i^2,$$

where w_i's are positive constants and $r_i = y_i - \mu_i(\hat{\beta}_n)$ is the ith residual. In particular, $\hat{\Sigma}_n$ is shown to be the same as the estimators obtained by using the *heteroscedastic bootstrap* (Beran (1986) and Efron (1986)) and the *linear jackknife* (Fox, Hinkley and Larntz (1980)). The consistency-robustness of $\hat{\Sigma}_n$ is proved in Section 5.

In the special case of $\phi_i \equiv \phi$, Σ_n reduces to $\phi\sum_{i=1}^n l_i(\beta)v_i(\beta)$, where $v_i(\beta)=b''(\theta_i)$. An alternative estimator is then $\hat{\phi}_n\sum_{i=1}^n l_i(\hat{\beta}_n)v_i(\hat{\beta}_n)$, where $\hat{\phi}_n=n^{-1}\sum_{i=1}^n w_i r_i^2/v_i(\hat{\beta}_n)$. The consistency of this estimator is studied in Section 6.

2. Notations and assumptions

Throughout the paper, the minimum eigenvalue, the maximum eigenvalue, the transpose, and the trace of a $p \times p$ matrix A are denoted by $\lambda_{\min}(A)$, $\lambda_{\max}(A)$, A^τ, and $tr(A)$, respectively. The Euclidean norm of A is defined to be $\|A\|=[tr(A^\tau A)]^{1/2}$. For any positive definite matrix A, let $A^{1/2L}$ ($A^{1/2R}$) be a left (the corresponding right) square root of A, i.e., $A=A^{1/2L}A^{1/2R}$, $(A^{1/2L})^\tau=A^{1/2R}$. Define $A^{-1/2L}=(A^{1/2L})^{-1}$ and $A^{-1/2R}=(A^{1/2R})^{-1}$. The left (right) square roots are unique up to an orthogonal transformation from the right (from the left). Note that $A^{1/2L}$ is not necessarily symmetric. When $A^{1/2L}$ is symmetric, we write $A^{1/2}=A^{1/2L}=A^{1/2R}$.

We will denote the *true* but unknown parameter by β_0 and the admissible set of the regressors x_i by **X**, i.e., $x_i \in \mathbf{X}$ for all i. Let $\mu(\theta)=b'(\theta)$ and $v(\theta)=b''(\theta)$ (when $\theta \in \Theta^0$, all derivatives of $b(\theta)$ exist). If $v(\theta)>0$ for $\theta \in \Theta^0$, μ restricted to Θ^0 is an injective function. Let $\eta(t)=\mu^{-1}[g^{-1}(t)]$, where g is the link function, $h(t)=\eta'(t)$ and $\zeta(t)=[h(t)]^2 v[\eta(t)]$. Throughout the paper, $\mathbf{N}(\varepsilon)$ denotes the set $\{ \beta\colon \|\beta-\beta_0\| \le \varepsilon \}$ for a positive ε.

The modeling assumptions are as follows:

(M1) $v(\theta)>0$ for $\theta \in \Theta^0$, η is twice continuously differentiable and $h(t)\neq 0$.

(M2) The admissible set for parameter β, denoted by **B**, is the interior of

$$\{ \beta\colon \eta(x^\tau\beta) \in \Theta^0 \text{ for } all\ x \in \mathbf{X} \}.$$

(M3) $0 < inf_i\, \zeta(x_i^\tau\beta_0) \le sup_i\, \zeta(x_i^\tau\beta_0) < \infty$.

When $\eta(t)\equiv t$ (i.e., η is the identity function), g is called the natural link function and **B** is convex.

Let $\zeta_i(\beta)=\zeta(x_i^\tau\beta)$, $v_i(\beta)=v[\eta(x_i^\tau\beta)]$, $h_i(\beta)=h(x_i^\tau\beta)$,

$$D_n = \sum_{i=1}^n x_i x_i^\tau,$$

$$M_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \zeta_i(\beta), \tag{2.1}$$

and

$$F_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \zeta_i(\beta)\phi_i. \tag{2.2}$$

Denote $M_n(\beta_0)$ and $F_n(\beta_0)$ by M_n and F_n, respectively.

The design $\{ x_i \}_{i=1}^n$ assumptions are as follows:

(D1) D_n is positive definite for sufficiently large n and

$$\lim_{n\to\infty} \max_{i\le n} x_i^\tau D_n^{-1} x_i = 0. \tag{2.3}$$

(D2) There exists a constant $\delta\in(0,1]$ such that

$$limsup_{n\to\infty} [\lambda_{\max}(D_n)]^{½(1+\delta)}/\lambda_{\min}(D_n) < \infty.$$

(D3) There exists a constant $\alpha > 0$ such that

$$limsup_{n\to\infty} \max_{i\le n} \|x_i\|^2/[\lambda_{\max}(D_n)]^{(\delta-\alpha)/(1+\alpha)} < \infty, \quad \|x\|=(x^\tau x)^{½}.$$

Note that equation (2.3) is the Lindeberg's condition for the classical linear model ($\mu_i = x_i^\tau\beta$). It implies that

$$\lim_{n\to\infty} \lambda_{\min}(D_n) = \infty, \tag{2.4}$$

which is a necessary and sufficient condition for the consistency of $\hat\beta_n$ in the classical linear model. (D1) and (M3) together imply that

$$\lim_{n\to\infty} \max_{i\le n} \zeta_i(\beta_0) x_i^\tau M_n^{-1} x_i = 0.$$

Assumption (D2) was discussed in Wu (1981). (D2) and (M3) together imply that

$$limsup_{n\to\infty} [\lambda_{\max}(M_n)]^{½(1+\delta)}/\lambda_{\min}(M_n) < \infty.$$

Fahrmeir and Kaufmann (1985) proved the asymptotic normality of $\hat\beta_n$ for $\phi_i \equiv \phi$ under the following continuity assumption:

(C) $\|M_n^{-½L} M_n(\beta) M_n^{-½R} - I\| \to 0$ as $n\to\infty$ and $\beta\to\beta_0$,

where I is the $p\times p$ identity matrix. Since (C) involves the parameter β, it is desirable to establish sufficient conditions for (C) in terms of the functions v and η and the design. Note that

$$M_n(\beta)-M_n = \sum_{i=1}^{n} x_i x_i^{\tau} \zeta_i(\beta_0)[\zeta_i(\beta)/\zeta_i(\beta_0)-1]$$

and

$$\|M_n^{-\frac{1}{2}L} M_n(\beta) M_n^{-\frac{1}{2}R} - I\| = \|M_n^{-\frac{1}{2}L}(M_n(\beta)-M_n)M_n^{-\frac{1}{2}R}\|$$

$$\le \max_{i\le n} |\zeta_i(\beta)/\zeta_i(\beta_0)-1| \; \|M_n^{-\frac{1}{2}L} M_n M_n^{-\frac{1}{2}R}\| = p^{\frac{1}{2}} \max_{i\le n} |\zeta_i(\beta)/\zeta_i(\beta_0)-1|.$$

Hence (C) is implied by

$$\lim_{n\to\infty,\, \beta\to\beta_0} \max_{i\le n} |\zeta_i(\beta)/\zeta_i(\beta_0)-1| = 0,$$

which is implied by (M3) and

$$\{\ \zeta_i(\beta)\ \}_{i=1}^{\infty} \text{ is equicontinuous at } \beta_0. \tag{2.5}$$

When the admissible set of regressors $\mathbf{X}$ is a compact subset of $\mathbf{R}^p$, the above conditions simplify considerably. Since $\{\ x_i^{\tau}\beta_0\ \}_{i=1}^{\infty}$ is compact and $sup_i \|x_i\|<\infty$ when $\mathbf{X}$ is compact, (M3) and (D3) are satisfied and (D1) is implied by (2.4). From the continuity of ζ, the compactness of $\mathbf{X}$ also implies (2.5) and therefore (C) holds.

The asymptotic results in this paper are established under one of the following three groups of assumptions:

(A1) Assumptions (M1)-(M3), (D1) and (C).

(A2) Assumptions (M1)-(M3), (D1)-(D3) and (C).

(A3) $\mathbf{X}$ is compact and Assumptions (M1), (M2), (D1) and (D2).

From the above discussion, Assumption (A3) is stronger than (A1) and (A2).

We need to discuss one more condition before stating the main results. Let $\{\ f_i\ \}_{i=1}^{\infty}$ be a sequence of functions defined on $\mathbf{B}$ and $\mathbf{C}$ be a compact subset of $\mathbf{B}$. We say that $\{\ f_i\ \}_{i=1}^{\infty}$ is bounded on $\mathbf{C}$ if

$$sup_i\, sup_{\beta\in\mathbf{C}} |f_i(\beta)| < \infty,$$

and that $\{\ f_i\ \}_{i=1}^{\infty}$ is Lipschitz continuous on $\mathbf{C}$ if

$$sup_{\beta\ne\gamma,\ \beta,\ \gamma\in\mathbf{C}} \frac{|f_i(\beta)-f_i(\gamma)|}{\|\beta-\gamma\|} \le M\, sup_{\beta\in\mathbf{C}} |f_i(\beta)|$$

for all i, where M is a constant independent of i.

Let f be defined on $\eta^{-1}(\Theta^0)$ and $f_i(\beta)=f(x_i^\tau\beta)$. If f is continuously differentiable on N(ε) and **X** is compact, then $\{ f_i(\beta) \}_{i=1}^{\infty}$ is bounded on N(ε) and

$$sup_{\beta\neq\gamma,\, \beta,\, \gamma\in N(\varepsilon)} \frac{|f(x_i^\tau\beta)-f(x_i^\tau\gamma)|}{\|\beta-\gamma\|} \leq sup_{\beta\in N(\varepsilon)} |f'(x_i^\tau\beta)|\ \|x_i\| \leq M\ sup_{\beta\in N(\varepsilon)} |f(x_i^\tau\beta)|$$

with

$$M = sup_i sup_{\beta\in N(\varepsilon)} |f'(x_i^\tau\beta)|\ sup_i \|x_i\| \,/\, inf_i sup_{\beta\in N(\varepsilon)} |f(x_i^\tau\beta)|,$$

provided $inf_i sup_{\beta\in N(\varepsilon)} |f(x_i^\tau\beta)|>0$. Hence $\{ f_i(\beta) \}_{i=1}^{\infty}$ is Lipschitz continuous on N(ε).

3. Asymptotic normality and consistency-robustness of $\hat{\beta}_n$

The technique used in our proof of the asymptotic normality of $\hat{\beta}_n$ is different from Fahrmeir and Kaufmann's (1985, Theorem 3). Their proof relies on the assumption that the distribution of y_i belongs to the exponential family, whereas our results hold as long as

$$d\mu(\theta)/d\theta = v(\theta), \tag{3.1}$$

by examining the proofs of Theorems 1 and 2. In fact, the density defined in (1.1) does not belong to the exponential family. Our results hold for a general case where the log likelihood of y, denoted by $l(\mu, y)$, is determined by the system of partial differential equations

$$\partial l(\mu, y)/\partial\mu = \Sigma^{-1}(y-\mu),$$

where $\Sigma=Var(y)$. In this case (1.3) is only a log quasi-likelihood equation (see Wedderburn (1974) and McCullagh (1983)). Our design assumption is slightly stronger than Fahrmeir and Kaufmann's, i.e., they assume (2.4), which is implied by our assumption (D1). However, since (D1) is equivalent to (2.4) and

$$\lim_{n\to\infty} x_n^\tau D_n^{-1} x_n = 0,$$

the difference between (D1) and (2.4) is small. In the classical linear models, (D1) is also necessary for the asymptotic normality of $\hat{\beta}_n$ (Huber (1981)).

We start with the case where g is the natural link function.

3.1. Natural link functions. Let $e_i(\beta)= y_i-\mu_i(\beta)$ and $e_i=e_i(\beta_0)$. Equation (1.3) becomes

$$s_n(\beta) = \sum_{i=1}^n x_i h_i(\beta) e_i(\beta) = 0,$$

where $s_n(\beta)$ is called the score function. Let $-H_n(\beta)$ be a $p\times p$ matrix whose ith row is the gradient of the ith component of $s_n(\beta)$. When g is the natural link function, $h(t)\equiv 1$ and $s_n(\beta)$ reduces to $\sum_{i=1}^n x_i e_i(\beta)$, and

$$H_n(\beta) = M_n(\beta),$$

where $M_n(\beta)$ is defined in (2.1). Since $M_n(\beta)$ is positive definite for large n (under (D1)), if $\hat{\beta}_n$ exists then it is uniquely defined (as a solution of (1.3)). An slight modification of the proof of Theorem 1 of Fahrmeir and Kaufmann (1985) leads us to the following result.

Lemma 1. Under Assumption (A1), there exists a sequence of random variables $\hat{\beta}_n$ such that

$$P(\, s_n(\hat{\beta}_n) = 0 \,) \rightarrow 1 \tag{3.2}$$

and $\hat{\beta}_n$ is weakly consistent, i.e.,

$$\hat{\beta}_n \rightarrow_p \beta_0 , \tag{3.3}$$

where $\rightarrow_p$ denotes convergence in probability.

Theorem 1. Under Assumption (A1),

$$\Sigma_n^{-\frac{1}{2}L}(\hat{\beta}_n-\beta_0) \rightarrow_d N(0, I), \tag{3.4}$$

where $\rightarrow_d$ denotes convergence in distribution and

$$\Sigma_n = M_n^{-1} F_n M_n^{-1}. \tag{3.5}$$

Proof. Denote $s_n(\beta_0)$ by s_n. There is an $\varepsilon > 0$ such that $\mathbf{N}(\varepsilon)\subset\mathbf{B}$. From Lemma 1, $P(\, s_n(\hat{\beta}_n)=0$ and $\hat{\beta}_n \in \mathbf{N}(\varepsilon)\,)\rightarrow 1$. Hence we focus on the set $\{\, s_n(\hat{\beta}_n)=0$ and $\hat{\beta}_n \in \mathbf{N}(\varepsilon)\,\}$. By the mean-value theorem,

$$s_n = M_n(\beta_n^*)(\hat{\beta}_n - \beta_0),$$

where β_n^* is on the line segment between $\hat{\beta}_n$ and β_0. Then

$$M_n^{-\frac{1}{2}L} s_n = M_n^{-\frac{1}{2}L} M_n(\beta_n^*) M_n^{-\frac{1}{2}R} M_n^{\frac{1}{2}R}(\hat{\beta}_n - \beta_0).$$

From (1.2) and (2.2),

$$E(M_n^{-\frac{1}{2}L} s_n s_n^\tau M_n^{-\frac{1}{2}R}) = M_n^{-\frac{1}{2}L} \sum_{i=1}^n x_i x_i^\tau h_i^2(\beta_0) E e_i^2 M_n^{-\frac{1}{2}R} = M_n^{-\frac{1}{2}L} F_n M_n^{-\frac{1}{2}R} \le (sup_i \phi_i)\, I.$$

Hence $M_n^{-\frac{1}{2}L} s_n = O_p(1)$. By Lemma 1 and (C), $M_n^{-\frac{1}{2}L} M_n(\beta_n^*) M_n^{-\frac{1}{2}R} \to_p I$ and therefore

$$M_n^{\frac{1}{2}R}(\hat{\beta}_n - \beta_0) = M_n^{-\frac{1}{2}L} s_n + o_p(1). \tag{3.6}$$

Since $(inf_i \phi_i) M_n^{-1} \le \Sigma_n \le (sup_i \phi_i) M_n^{-1}$,

$$(sup_i \phi_i)^{-1} p \le tr(M_n^{-\frac{1}{2}L} \Sigma_n^{-1} M_n^{-\frac{1}{2}R}) = \| \Sigma_n^{-\frac{1}{2}L} M_n^{-\frac{1}{2}R} \|^2 \le (inf_i \phi_i)^{-1} p. \tag{3.7}$$

Hence from (3.6) and (3.7),

$$\Sigma_n^{-\frac{1}{2}L}(\hat{\beta}_n - \beta_0) = \Sigma_n^{-\frac{1}{2}L} M_n^{-1} s_n + o_p(1).$$

From (D1) and (M3), $\max_{i \le n} x_i^\tau M_n^{-1} x_i \to 0$. Thus,

$$\Sigma_n^{-\frac{1}{2}L} M_n^{-1} s_n \to_d N(0, I)$$

from Lemma 2 stated below. This completes the proof. □

Lemma 2. Let $\{ \xi_i \}_{i=1}^\infty$ be a sequence of independent random variables with $E\xi_i = 0$ and $sup_i E|\xi_i|^{2+\varepsilon} < \infty$ for a constant $\varepsilon > 0$. Let $\{ c_i \}_{i=1}^\infty$ be a sequence of constants satisfying

$$\lim_{n\to\infty} \max_{i \le n} c_i^2 / \sum_{i=1}^n c_i^2 = 0.$$

Then

$$\sum_{i=1}^n c_i \xi_i / [Var(\sum_{i=1}^n c_i \xi_i)]^{\frac{1}{2}} \to_d N(0,1).$$

Proof. This can be shown by directly checking the Lindeberg's condition. □

3.2. Nonnatural link functions. For the case where g is a nonnatural link function, the following lemma is very useful. The proof of this lemma is in Wu (1981).

Lemma 3. Let $\mathbf{C}$ be a compact subset of $\mathbf{R}^p$, $\{ f_i \}_{i=1}^{\infty}$ be Lipschitz continuous on $\mathbf{C}$ (see Section 2), and $\{ \xi_i \}_{i=1}^{\infty}$ be independent random variables with $E\xi_i=0$ and $sup_i E\xi_i^2 < \infty$.
(i) If $d_n=\sum_{i=1}^n sup_{s\in \mathbf{C}} |f_i(s)|^2 \to\infty$ as $n\to\infty$, then for any $\alpha > 0$,

$$sup_{s\in \mathbf{C}} |\sum_{i=1}^n f_i(s)\xi_i|/d_n^{(1+\alpha)/2} \to 0 \ \ a.s.$$

(ii) If $d_n \to d < \infty$ and a_n is a sequence of positive constants satisfying $a_n \to\infty$, then

$$sup_{s\in \mathbf{C}} |\sum_{i=1}^n f_i(s)\xi_i|/a_n \to 0 \ \ a.s.$$

When g is a nonnatural link function,

$$H_n(\beta) = M_n(\beta) - R_n(\beta),$$

where

$$R_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \psi_i(\beta) e_i(\beta),$$

$\psi_i(\beta)=\psi(x_i^\tau\beta)$ and $\psi(t)=\eta''(t)$. In this case, the uniqueness of the solutions of (1.3) can not be guaranteed. However, examining the proofs in the previous subsection, it is evident that the assertions in Theorem 1 and Lemma 1 are still true for a sequence of solutions of (1.3) as long as

$$\| M_n^{-\frac{1}{2}L} H_n(\beta) M_n^{-\frac{1}{2}R} - I \| \to 0 \ \ a.s.$$

as $n\to\infty$ and $\|\beta-\beta_0\|\to 0$. This follows from (C) and

$$\| M_n^{-\frac{1}{2}L} R_n(\beta) M_n^{-\frac{1}{2}R} \| \to 0 \ \ a.s. \tag{3.8}$$

as $n\to\infty$ and $\|\beta-\beta_0\|\to 0$.

Lemma 4. Assume (2.4), (M3), (D2) and (D3). Suppose that $\{ \psi_i(\beta) \}_{i=1}^{\infty}$ is bounded and Lipschitz continuous on N(ε) and $\{ \mu_i(\beta) \}_{i=1}^{\infty}$ is equicontinuous at β_0. Then (3.8) holds.

Remark. If the range of x_i is compact, then $\{ \psi_i(\beta) \}_{i=1}^{\infty}$ is bounded on N(ε) and $\{ \mu_i(\beta) \}_{i=1}^{\infty}$ is equicontinuous at β_0. Conditions under which $\{ \psi_i(\beta) \}_{i=1}^{\infty}$ is Lipschitz continuous on N(ε) are discussed in Section 2.

Proof. Let $\tau_n = \lambda_{\min}(D_n)$,

$$W_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \psi_i(\beta) e_i$$

and

$$U_n(\beta) = \sum_{i=1}^n x_i x_i^\tau \psi_i(\beta)[\mu_i(\beta_0) - \mu_i(\beta)].$$

Then $R_n(\beta) = W_n(\beta) + U_n(\beta)$. From $\| M_n^{-1} \| \le c_1 \tau_n^{-1}$, where c_1 is a constant, we have

$$\| M_n^{-\frac{1}{2}L} W_n(\beta) M_n^{-\frac{1}{2}R} \| \le \| M_n^{-1} \| \, \| W_n(\beta) \| \le c_1 \tau_n^{-1} \| W_n(\beta) \| . \tag{3.9}$$

Let x_{ik} be the kth component of x_i and $g_{n,kj} = \sum_{i=1}^n x_{ik}^2 x_{ij}^2$ for any (k, j). Then from (D3),

$$\max_{k,j} g_{n,kj} \le \max_{i \le n} \| x_i \|^2 \sum_{i=1}^n \| x_i \|^2$$

$$\le \max_{i \le n} \| x_i \| [p \lambda_{\max}(D_n)] \le c_2 [\lambda_{\max}(D_n)]^{(1+\delta)/(1+\alpha)},$$

where c_2 is a constant. Then from (2.4), (D2) and Lemma 3, for any (k, j),

$$sup_{\beta \in N(\varepsilon)} | \sum_{i=1}^n x_{ik} x_{ij} \psi_i(\beta) e_i | / \tau_n \to 0 \;\; a.s.,$$

which together with (3.9) imply

$$sup_{\beta \in N(\varepsilon)} \| M_n^{-\frac{1}{2}L} W_n(\beta) M_n^{-\frac{1}{2}R} \| \to 0 \;\; a.s. \tag{3.10}$$

Also, there is a constant $c_0 > 0$ such that for all n,

$$\| M_n^{-\frac{1}{2}L} U_n(\beta) M_n^{-\frac{1}{2}R} \| \le c_0 \max_{i \le n} | \mu_i(\beta) - \mu_i(\beta_0) | \to 0 \tag{3.11}$$

as $\| \beta - \beta_0 \| \to 0$, since $\{ \mu_i(\beta) \}_{i=1}^\infty$ is equicontinuous at β_0. Then (3.8) follows from (3.10) and (3.11). □

From Lemma 4 and the above discussion, we have

Theorem 2. Assume that $\{ \psi_i(\beta) \}_{i=1}^\infty$ is bounded and Lipschitz continuous on $N(\varepsilon)$ for an $\varepsilon > 0$ and $\{ \mu_i(\beta) \}_{i=1}^\infty$ is equicontinuous at β_0. Then under Assumption (A2), the assertions in Lemma 1 are true and for any $\{ \hat{\beta}_n \}_{n=1}^\infty$ satisfying (3.2) and (3.3),

$$\Sigma_n^{-\frac{1}{2}L}(\hat{\beta}_n - \beta_0) \to_d N(0, I).$$

3.3. Strong consistency-robustness. The weak consistency-robustness of $\hat{\beta}_n$ has been established above. To prove the strong consistency-robustness of $\hat{\beta}_n$, we need to assume that the regressors have a compact range.

Theorem 3. Assuming (A3), the following two assertions hold.

(i) If g is the natural link function, then there exists a sequence of random variables $\hat{\beta}_n$ and a random number n_0 such that

$$P(\, s_n(\hat{\beta}_n)=0 \text{ for } all \;\, n \geq n_0\,) = 1 \tag{3.12}$$

and

$$\hat{\beta}_n \to \beta_0 \;\; a.s. \tag{3.13}$$

(ii) If g is a nonnatural link function with $\{\, \psi_i(\beta)\,\}_{i=1}^{\infty}$ being Lipschitz continuous on $\mathrm{N}(\varepsilon)$ for an $\varepsilon > 0$, then (3.12) and (3.13) hold.

Proof. (i) When **X**, the range of x_i, is compact, (D2) implies

$$\lambda_{\min}[M_n(\beta)] \geq c\,[\lambda_{\max}(M_n)]^{\frac{1}{2}(1+\delta)}, \qquad \beta \in \mathrm{N}(\varepsilon), \;\, n \geq n_1 \tag{3.14}$$

with some positive constants c, ε and n_1. Then the proof of (3.12) and (3.13) is the same as that of Theorem 2 of Fahrmeir and Kaufmann (1985).

(ii) We first show that almost surely,

$$H_n(\beta) \geq c_1 M_n(\beta), \qquad \beta \in \mathrm{N}(\varepsilon_0), \;\, n \geq n_2 \tag{3.15}$$

for some constants $\varepsilon_0 > 0$ and $c_1 > 0$ and a random number n_2. From the conditions in (ii) and the compactness of **X**, there is an $\varepsilon_1 > 0$ such that

$$\rho = sup_i \, \|x_i\| \; sup_i sup_{\beta \in \mathrm{N}(\varepsilon_1)} |v_i(\beta) h_i(\beta)| \; sup_i sup_{\beta \in \mathrm{N}(\varepsilon_1)} |\psi_i(\beta)/\zeta_i(\beta)| < \infty.$$

Choose $\varepsilon_0 < \min(\, \varepsilon_1, p^{-\frac{1}{2}} \rho^{-1}\,)$. Using the same notation as in the proof of Lemma 4, we have

$$sup_{\beta \in \mathrm{N}(\varepsilon_0)} \|M_n^{-\frac{1}{2}L}(\beta) U_n(\beta) M_n^{-\frac{1}{2}R}(\beta)\| \leq p^{\frac{1}{2}} sup_i sup_{\beta \in \mathrm{N}(\varepsilon_0)} |\psi_i(\beta)/\zeta_i(\beta)| \, |\mu_i(\beta) - \mu_i(\beta_0)|$$

$$\le \varepsilon_0 p^{1/2} sup_i \|x_i\| \; sup_i sup_{\beta\in N(\varepsilon_0)} |v_i(\beta)h_i(\beta)| \; sup_i sup_{\beta\in N(\varepsilon_0)} |\psi_i(\beta)/\zeta_i(\beta)| \le \varepsilon_0 p^{1/2} \rho < 1.$$

Hence for $\beta\in N(\varepsilon_0)$ and all n,

$$M_n^{-1/2L}(\beta)U_n(\beta)M_n^{-1/2R}(\beta) \le \varepsilon_0 p^{1/2} \rho \, I.$$

From the compactness of **X**, there is a constant $c_2>0$ such that

$$sup_{\beta\in N(\varepsilon_0)} \|M_n^{-1/2L}(\beta)M_n^{1/2L}\| \le c_2.$$

Then from (3.10), almost surely,

$$M_n^{-1/2L}(\beta)W_n(\beta)M_n^{-1/2R}(\beta) \le \varepsilon_2 I, \qquad \beta\in N(\varepsilon_0),\ n \ge n_2$$

for a constant $\varepsilon_2 < 1-\varepsilon_0 p^{1/2}\rho$ and a random number n_2. Let $c_1=1-(\varepsilon_0 p^{1/2}\rho+\varepsilon_2)$. Then almost surely,

$$M_n^{-1/2L}(\beta)R_n(\beta)M_n^{-1/2R}(\beta) \le (\varepsilon_0 p^{1/2}\rho+\varepsilon_2)\, I, \qquad \beta\in N(\varepsilon_0),\ n \ge n_2$$

and

$$M_n^{-1/2L}(\beta)H_n(\beta)M_n^{-1/2R}(\beta) \ge c_1 I, \qquad \beta\in N(\varepsilon_0),\ n \ge n_2.$$

Hence (3.15) holds. Then from (3.14), almost surely,

$$\lambda_{min}[H_n(\beta)] \ge c\,[\lambda_{max}(M_n)]^{1/2(1+\delta)} \qquad \beta\in N(\varepsilon_0),\ n \ge n_2,$$

where $c > 0$ is a constant. The rest of the proof is the same as that of Theorem 2 of Fahrmeir and Kaufmann (1985). □

4. Estimators of the asymptotic covariance matrix

We have shown in Section 3 that $\hat\beta_n$ has an asymptotic normal distribution with mean β_0 and covariance matrix Σ_n (3.5). For statistical inference based on $\hat\beta_n$, a consistency-robust estimator of Σ_n is required. More precisely, regardless of equality or lack of equality of ϕ_i's, we need an estimator $\hat\Sigma_n$ such that

$$\Sigma_n^{-1/2L}\hat\Sigma_n\Sigma_n^{-1/2R} - I \to_p 0 \;\; (\text{or} \to 0 \;\; a.s.). \tag{4.1}$$

Note that (3.4) is true for any choice of the square root of Σ_n. Let $\hat{\Sigma}_n^{1/2L}$ and $\Sigma_n^{1/2L}$ be the Cholesky square roots of $\hat{\Sigma}_n$ and Σ_n, respectively. Then as argued in Fahrmeir and Kaufmann (1985, Remark (iii) after Theorem 3), (4.1) implies $\Sigma_n^{-1/2L}\hat{\Sigma}_n^{1/2L} \to_p I$ and therefore

$$\hat{\Sigma}_n^{-1/2L}(\hat{\beta}_n - \beta_0) \to_d N(0, I).$$

When it is desired to make inference about $l^\tau\beta_0$, where l is a known p-vector, the following discussion shows that the computation of the square root of $\hat{\Sigma}_n$ can be avoided. Note that (3.4) is equivalent to

$$l^\tau(\hat{\beta}_n - \beta_0)/(l^\tau\Sigma_n l)^{1/2} \to_d N(0,1) \quad \text{for } any \ l \in \mathbf{R}^p,\ l \neq 0.$$

Then, for statistical inference, we need

$$(l^\tau\hat{\Sigma}_n l)/(l^\tau\Sigma_n l) \to_p 1.$$

But this is guaranteed by (4.1), since (4.1) implies

$$l_n^\tau(\Sigma_n^{-1/2L}\hat{\Sigma}_n\Sigma_n^{-1/2R} - I)l_n \to_p 0$$

for any vector l_n satisfying $\|l_n\|=1$ and

$$(l^\tau\hat{\Sigma}_n l)/(l^\tau\Sigma_n l) - 1 = [l^\tau\Sigma_n^{1/2L}(\Sigma_n^{-1/2L}\hat{\Sigma}_n\Sigma_n^{-1/2R} - I)\Sigma_n^{1/2R}l]/(l^\tau\Sigma_n l) = l_n^\tau(\Sigma_n^{-1/2L}\hat{\Sigma}_n\Sigma_n^{-1/2R} - I)l_n$$

with $l_n = \Sigma_n^{1/2R}l/(l^\tau\Sigma_n l)^{1/2}$.

Let $\Sigma_n(\beta) = M_n^{-1}(\beta)F_n(\beta)M_n^{-1}(\beta)$. Then

$$\Sigma_n = \Sigma_n(\beta_0) = M_n^{-1}(\beta_0)\sum_{i=1}^n x_i x_i^\tau h_i^2(\beta_0)\sigma_i^2 M_n^{-1}(\beta_0). \tag{4.2}$$

A natural estimator of Σ_n is obtained by replacing σ_i^2 in (4.2) by an estimator $\hat{\sigma}_i^2$ and β_0 by $\hat{\beta}_n$. If σ_i^2 (or ϕ_i) are neither known nor functions of β, consistent estimators of σ_i^2 are not available in general. Nevertheless, one may use estimators based on residuals, i.e., $\hat{\sigma}_i^2 = w_i r_i^2$, where w_i are positive constants. The resulting estimator of Σ_n is then

$$\hat{\Sigma}_n = M_n^{-1}(\hat{\beta}_n)\sum_{i=1}^n x_i x_i^\tau h_i^2(\hat{\beta}_n)w_i r_i^2 M_n^{-1}(\hat{\beta}_n). \tag{4.3}$$

It is proved in Section 5 that $\hat{\Sigma}_n$ defined in (4.3) is consistency-robust (i.e., (4.1) holds) for any w_i satisfying $\lim_{n\to\infty}\max_{i\leq n}|w_i - 1| = 0$. Let us illustrate that the same estimator $\hat{\Sigma}_n$ (4.3) may be obtained by using two alternative methods as follows.

(1) *The heteroscedastic bootstrap.* This method is an extension of the bootstrap method to the case where the observations have unequal variances (Beran (1986) and Efron (1986)). In view of (3.6), $\hat{\beta}_n - \beta_0$ is approximated by $M_n^{-1} s_n$, which has covariance matrix Σ_n under a probability structure with parameters β, $\{ \sigma_i^2 \}_{i=1}^{\infty}$. For given y, one may generate "data" $y_1^*, \ldots, y_n^*$ from the same probability structure but with the parameters being estimated by $\hat{\beta}_n$, $\{ \hat{\sigma}_i^2 \}_{i=1}^{\infty}$. Then the resample analog of $M_n^{-1} s_n$ is

$$M_n^{-1}(\hat{\beta}_n) \sum_{i=1}^n x_i h_i(\hat{\beta}_n)[y_i^* - \mu_i(\hat{\beta}_n)]$$

and its covariance matrix under the bootstrap distribution is

$$Var_*[M_n^{-1}(\hat{\beta}_n) \sum_{i=1}^n x_i h_i(\hat{\beta}_n) y_i^*] = M_n^{-1}(\hat{\beta}_n) \sum_{i=1}^n x_i x_i^\tau h_i^2(\hat{\beta}_n) Var_*(y_i^*) M_n^{-1}(\hat{\beta}_n),$$

which is identical to $\hat{\Sigma}_n$ (4.3) since $Var_*(y_i^*) = \hat{\sigma}_i^2$.

(2) *The linear jackknife.* Fox, Hinkley and Larntz (1980) introduced the linear jackknife for estimating the covariances of the least squares estimators in nonlinear regression models. The resulting estimators are consistent (Shao (1988)). The linear jackknife in GLIM is described as follows. Let $\hat{\beta}_{n,j}$ be the WLSE of β obtained by deleting the data (y_j, x_j), $j=1,\ldots,n$. From (3.6),

$$\hat{\beta}_n - \beta_0 \approx M_n^{-1} s_n = M_n^{-1} \sum_{i=1}^n x_i h_i(\beta_0) e_i$$

and

$$\hat{\beta}_{n,j} - \beta_0 \approx M_{n,j}^{-1} \sum_{i \neq j} x_i h_i(\beta_0) e_i,$$

where $M_{n,j} = \sum_{i \neq j} x_i x_i^\tau \zeta_i(\beta_0)$. Then

$$\hat{\beta}_n - \hat{\beta}_{n,j} \approx M_n^{-1} x_j h_j(\beta_0) e_j - M_n^{-1} x_j x_j^\tau \zeta_j(\beta_0) M_{n,j}^{-1} \sum_{i \neq j} x_i h_i(\beta_0) e_i, \tag{4.4}$$

since $M_n^{-1} - M_{n,j}^{-1} = M_n^{-1}(M_{n,j} - M_n) M_{n,j}^{-1} = -M_n^{-1} x_j x_j^\tau \zeta_j(\beta_0) M_{n,j}^{-1}$. Note that

$$E \| M_n^{-1} x_j h_j(\beta_0) e_j \|^2 = \phi_j \zeta_j^2(\beta_0) x_j^\tau M_n^{-2} x_j$$

and

$$E \| M_n^{-1} x_j x_j^\tau \zeta_j(\beta_0) M_{n,j}^{-1} \sum_{i \neq j} x_i h_i(\beta_0) e_i \|^2 \leq (sup_i \phi_i) \zeta_j^2(\beta_0) \; tr(M_n^{-1} x_j x_j^\tau M_{n,j}^{-1} x_j x_j^\tau M_n^{-1})$$

$$= (sup_i\phi_i)\zeta_j^2(\beta_0)(x_j^\tau M_{n,j}^{-1}x_j)(x_j^\tau M_n^{-2}x_j),$$

i.e., the second term on the right hand side of (4.4) is of a lower order than the first term. Thus,

$$\hat{\beta}_n-\hat{\beta}_{n,j} \approx M_n^{-1}x_jh_j(\beta_0)e_j.$$

Replacing β_0 by $\hat{\beta}_n$ and e_j by r_j, we have

$$\hat{\beta}_n-\hat{\beta}_{n,j} \approx M_n^{-1}(\hat{\beta}_n)x_jh_j(\hat{\beta}_n)r_j,$$

i.e., $M_n^{-1}(\hat{\beta}_n)x_jh_j(\hat{\beta}_n)r_j$ is the dominating component of $\hat{\beta}_n-\hat{\beta}_{n,j}$. Note that the weighted jackknife estimator of the asymptotic covariance matrix of $\hat{\beta}_n$ is (Wu (1986)):

$$\Sigma_{i=1}^n w_i(\hat{\beta}_n-\hat{\beta}_{n,j})(\hat{\beta}_n-\hat{\beta}_{n,j})^\tau.$$

Taking the dominating component of $\hat{\beta}_n-\hat{\beta}_{n,j}$, the linear jackknife estimator is

$$M_n^{-1}(\hat{\beta}_n)\Sigma_{i=1}^n x_ix_i^\tau h_i^2(\hat{\beta}_n)w_ir_i^2M_n^{-1}(\hat{\beta}_n),$$

which is of the same type as the estimators defined in (4.3).

Note that in the classical linear model where $h\equiv1$, $v\equiv1$, and $\phi_i\equiv\sigma_i^2$, the linear jackknife is exactly the same as the jackknife.

5. Consistency-robustness of $\hat{\Sigma}_n$

We prove the consistency-robustness of $\hat{\Sigma}_n$ (as defined in (4.3)) first for the special case of classical linear models, then for the GLIM with natural link functions and finally for the GLIM with nonnatural link functions. Throughout this and the next section, we will assume that the coefficients w_i in (4.3) satisfy

$$\lim_{n\to\infty} \max_{i\le n} |w_i-1| = 0. \tag{5.1}$$

5.1. The classical linear model. The results in this subsection are not special cases of the results in Subsections 5.2 and 5.3, since we do not need to assume that the distribution of y_i is of the form (1.1) in the classical linear model. Only

$$sup_i Ey_i^4 < \infty \tag{5.2}$$

is assumed, which is satisfied if y_i has a density given in (1.1). Our results are also extensions of those in Shao and Wu (1987) for "delete-one" version jackknife estimators, where $\lambda_{\min}(D_n) \geq cn$ for a constant $c > 0$ is assumed.

Recall that in the classical linear model, $\mu_i = x_i^\tau \beta$,

$$\Sigma_n = D_n^{-1} \sum_{i=1}^n x_i x_i^\tau \sigma_i^2 D_n^{-1}, \qquad \sigma_i^2 = Var(y_i)$$

and

$$\hat{\Sigma}_n = D_n^{-1} \sum_{i=1}^n x_i x_i w_i r_i^2 D_n^{-1}, \qquad r_i = y_i - x_i^\tau \hat{\beta}_n.$$

Theorem 4. (i) Assume (5.2) and (D1)-(D3). Then

$$\Sigma_n^{-\frac{1}{2}L} \hat{\Sigma}_n \Sigma_n^{-\frac{1}{2}R} - I \to_p 0.$$

(ii) If in addition, **X** is compact, then

$$\Sigma_n^{-\frac{1}{2}L} \hat{\Sigma}_n \Sigma_n^{-\frac{1}{2}R} - I \to 0 \;\; a.s.$$

Proof. In view of (5.1), without loss of generality, we assume $w_i \equiv 1$.

(i) It is known that $\hat{\beta}_n \to \beta_0$ *a.s.* under (D1). Note that F_n (2.2) reduces to

$$F_n = \sum_{i=1}^n x_i x_i^\tau \sigma_i^2.$$

Let

$$G_n = \sum_{i=1}^n x_i x_i^\tau e_i^2,$$

where $e_i = y_i - x_i^\tau \beta_0$. Then by Kolmogorov's strong law of large numbers (e.g., Wu (1981, Lemma 2)) and (5.2),

$$\sum_{i=1}^n x_{ik} x_{ij} (e_i^2 - \sigma_i^2) / (\sum_{i=1}^n x_{ik}^2 x_{ij}^2)^{\frac{1}{2}(1+\alpha)} \to 0 \;\; a.s.$$

for any (k, j), where $\alpha > 0$ is given in (D3). Then by (D2) and (D3),

$$\| G_n - F_n \| / \lambda_{\min}(D_n) \to 0 \;\; a.s. \tag{5.3}$$

Write

$$\Sigma_n^{-\frac{1}{2}L}\hat{\Sigma}_n\Sigma_n^{-\frac{1}{2}R}-I=\Sigma_n^{-\frac{1}{2}L}(A_n+B_n+C_n)\Sigma_n^{-\frac{1}{2}R}-I,$$

where $A_n=D_n^{-1}G_nD_n^{-1}$, $B_n=D_n^{-1}\sum_{i=1}^n x_ix_i^\tau u_i^2D_n^{-1}$ with $u_i=r_i-e_i$, and $C_n=2D_n^{-1}\sum_{i=1}^n x_ix_i^\tau e_iu_iD_n^{-1}$. Then by the Cauchy-Schwarz inequality, the result follows if

$$\Sigma_n^{-\frac{1}{2}L}A_n\Sigma_n^{-\frac{1}{2}R}-I\to_p 0 \tag{5.4}$$

and

$$\Sigma_n^{-\frac{1}{2}L}B_n\Sigma_n^{-\frac{1}{2}R}\to_p 0. \tag{5.5}$$

From

$$\Sigma_n^{-\frac{1}{2}L}A_n\Sigma_n^{-\frac{1}{2}R}-I=\Sigma_n^{-\frac{1}{2}L}(A_n-\Sigma_n)\Sigma_n^{-\frac{1}{2}R}$$

$$=\Sigma_n^{-\frac{1}{2}L}D_n^{-1}(G_n-F_n)D_n^{-1}\Sigma_n^{-\frac{1}{2}R}=\Sigma_n^{-\frac{1}{2}L}D_n^{-\frac{1}{2}}D_n^{-\frac{1}{2}}(G_n-F_n)D_n^{-\frac{1}{2}}D_n^{-\frac{1}{2}}\Sigma_n^{-\frac{1}{2}R},$$

(5.4) follows if

$$\|\Sigma_n^{-\frac{1}{2}L}D_n^{-\frac{1}{2}}\|=O(1) \tag{5.6}$$

and

$$\|D_n^{-\frac{1}{2}}(G_n-F_n)D_n^{-\frac{1}{2}}\|\to_p 0.$$

Since $F_n^{-1}\le cD_n^{-1}$ for a constant $c>0$,

$$\|\Sigma_n^{-\frac{1}{2}L}D_n^{-\frac{1}{2}}\|^2=tr(D_n^{-\frac{1}{2}}\Sigma_n^{-1}D_n^{-\frac{1}{2}})=tr(D_n^{-\frac{1}{2}}D_nF_n^{-1}D_nD_n^{-\frac{1}{2}})\le cp.$$

Hence (5.6) holds. Also,

$$\|D_n^{-\frac{1}{2}}(G_n-F_n)D_n^{-\frac{1}{2}}\|\le\|D_n^{-\frac{1}{2}}\|^2\|G_n-F_n\|=tr(D_n^{-1})\|G_n-F_n\|$$

$$\le\|G_n-F_n\|/\lambda_{\min}(D_n)\to 0\ \ a.s.$$

by (5.3). This proves

$$\Sigma_n^{-\frac{1}{2}L}A_n\Sigma_n^{-\frac{1}{2}R}-I\to 0\ \ a.s., \tag{5.7}$$

which implies (5.4). Note that

$$u_i^2 = [x_i^\tau(\hat\beta_n - \beta_0)]^2 = [x_i^\tau \Sigma_n^{\frac{1}{2}L} \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0)]^2$$

$$\le (x_i^\tau \Sigma_n x_i) \| \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0) \|^2 \le c\,(x_i^\tau D_n^{-1} x_i) \| \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0) \|^2$$

for a constant $c > 0$. From $E \| \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0) \|^2 = p$, $\| \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0) \|^2 = O_p(1)$. Then by (D1),

$$\max_{i \le n} u_i^2 \le c \, \max_{i \le n} x_i^\tau D_n^{-1} x_i \| \Sigma_n^{-\frac{1}{2}L}(\hat\beta_n - \beta_0) \|^2 \to_p 0.$$

Thus by (5.6),

$$\| \Sigma_n^{-\frac{1}{2}L} B_n \Sigma_n^{-\frac{1}{2}R} \| \le \max_{i \le n} u_i^2 \| \Sigma_n^{-\frac{1}{2}L} D_n^{-1} \Sigma_n^{-\frac{1}{2}R} \| \le \max_{i \le n} u_i^2 \| \Sigma_n^{-\frac{1}{2}L} D_n^{-\frac{1}{2}} \|^2 \to_p 0.$$

This proves (5.5) and therefore completes the proof of (i).

(ii) From the proof of (i) and (5.6)-(5.7), we only need to show that

$$\max_{i \le n} u_i^2 \to 0 \;\; a.s.$$

Since **X** is compact, $sup_i \| x_i \|^2 < \infty$. Then

$$\max_{i \le n} u_i^2 \le sup_i \| x_i \|^2 \| \hat\beta_n - \beta_0 \|^2 \to 0 \;\; a.s.$$

by the strong consistency of $\hat\beta_n$. □

5.2. The GLIM: natural link functions. We return to the GLIM with natural link functions, i.e., $h \equiv 1$ and therefore

$$\Sigma_n = M_n^{-1} \sum_{i=1}^n x_i x_i^\tau \sigma_i^2 M_n^{-1}$$

and

$$\hat\Sigma_n = M_n^{-1}(\hat\beta_n) \sum_{i=1}^n x_i x_i^\tau w_i r_i^2 M_n^{-1}(\hat\beta_n).$$

Theorem 5. (i) Suppose that Assumption (A2) holds. Assume either

$$\{ \mu_i(\beta) \}_{i=1}^\infty \text{ is equicontinuous at } \beta_0 \tag{5.8}$$

or

$$\{ v_i(\beta) \}_{i=1}^\infty \text{ is bounded on } \mathbf{N}(\varepsilon) \text{ for an } \varepsilon > 0. \tag{5.9}$$

Then

$$\Sigma_n^{-\frac{1}{2}L}\hat{\Sigma}_n\Sigma_n^{-\frac{1}{2}R} - I \to_p 0.$$

(ii) Suppose that Assumption (A3) holds. Then

$$\Sigma_n^{-\frac{1}{2}L}\hat{\Sigma}_n\Sigma_n^{-\frac{1}{2}R} - I \to 0 \ \ a.s.$$

proof. Again we assume that $w_i \equiv 1$. Let $u_i = r_i - e_i = \mu_i(\beta_0) - \mu_i(\hat{\beta}_n)$. Examining the proof of Theorem 4, the result in (i) (or in (ii)) follows from

$$\| \Sigma_n^{-\frac{1}{2}L} M_n^{-\frac{1}{2}}(\hat{\beta}_n) \| = O_p(1) \ \ (\text{or } O(1) \ \ a.s.), \tag{5.10}$$

$$tr[M_n^{-1}(\hat{\beta}_n)] = O_p(\tau_n^{-1}) \ \ (\text{or } O(\tau_n^{-1}) \ \ a.s.), \tag{5.11}$$

where $\tau_n = \lambda_{\min}(D_n)$, and

$$\max_{i \le n} u_i^2 = o_p(1) \ \ (\text{or } o(1) \ \ a.s.). \tag{5.12}$$

From Lemma 1 and Theorem 3, the conditions in (i) (or in (ii)) imply $\hat{\beta}_n \to_p \beta_0$ (or $\hat{\beta}_n \to \beta_0$ $a.s.$). Note that

$$\| \Sigma_n^{-\frac{1}{2}L} M_n^{-\frac{1}{2}}(\hat{\beta}_n) \|^2 = tr[M_n^{-\frac{1}{2}}(\hat{\beta}_n)\Sigma_n^{-1} M_n^{-\frac{1}{2}}(\hat{\beta}_n)]$$

$$\le c \ tr[M_n^{-\frac{1}{2}}(\hat{\beta}_n) M_n M_n^{-\frac{1}{2}}(\hat{\beta}_n)] = c \ tr[M_n^{\frac{1}{2}} M_n^{-1}(\hat{\beta}_n) M_n^{\frac{1}{2}}]$$

for a constant $c > 0$. Then (5.10) follows from (C) and the weak (or strong) consistency of $\hat{\beta}_n$. Similarly, (5.11) holds since

$$tr[M_n^{-1}(\hat{\beta}_n)] \le \tau_n^{-1} tr[M_n^{-\frac{1}{2}}(\hat{\beta}_n) M_n M_n^{-\frac{1}{2}}(\hat{\beta}_n)].$$

If (5.8) holds, (5.12) follows from the weak (or strong) consistency of $\hat{\beta}_n$. Note that the compactness of **X** implies (5.8). Hence it remains to show that

$$\max_{i \le n} u_i^2 \to_p 0$$

under (5.9). By the mean-value theorem,

$$u_i^2 = v_i^2(\beta^*)[x_i^\tau(\hat{\beta}_n - \beta_0)]^2,$$

where β^* is on the line segment between β_0 and $\hat{\beta}_n$. From Theorem 1, $\Sigma_n^{-\frac{1}{2}L}(\hat{\beta}_n - \beta_0) = O_p(1)$. Hence by (D1),

$$\max_{i\le n}[x_i^\tau(\hat\beta_n-\beta_0)]^2 \le c\ \max_{i\le n} x_i^\tau D_n^{-1}x_i\,\|\Sigma_n^{-\frac12 L}(\hat\beta_n-\beta_0)\|^2 \to_p 0,$$

where $c>0$ is a constant. From (5.9) and the weak consistency of $\hat\beta_n$, $\max_{i\le n} v_i^2(\beta^*)=O_p(1)$. This completes the proof. □

5.3. The GLIM: nonnatural link functions. As we discussed early, the solutions of (1.3) may not be unique for nonnatural link functions. We say that $\{\hat\beta_n\}_{n=1}^\infty$ is a sequence of weakly (or strongly) consistent solutions of (1.3) iff (3.2) and (3.3) (or (3.12) and (3.13)) hold for $\hat\beta_n$. The existence of a weakly (or strongly) consistent sequence of solutions of (1.3) is proved in Section 3.

Theorem 6. (i) Suppose that (5.8) and Assumption (A2) hold. Also, $\{h_i^2(\beta)\}_{i=1}^\infty$ and $\{\psi_i(\beta)\}_{i=1}^\infty$ are bounded and Lipschitz continuous on N(ε) for an $\varepsilon>0$. Then for any weakly consistent sequence of solutions $\hat\beta_n$,

$$\Sigma_n^{-\frac12 L}\hat\Sigma_n\Sigma_n^{-\frac12 R}-I\to_p 0.$$

(ii) Suppose that Assumption (A3) holds and $\{\psi_i(\beta)\}_{i=1}^\infty$ is bounded and Lipschitz continuous on N(ε). Then for any strongly consistent sequence of solutions $\hat\beta_n$,

$$\Sigma_n^{-\frac12 L}\hat\Sigma_n\Sigma_n^{-\frac12 R}-I\to 0\ \ a.s.$$

Proof. Assume $w_i\equiv 1$. Let

$$A_n=M_n^{-1}(\hat\beta_n)\sum_{i=1}^n x_ix_i^\tau h_i^2(\hat\beta_n)e_i^2M_n^{-1}(\hat\beta_n)$$

and

$$B_n=M_n^{-1}(\hat\beta_n)\sum_{i=1}^n x_ix_i^\tau h_i^2(\hat\beta_n)u_i^2M_n^{-1}(\hat\beta_n).$$

From Theorems 2 and 3, Condition (C) and the proof of Theorem 5,

$$\|\Sigma_n^{-\frac12 L}B_n\Sigma_n^{-\frac12 R}\|\to_p 0\ \ (\text{or}\to 0\ \ a.s.)$$

under the conditions in (i) (or in (ii)). Then it suffices to show

$$\Sigma_n^{-\frac12 L}A_n\Sigma_n^{-\frac12 R}-I\to_p 0\ \ (\text{or}\to 0\ \ a.s.)$$

under the conditions in (i) (or in (ii)). Assume the conditions in (i). By Lemma 3, (D2) and (D3),

$$sup_{\beta\in N(\varepsilon)} |\sum_{i=1}^n x_i x_i^\tau h_i^2(\beta)(e_i^2-\sigma_i^2)|/\lambda_{\min}(D_n) \to 0 \ \ a.s.$$

which implies, via (C) and the weak (or strong) consistency of $\hat\beta_n$,

$$\Delta_n = \Sigma_n^{-\frac{1}{2}L} M_n^{-1}(\hat\beta_n)\sum_{i=1}^n x_i x_i^\tau h_i^2(\hat\beta_n)(e_i^2-\sigma_i^2)M_n^{-1}(\hat\beta_n)\Sigma_n^{-\frac{1}{2}R} \to_p 0 \ \ (\text{or} \to 0 \ \ a.s.).$$

Since

$$\Sigma_n^{-\frac{1}{2}L} A_n \Sigma_n^{-\frac{1}{2}R} - I = \Delta_n + \Sigma_n^{-\frac{1}{2}L} M_n^{-1}(\hat\beta_n)\sum_{i=1}^n x_i x_i^\tau [h_i^2(\hat\beta_n)-h_i^2(\beta_0)]\sigma_i^2 M_n^{-1}(\hat\beta_n)\Sigma_n^{-\frac{1}{2}R},$$

what remains to be shown is that

$$M_n^{-\frac{1}{2}}(\hat\beta_n)\sum_{i=1}^n x_i x_i^\tau [h_i^2(\hat\beta_n)-h_i^2(\beta_0)]\sigma_i^2 M_n^{-\frac{1}{2}}(\hat\beta_n) \to_p 0 \ \ (\text{or} \to 0 \ \ a.s.). \qquad (5.13)$$

If $\{ h_i^2(\beta) \}_{i=1}^\infty$ is bounded and Lipschitz continuous on $\mathbf{N}(\varepsilon)$, then there is a constant $c > 0$ such that

$$\max_{i\le n} |h_i^2(\hat\beta_n)-h_i^2(\beta_0)| \le c \,\|\hat\beta_n-\beta_0\| \qquad (5.14)$$

for $\hat\beta_n \in \mathbf{N}(\varepsilon)$. From the remarks in the end of Section 2, the compactness of $\mathbf{X}$ implies that $\{ h_i^2(\beta) \}_{i=1}^\infty$ is bounded and Lipschitz continuous on $\mathbf{N}(\varepsilon)$. Hence the proof is completed by noting that (5.13) follows from (C), (5.14) and the weak (or strong) consistency of $\hat\beta_n$. □

6. The special case of $\phi_i \equiv \phi$

From $\sigma_i^2=\phi_i v_i(\beta_0)$, estimates of ϕ_i are $\hat\phi_i=\hat\sigma_i^2/v_i(\hat\beta_n)$. When $\phi_i\equiv\phi$, Σ_n reduces to ϕM_n^{-1}. Thus, consistent estimator of Σ_n can be obtained by estimating ϕ consistently. A natural estimator of ϕ is the average of $\hat\phi_i$:

$$\hat\phi = n^{-1}\sum_{i=1}^n z_i(\hat\beta_n)\hat\sigma_i^2 = n^{-1}\sum_{i=1}^n z_i(\hat\beta_n) w_i r_i^2,$$

where $z_i(\beta)=1/v_i(\beta)$ and $\hat\beta_n$ is any sequence of weakly (or strongly) consistent solutions of (1.3). For the classical linear model where $z_i\equiv 1$ and $\phi=Var(y_i)$, $\hat\phi$ reduces to the sum of residual squares if $w_i\equiv(1-p/n)^{-1}$. The consistency of $\hat\phi$ is proved in the following theorem. Thus, a consistent estimator of Σ_n is $\hat\phi\, M_n^{-1}(\hat\beta_n)$. It should be remarked that $\hat\phi\, M_n^{-1}(\hat\beta_n)$ is not consistent if some of ϕ_i are not equal, i.e., $\hat{\phi} M_n^{-1}(\hat{\beta}_n)$ is not consistency-robust.

Theorem 7. (i) Suppose that (5.8) and Assumption (A2) hold. Also, $\{ z_i(\beta) \}_{i=1}^{\infty}$ and $\{ \psi_i(\beta) \}_{i=1}^{\infty}$ are bounded and Lipschitz continuous on $\mathbf{N}(\varepsilon)$ for an $\varepsilon > 0$. Then

$$\hat{\phi} \to_p \phi.$$

(ii) Suppose that Assumption (A3) holds and $\{ \psi_i(\beta) \}_{i=1}^{\infty}$ is bounded and Lipschtiz continuous on $\mathbf{N}(\varepsilon)$. Then

$$\hat{\phi} \to \phi \;\; a.s.$$

Proof. Assume $w_i \equiv 1$. Write

$$\hat{\phi} = n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)e_i^2 + n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)u_i^2 + 2n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)e_i u_i.$$

From the proof of Theorem 5,

$$\max_{i\le n} u_i^2 \to_p 0 \;\; (\text{or} \to 0 \;\; a.s.)$$

under the conditions in (i) (or in (ii)). Since $\{ z_i(\beta) \}_{i=1}^{\infty}$ is bounded on $\mathbf{N}(\varepsilon)$, $z_i(\hat{\beta}_n) \le c$ for a constant $c > 0$ and $\hat{\beta}_n \in \mathbf{N}(\varepsilon)$. Thus,

$$n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)u_i^2 \to_p 0 \;\; (\text{or} \to 0 \;\; a.s.)$$

under the conditions in (i) (or in (ii)). From Lemma 3,

$$sup_{\beta\in \mathbf{N}(\varepsilon)} n^{-1} \left| \sum_{i=1}^{n} z_i(\beta)(e_i^2 - \sigma_i^2) \right| \to 0 \;\; a.s.,$$

which implies

$$n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)(e_i^2 - \sigma_i^2) \to_p 0 \;\; (\text{or} \to 0 \;\; a.s.)$$

under the conditions in (i) (or in (ii)). Hence the proof is completed if we can show

$$n^{-1}\sum_{i=1}^{n} z_i(\hat{\beta}_n)\sigma_i^2 - \phi \to_p 0 \;\; (\text{or} \to 0 \;\; a.s.). \tag{6.1}$$

Since $\{ z_i(\beta) \}_{i=1}^{\infty}$ is bounded and Lipschitz continuous on $\mathbf{N}(\varepsilon)$,

$$\max_{i\le n} |z_i(\hat{\beta}_n) - z_i(\beta_0)| \le c \, \|\hat{\beta}_n - \beta_0\|$$

for $\hat{\beta}_n \in N(\varepsilon)$, where c is a constant. Hence (6.1) follows from $\phi = n^{-1}\sum_{i=1}^{n} z_i(\beta_0)\sigma_i^2$ and the weak (or strong) consistency of $\hat{\beta}_n$. □

REFERENCES

Beran, R. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14**, 1295-1298.

Bradley, E. L. (1973). The equivalence of maximum likelihood and weighted least squares estimates in the exponential family. *J. Amer. Statist. Assco.* **68**, 199-200.

Efron, B. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14**, 1301-1304.

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

Fox, T., Hinkley, D. and Larntz, K. (1980). Jackknifing in nonlinear regression. *Technometrics* **22**, 29-33.

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

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

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.

Shao, J. (1988). Asymptotic theory in heteroscedastic nonlinear models. Tech. Report #88-2, Department of Statistics, Purdue University.

Shao, J. and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15**, 1563-1579.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrka* **63**, 27-32.

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.

Department of Statistics
Purdue University
West Lafayette, IN 47907