The Annals of Statistics
1985, Vol. 13, No. 1, 342–368

CONSISTENCY AND ASYMPTOTIC NORMALITY OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN GENERALIZED LINEAR MODELS

BY LUDWIG FAHRMEIR AND HEINZ KAUFMANN

University of Regensburg

Generalized linear models are used for regression analysis in a number of cases, including categorical responses, where the classical assumptions are violated. The statistical analysis of such models is based on the asymptotic properties of the maximum likelihood estimator. We present mild general conditions which, respectively, assure weak or strong consistency or asymptotic normality. Most of the previous work has been concerned with natural link functions. In this case our normality condition, though obtained by a different approach, is closely related to a condition of Haberman (1977a). Examples show how the general conditions reduce to weak requirements for special exponential families. Further, for regressors with a compact range, sufficient conditions are given which do not involve the unknown parameter, and are therefore easy to check in practice. Responses with a bounded range, e.g. categorical responses, and stochastic regressors also are treated.

1. Introduction. Generalized linear models (Nelder and Wedderburn, 1972) are regression models for a number of cases where the classical assumptions are not met. Let the data consist of a sequence $\{(\mathbf{y}_n, \mathbf{Z}_n)\}$, where $\mathbf{y}_n$, the responses, are independent q-dimensional random variables, and $\mathbf{Z}_n$, the regressors, are $p \times q$-matrices of known constants. The distribution of the response $\mathbf{y}_n$ is assumed to belong to a natural exponential family (univariate examples are the normal, binomial, Poisson, exponential and gamma distribution, multivariate examples the multinormal and the multinomial distribution). The mean of the response $\mathbf{y}_n$ is related to a linear combination $\mathbf{Z}_n'\beta$ of the regressors by a one-to-one mapping, the link function. Depending on the special exponential family, there is one "natural" link function. Multinomial (categorical, quantal) response models are an important example. Here with the natural link function the logit model is obtained.

Usually the p-vector β of coefficients of the linear combinations $\{\mathbf{Z}_n'\beta\}$ has to be estimated from a finite sample of n observations, e.g. by the maximum likelihood (ML) method. The methods for the analysis of a generalized linear model (e.g. McCullagh and Nelder, 1983; Fahrmeir and Kredler, 1984) heavily rely on the asymptotic properties of the maximum likelihood estimator (MLE) as $n \to \infty$.

Conditions to assure weak consistency and asymptotic normality of the MLE for natural link functions have previously been given by Haberman (1977a, admitting $p \to \infty$, too), Andersen (1980, without proof), Nordberg (1980) and

Received November 1983; revised July 1984.
AMS 1980 *subject classifications.* Primary 62F12, secondary 62H12.
Key words and phrases. Generalized linear models, categorical response models, maximum likelihood estimator, consistency, asymptotic normality.

especially for the logit model by Haberman (1974), McFadden (1974), Gourieroux and Monfort (1981, weak and strong consistency). For nonnatural link functions, much less work has been done. McCullagh (1983) states asymptotic results, without giving rigorous proofs or exact assumptions, and Mathieu (1981) requires rather strong conditions.

All authors assume the Fisher information of the first n observations, $\mathbf{F}_n(\beta)$ say, to be divergent:

$$\lambda_{\min}\mathbf{F}_n(\beta) \to \infty, \tag{1.1}$$

(here and in the sequel $\lambda_{\min}$ ($\lambda_{\max}$) denotes the smallest (largest) eigenvalue of a symmetric matrix). In the classical linear regression model with i.i.d. errors, a necessary and sufficient condition for weak (Drygas, 1976) and strong (Lai, Robbins and Wei, 1979) consistency of the least squares estimator is

$$\lambda_{\min} \sum_{i=1}^{n} \mathbf{Z}_i\mathbf{Z}_i' \to \infty \tag{1.2}$$

which is closely related to (1.1). For nonnormally distributed errors, only a weak negligibility condition has to be added to obtain asymptotic normality (Eicker, 1963). Therefore, it is natural to look for mild additional conditions which, together with (1.1), assure consistency and asymptotic normality of the MLE in generalized linear models.

Natural link functions are treated in Section 3. Weak consistency is shown under (1.1) together with an additional condition (C), requiring only that, in a neighborhood of the true parameter, information does not decrease too rapidly. Strong consistency is shown under (1.1) and a condition (S_δ) which bounds the eigenvalue ratio of the successive information matrices. Asymptotic normality is established under (1.1) and a continuity condition (N) on the sequence of information matrices. It should be noted that we show the (multivariate) asymptotic normality of the MLE normalised by a square root of the information matrix, i.e. $\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$, whereas some authors (Haberman, 1977a; Friedman, 1982) consider (univariate) asymptotic normality of any normed linear functional of the MLE. Although both formulations of asymptotic normality turn out to be equivalent (see Remark (ii) of Subsection 3.1), we found matrix normalisation to be more convenient in establishing asymptotic efficiency or the asymptotic χ^2-distribution of various test statistics.

Using fixed point theorems, Haberman (1977a, Condition 2) adds a mild requirement to (1.1) to obtain weak consistency and asymptotic normality. In Subsection 3.1 we present Haberman's normality condition in a form that allows a comparison with our normality condition. Though we use a different approach (see Sweeting, 1980; Hall and Heyde, 1980, page 162 ff.; Basawa and Scott, 1983, for related work) leading to simpler proofs, we arrive at conditions which have a strong relationship to Haberman's Condition 2. It can be seen that, under (1.1), his Condition 2 implies our normality condition (N). Although our assumption is simpler to interpret and to check, it seems only slightly weaker (see Remark (ii) in Subsection 3.1). Haberman (1977a) presents no separate condition for weak consistency comparable to our condition (C), which is weaker than (N), and strong consistency is not considered.

The conditions assumed by the other authors are all relatively strong and imply ours. They require a compact admissible set for the regressors, furthermore the smallest and the largest eigenvalue of the information matrix are demanded to grow at the same rate (Gourieroux and Monfort, 1981; implicitly in Nordberg, 1980), or, even stronger, it is assumed that $\mathbf{F}_n(\beta)/n \to \mathbf{V}(\beta)$ positive definite (McFadden, 1974; Andersen, 1980).

In Subsection 3.3 verification of the conditions is discussed. Examples (i), (ii), (iii) deal with specific exponential families and point out how the general conditions can be utilized in special cases. For the Poisson model, they reduce to (1.1) and a negligibility condition of the Feller-type, for gamma distributed responses consistency and asymptotic normality is achieved under minimal conditions. Example (iii) shows that the consistency condition (C) is weaker than the normality condition (N).

Thereafter, some corollaries present results of general interest. Corollary 1 states that the assumption of a compact admissible set for the regressors and Condition (1.2) alone assure weak consistency and asymptotic normality of the MLE. These conditions do not involve the parameter β and therefore may be checked easily in applications. Corollary 2 treats the case of responses with a bounded range, e.g. the logit model. As for the Poisson model, (C) and (N) are implied by (1.1) and a negligibility condition of the Feller-type. In Corollary 3, stochastic regressors are considered, where the matrices $\mathbf{Z}_n$, $n = 1, 2, \cdots$, are observations of random matrices. This situation is frequently met in practice. Specifically, we shall assume that the pairs $(\mathbf{y}_n, \mathbf{Z}_n)$, $n = 1, 2, \cdots$, are independent and identically distributed. Lack of further knowledge about the common distribution of $(\mathbf{y}_n, \mathbf{Z}_n)$ leads to the conditional approach: the parameter β is estimated by maximizing the conditional likelihood, given $\mathbf{Z}_1, \cdots, \mathbf{Z}_n$. Weak assumptions on the marginal distribution of $\mathbf{Z}_n$, in particular not requiring a compact range for the regressors, are demonstrated to be sufficient for consistency and asymptotic normality of the (conditional) MLE. Although concerned with a more general case and using a different approach, a paper of Haberman (1984) is also of relevance in connection with conditional ML estimation.

Nonnatural link functions are treated in Section 4. We first discuss how the general consistency and normality conditions of Section 3 can be extended to cover this more general situation. Thereupon, we study the same three cases of general interest as for natural link functions. Again, for regressors with a compact range, Condition (1.2) alone assures weak consistency and asymptotic normality, if the link function is twice continuously differentiable. For responses with a bounded range and stochastic regressors, further mild assumptions are added to the requirements for natural link functions. All these conditions are far weaker and easier to check than those of Mathieu (1981), who requires a convergence condition, analogous to $\mathbf{F}_n(\beta)/n \to \mathbf{V}(\beta)$ as above, a Liapounov condition for the score function, and, tacitly, a compact admissible set for the regressors.

Finally, in an example, we treat the binomial model. It is demonstrated that weak consistency and asymptotic normality hold under the same negligibility condition of Section 3 as for the logit model, if additional weak assumptions on

the derivatives of the link function are met. Such assumptions hold e.g. for the probit model.

2. Generalized linear models.

2.1 Definition and some properties. A family of distributions $P_{\mathbf{y}|\theta}$ of a q-dimensional random variable $\mathbf{y}$, $\theta \in \Theta \subset \mathbb{R}^q$, which have densities

$$f(\mathbf{y} \mid \theta) = c(\mathbf{y})\exp(\theta'\mathbf{y} - b(\theta)), \quad c \geq 0 \text{ measurable},$$

with respect to a σ-finite measure ν is called a *natural exponential family* with *natural parameter* θ. We assume Θ to be the *natural parameter space*, i.e. the set of all θ satisfying $0 < \int c(\mathbf{y})\exp(\theta'\mathbf{y})\, d\nu < \infty$. Then Θ is convex, and in the interior Θ^0 of Θ, all derivatives of $b(\theta)$ and all moments of $\mathbf{y}$ exist (we assume $\Theta^0 \neq \emptyset$). In particular we have $E_\theta \mathbf{y} = \partial b(\theta)/\partial\theta = \mu(\theta)$ and $\operatorname{cov}_\theta \mathbf{y} = \partial^2 b(\theta)/\partial\theta\, \partial\theta' = \Sigma(\theta)$, say. The covariance matrix $\Sigma(\theta)$ is supposed to be positive definite in the interior Θ^0 of Θ, implying that the restriction of μ to Θ^0 is injective. Let M denote the image $\mu(\Theta^0)$ of Θ^0.

Generalized linear models are characterized by the following structure:

(i) The $\{\mathbf{y}_n\}$ are independent with densities

(2.1) $$f(\mathbf{y}_n \mid \theta_n) = c(\mathbf{y}_n)\exp(\theta_n'\mathbf{y}_n - b(\theta_n)), \quad n = 1, 2, \cdots,$$

of the natural exponential type, $\theta_n \in \Theta^0$.

(ii) The matrix $\mathbf{Z}_n$ influences $\mathbf{y}_n$ in form of a linear combination $\gamma_n = \mathbf{Z}_n'\beta$, where β is a p-dimensional parameter.

(iii) The linear combination is related to the mean $\mu(\theta_n)$ of $\mathbf{y}_n$ by the injective link function $\mathbf{g}$: $M \to \mathbb{R}^q$, $\gamma_n = \mathbf{g}(\mu(\theta_n))$.

REMARKS. (i) As in the original definition of Nelder and Wedderburn (1972), an additional nuisance parameter may be introduced in (2.1). The MLE of β remains the same, but the information matrix has to be multiplied by an unknown scale factor, which can be estimated consistently. Thus, without loss of generality, we confine ourselves to the simpler form (2.1).

(ii) In this paper the link function $\mathbf{g}$ is defined as in the program GLIM or in McCullagh and Nelder (1983). For theoretical purposes it is more convenient to relate $\gamma_n = \mathbf{Z}_n'\beta$ to the natural parameter θ_n by the injective function $\mathbf{u} = (\mathbf{g} \circ \mu)^{-1}$, i.e. $\theta_n = \mathbf{u}(\mathbf{Z}_n'\beta)$, as in the original definition.

(iii) Of special importance are natural link functions $\mathbf{g} = \mu^{-1}$, $\mathbf{u} = \mathrm{id}$. Then we obtain a linear model $\theta_n = \mathbf{Z}_n'\beta$ for the natural parameter.

(iv) In the above definition the elements of the matrices $\{\mathbf{Z}_n\}$ are known constants. For stochastic regressors we assume the pairs $\{(\mathbf{y}_n, \mathbf{Z}_n)\}$ to be i.i.d.

Then the definition is to be understood conditionally, i.e. (2.1) is the conditional density of $\mathbf{y}_n$ given $\mathbf{Z}_n$, and the $\{\mathbf{y}_n\}$ are conditionally independent.

EXAMPLES. We mention a few distributions and models.

(i) Univariate ($q = 1$) models. For the normal distribution and the natural link function $g(\mu) = \mu$ we obtain the classical linear regression model. Nonnatural link functions are e.g. μ^α, $\ln \mu$.

In the binomial case $P(y_n = 1) = \pi_n$, $P(y_n = 0) = 1 - \pi_n$ the natural link function is $g(\pi_n) = \text{logit}(\pi_n) = \ln \pi_n - \ln(1 - \pi_n)$, inducing the binary logit model $\text{logit}(\pi_n) = \mathbf{z}_n'\beta$ or, equivalently, $\pi_n = \exp(\mathbf{z}_n'\beta)/(1 + \exp(\mathbf{z}_n'\beta))$. Choice of $g(\pi_n) = \phi^{-1}(\pi_n)$, where ϕ is the standard normal distribution function, leads to the binary probit model $\pi_n = \phi(\mathbf{z}_n'\beta)$. The linear binary model $\pi_n = \mathbf{z}_n'\beta$ is obtained from setting $g(\pi_n) = \pi_n$. All these binary models may be used for regression analysis with binary dependent variables.

Poisson distributed responses y_n with mean λ_n are used, for example, in the analysis of multidimensional contingency tables. The natural link function $g(\lambda_n) = \ln \lambda_n$ provides the log linear Poisson model $\ln \lambda_n = \mathbf{z}_n'\beta$, the function $g(\lambda_n) = \lambda_n$ leads to the linear Poisson model $\lambda_n = \mathbf{z}_n'\beta$.

Regression models with gamma distributed responses y_n are used in the analysis of lifetimes depending on exogenous variables. The natural link function is $g(\mu) = -r/\mu$, where r is the shape parameter.

(ii) The most interesting multivariate ($q > 1$) examples are multinomial models, including the logit model.

2.2 *Regularity assumptions and the log likelihood function.* In the sequel β_0 denotes the true but unknown parameter which is to be estimated by the method of maximum likelihood, and β is any parameter in an admissible set $B \subset \mathbb{R}^p$. From now on we shall presume the following

REGULARITY ASSUMPTIONS. (i) B is open in $\mathbb{R}^p$ and, additionally, convex for natural link functions,

(ii) $\mathbf{Z}_n'\beta \in \mathbf{g}(M)$, $n = 1, 2, \cdots$, for all $\beta \in B$,

(iii) $\mathbf{g}$ resp. $\mathbf{u}$ is twice continuously differentiable, $\det(\partial \mathbf{u}/\partial \gamma) \neq 0$,

(iv) $\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i'$ has full rank for $n \geq n_0$, say.

Condition (ii) is necessary to have a generalized linear model for all β. For natural link functions, the convexity of B guarantees uniqueness of the maximum likelihood estimator if it exists; see the discussion after (2.3). The differentiability assumption on $\mathbf{g}$ is needed to guarantee that the second derivatives of the log likelihood are continuous. Condition (iv) and $\det(\partial \mathbf{u}/\partial \gamma) \neq 0$ will ensure that the information matrix $\mathbf{F}_n(\beta)$ is positive definite for all $\beta \in B$, $n \geq n_0$.

If there is an admissible set $\mathcal{Z}$ for the regressors, i.e. $\mathbf{Z}_n \in \mathcal{Z}$ for all n, then (ii) follows from $\mathbf{u}(\mathbf{Z}'\beta) \in \Theta^0$ for all $\beta \in B$ and $\mathbf{Z} \in \mathcal{Z}$. Then we obtain a largest

admissible set B^* for β as the interior of $\{\beta: \mathbf{u}(\mathbf{Z}'\beta) \in \Theta^0 \text{ for all } \mathbf{Z} \in \mathcal{Z}\}$. Since B^* is convex for natural link functions, the convexity assumption in (i) is a natural one.

The log likelihood of a sample $\mathbf{y}_1, \cdots, \mathbf{y}_n$ is given by

$$l_n(\beta) = \sum_{i=1}^n (\theta_i' \mathbf{y}_i - b(\theta_i)) - C, \quad \theta_i = \mathbf{u}(\mathbf{Z}_i'\beta), \quad i = 1, \cdots, n,$$

where C does not depend on β.

Setting $\mu_n(\beta) = \mu(\mathbf{u}(\mathbf{Z}_n'\beta))$, $\mathbf{\Sigma}_n(\beta) = \mathbf{\Sigma}(\mathbf{u}(\mathbf{Z}_n'\beta))$, $\mathbf{U}_n(\beta) = [\partial \mathbf{u}(\mathbf{Z}_n'\beta)/\partial\gamma]'$ and differentiating $l_n(\beta)$, we find the score function $\mathbf{s}_n(\beta)$ and the information matrix $\mathbf{F}_n(\beta)$ to be

$$\mathbf{s}_n(\beta) = \partial l_n(\beta)/\partial\beta = \sum_{i=1}^n \mathbf{Z}_i\mathbf{U}_i(\beta)(\mathbf{y}_i - \mu_i(\beta)),$$

$$\mathbf{F}_n(\beta) = \operatorname{cov}_\beta \mathbf{s}_n(\beta) = \sum_{i=1}^n \mathbf{Z}_i\mathbf{U}_i(\beta)\mathbf{\Sigma}_i(\beta)\mathbf{U}_i'(\beta)\mathbf{Z}_i'.$$

Further differentiation yields

$$\mathbf{H}_n(\beta) = -\partial^2 l_n(\beta)/\partial\beta\partial\beta' = \mathbf{F}_n(\beta) - \mathbf{R}_n(\beta),$$

say. The matrix $\mathbf{R}_n(\beta)$ is given by

$$\mathbf{R}_n(\beta) = \sum_{i=1}^n \sum_{r=1}^q \mathbf{Z}_i\mathbf{W}_{ir}(\beta)\mathbf{Z}_i'(y_{ir} - \mu_{ir}(\beta)), \tag{2.2}$$

where $\mathbf{W}_{ir}(\beta) = \partial^2 u_r(\mathbf{Z}_i'\beta)/\partial\gamma\partial\gamma'$, and $u_r(\gamma)$, y_{ir}, $\mu_{ir}(\beta)$ are the components of $\mathbf{u}(\gamma)$, $\mathbf{y}_i$, $\mu_i(\beta)$. It is easy to see that $E_\beta \mathbf{s}_n(\beta) = \mathbf{0}$, $E_\beta \mathbf{H}_n(\beta) = \mathbf{F}_n(\beta)$.

For natural link functions, these expressions simplify considerably:

$$\mathbf{s}_n(\beta) = \sum_{i=1}^n \mathbf{Z}_i(\mathbf{y}_i - \mu_i(\beta)), \quad \mathbf{F}_n(\beta) = \sum_{i=1}^n \mathbf{Z}_i\mathbf{\Sigma}_i(\beta)\mathbf{Z}_i', \quad \mathbf{H}_n(\beta) = \mathbf{F}_n(\beta). \tag{2.3}$$

The last equality is of great advantage: Since $\mathbf{F}_n(\beta)$ is positive definite, the likelihood function is concave so that the convexity of B ensures uniqueness of the maximum if it exists. Moreover, consistency and normality conditions on $\mathbf{H}_n(\beta)$, which arise from Taylor expansions, shrink to relatively mild conditions on $\mathbf{F}_n(\beta)$. For this reason, we treat the case of natural link functions separately in Section 3.

The MLE $\hat{\beta}_n$ based on a finite sample $\mathbf{y}_1, \cdots, \mathbf{y}_n$ does, in general, not exist for each possible sample, but only for a subset of the sample space. However, the MLE $\hat{\beta}_n$ can be extended to a random variable on the whole sample space, using techniques as in Witting and Noelle (1970, page 77); Kaufmann (1983, page 37f.), e.g. For convenience we assume that the sequences $\{\hat{\beta}_n\}$ are already random variables, i.e. measurable functions, defined on the whole sample space.

For notational simplicity, we shall mostly drop the argument β_0 in $\mathbf{s}_n(\beta_0)$, $\mathbf{F}_n(\beta_0)$, E_{β_0}, P_{β_0} etc. and write $\mathbf{s}_n$, $\mathbf{F}_n$, E, P etc.

3. Natural link functions.

3.1 *Theorems on consistency and asymptotic normality.* In the sequel we need square roots of positive definite matrices. Let $\mathbf{A}^{1/2}(\mathbf{A}^{T/2})$ be a left (the corresponding right) square root of the positive definite matrix $\mathbf{A}$, i.e. $\mathbf{A}^{1/2}\mathbf{A}^{T/2} = \mathbf{A}$. In addition, set $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$, $\mathbf{A}^{-T/2} = (\mathbf{A}^{T/2})^{-1}$. Note that left (right) square roots

are unique up to an orthogonal transformation from the right (from the left). Unique continuous "versions" of the square root are the Cholesky square root (e.g. Stoer, 1976, page 146) and the symmetric positive definite square root (e.g. Gourieroux and Monfort, 1981). The left Cholesky square root is defined as the (unique) lower triangular matrix with positive diagonal elements and can be computed without solving any eigenvalue problems.

Beyond the regularity assumptions of Subsection 2.2 we need some further assumptions to ensure consistency or asymptotic normality. Usually the following conditions on the sequence of information matrices have to be checked for all $\beta_0 \in B$; the constants involved may depend on β_0.

Define the sequence $N_n(\delta)$, $\delta > 0$, of neighborhoods of β_0 as

$$N_n(\delta) = \{\beta: \| \mathbf{F}_n^{T/2}(\beta - \beta_0) \| \leq \delta\}, \quad n = 1, 2, \cdots. \tag{3.1}$$

(D) *Divergence*: $\lambda_{\min}\mathbf{F}_n \to \infty$.

(C) *Boundedness from below*: for all $\delta > 0$,

$$\mathbf{F}_n(\beta) - c\mathbf{F}_n \text{ positive semidefinite}, \quad \beta \in N_n(\delta), \quad n \geq n_1,$$

with some constants $n_1 = n_1(\delta)$, $c > 0$ independent of δ.

(N) *Convergence and continuity*: for all $\delta > 0$,

$$\max_{\beta \in N_n(\delta)} \| \mathbf{V}_n(\beta) - \mathbf{I} \| \to 0,$$

where $\mathbf{V}_n(\beta) = \mathbf{F}_n^{-1/2}\mathbf{F}_n(\beta)\mathbf{F}_n^{-T/2}$ is the *normed information matrix.*

(S_δ) *Boundedness of the eigenvalue ratio*: there is a neighborhood $N \subset B$ of β_0 such that

$$\lambda_{\min}\mathbf{F}_n(\beta) \geq c(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta}, \quad \beta \in N, \quad n \geq n_1,$$

with some constants $c > 0$, $\delta > 0$, n_1.

Condition (C) is equivalent to the assertion that for any $\delta > 0$ there is a n_1 such that

$$\lambda'\mathbf{F}_n(\beta)\lambda \geq c\lambda'\mathbf{F}_n\lambda, \quad \text{for all} \quad \lambda \in \mathbb{R}^p, \quad \beta \in N_n(\delta), \quad n \geq n_1. \tag{3.2}$$

In other words, the information grows in the neighborhood $N_n(\delta)$ at least as fast as in β_0, uniformly in all directions and for all $n \geq n_1$. Conditions ($S_{1/2}$) and (D) imply (C): Under (D) the neighborhoods $N_n(\delta)$ shrink to β_0. Estimating the left side of (3.2) from below by $\lambda_{\min}\mathbf{F}_n(\beta)$ and the right side from above by $c\ \lambda_{\max}\mathbf{F}_n$, we see that ($S_{1/2}$) implies (C). Condition ($S_{1/2}$) is much stronger since it mixes different directions. Condition (N) is equivalent to the assertion that for any $\delta > 0$ and $\varepsilon > 0$ there is a n_1 such that

$$|\lambda'\mathbf{F}_n(\beta)\lambda - \lambda'\mathbf{F}_n\lambda| \leq \varepsilon\ \lambda'\mathbf{F}_n\lambda \quad \text{for all} \quad \lambda \in \mathbb{R}^p, \quad \beta \in N_n(\delta), \quad n \geq n_1. \tag{3.3}$$

In other words, the relative difference in information is arbitrarily small within the neighborhoods $N_n(\delta)$ of β_0, uniformly in all directions and for all $n \geq n_1$. Note that (N) does not depend on the version of the square root. From (3.3) and

(3.2) it is easy to see that (N) implies (C). Example (iii) in Subsection 3.4 shows that (C) is strictly weaker than (N).

Now we state the asymptotic results.

THEOREM 1. *Under* (D) *and* (C), *there is a sequence* $\{\hat{\beta}_n\}$ *of random variables with*

(i) $P(\mathbf{s}_n(\hat{\beta}_n) = \mathbf{0}) \to 1$ (*asymptotic existence*),

(ii) $\hat{\beta}_n \to_p \beta_0$ (*weak consistency*).

THEOREM 2. *Under* (D) *and* (S_δ) *with a* $\delta > 0$, *there is a sequence* $\{\hat{\beta}_n\}$ *of random variables and a random number* n_2 *with*

(i) $P(\mathbf{s}_n(\hat{\beta}_n) = \mathbf{0}$ *for all* $n \geq n_2) = 1$ (*asymptotic existence*),

(ii) $\hat{\beta}_n \to_{\text{a.s.}} \beta_0$ (*strong consistency*).

LEMMA 1. *Under* (D) *and* (N), *the normed score function is asymptotically normal*:

$$\mathbf{F}_n^{-1/2}\mathbf{s}_n \to_d N(\mathbf{0}, \mathbf{I}).$$

THEOREM 3. *Under* (D) *and* (N), *the normed MLE is asymptotically normal*:

$$\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I}).$$

REMARKS. (i) The conclusions of Lemma 1 and Theorem 3 are valid for any version of the square root. This follows from a general invariance property of sequences $\{\mathbf{x}_n\}$ of asymptotically normal variables: for any sequence $\{\mathbf{P}_n\}$ of nonstochastic, orthogonal matrices

$$\mathbf{x}_n \to_d N(\mathbf{0}, \mathbf{I}) \quad \text{implies} \quad \mathbf{P}_n\mathbf{x}_n \to_d N(\mathbf{0}, \mathbf{I}). \tag{3.4}$$

PROOF. If $\mathbf{x}_n \to_d N(\mathbf{0}, \mathbf{I})$, then the characteristic function $\varphi(\mathbf{x}_n; \lambda)$ of $\mathbf{x}_n$ converges to $\exp(-\|\lambda\|^2/2)$, uniformly on any compact set. Let $\lambda_n = \mathbf{P}_n\lambda$ for arbitrary, but fixed λ. From the uniformity of convergence on $\|\lambda_n\| = \|\lambda\|$, we have $\varphi(\mathbf{P}_n\mathbf{x}_n; \lambda) = \varphi(\mathbf{x}_n; \mathbf{P}_n\lambda) \to \exp(-\|\lambda\|^2/2)$, which in turn implies $\mathbf{P}_n\mathbf{x}_n \to N(\mathbf{0}, \mathbf{I})$.

(ii) Haberman (1977a) proves weak consistency and asymptotic normality of the MLE by a different approach, using fixed point theorems. It can be shown that his Condition 2 and, in his notation, $d_n f_n^2 \to 0$ are equivalent to condition (D) together with the following condition:

(N′) for all $\delta > 0$,

$$\sup_{\beta \in N_n(\delta), \beta \neq \beta_0} \frac{\|\mathbf{V}_n(\beta) - \mathbf{I}\|}{\|\mathbf{F}_n^{T/2}(\beta - \beta_0)\|} \to 0.$$

Compared to (N), the expression under $\sup(\cdot)$ is additionally divided by $\|\mathbf{F}_n^{T/2}(\beta - \beta_0)\|$. Since $\|\mathbf{F}_n^{T/2}(\beta - \beta_0)\| \leq \delta$, condition (N′) implies condition (N). However, the difference between the two conditions seems to be small, in that it

seems difficult to construct examples of practical relevance where condition (N) is fulfilled but condition (N′) is not. To understand this, consider the one-dimensional case $p = 1$. Here (N) resp. (N′) reduce to the requirement that, for any $\delta > 0$,

$$\max_{\beta \in N_n(\delta)} \frac{|F_n(\beta) - F_n|}{F_n} \to 0 \text{ resp. } \sup_{\beta \in N_n(\delta), \beta \neq \beta_0} \frac{|F_n(\beta) - F_n|}{F_n^{3/2} |\beta - \beta_0|} \to 0.$$

Under many circumstances, for instance if the function $F_n(\beta)$ is convex or concave in the interval $N_n(\delta)$, the supremum is, for both expressions, achieved at one of the boundary points. Up to the constant $\delta > 0$, the same sequence is required to converge to zero, and (N) and (N′) are equivalent. Nevertheless, condition (N) seems easier to be interpreted and to be checked.

Instead of norming by a square root of the information matrix, Haberman (1977a) states the asymptotic normality of any nontrivial linear functional, i.e. he states that for any $\lambda \neq \mathbf{0}$

$$\frac{\lambda'(\hat{\beta}_n - \beta_0)}{(\lambda' \mathbf{F}_n^{-1} \lambda)^{1/2}} \to_d N(0, 1). \tag{3.5}$$

However, both assertions are equivalent: If $\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$, it follows from (3.4) that $\lambda_n' \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$ for any sequence $\{\lambda_n\}$ with $\lambda_n' \lambda_n = 1$. The choice $\lambda_n = \mathbf{F}_n^{-1/2} \lambda / (\lambda' \mathbf{F}_n^{-1} \lambda)^{1/2}$, $n = 1, 2 \cdots$ yields (3.5). Conversely, define, for an arbitrary λ with $\lambda' \lambda = 1$, the sequence $\{\lambda_n\}$ in the same manner. Choose an orthogonal transformation $\mathbf{P}_n$ with $\lambda_n = \mathbf{P}_n \lambda$, $n = 1, 2, \cdots$. From (3.5), it follows that $\lambda_n' \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) = \lambda' \mathbf{P}_n \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(0, 1)$. This implies $\mathbf{P}_n \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$, and (3.4) gives the conclusion.

We found that matrix norming is useful in deducing asymptotic efficiency of the MLE in Fisher's or Wolfowitz' sense via the LAN condition (e.g. Basawa and Scott, 1983, page 23) and may lead directly to χ^2-distributions in tests of linear hypotheses.

(iii) In applications it is necessary to replace the norming matrix $\mathbf{F}_n^{T/2}$ by $\mathbf{F}_n^{T/2}(\hat{\beta}_n)$. Using (3.8) as in the proof of Theorem 3, it can be seen that this is possible if instead of (N) the following stronger condition holds:

(Q) for all $\delta > 0$, $\sup_{\beta \in N_n(\delta)} \| \mathbf{F}_n^{-1/2} \mathbf{F}_n^{1/2}(\beta) - \mathbf{I} \| \to 0$.

It is easy to see that (Q) implies (N). If $\mathbf{A} \mapsto \mathbf{A}^{1/2}$ is the Cholesky square root, (N) conversely implies (Q): if $\mathbf{F}_n^{1/2}$, $\mathbf{F}_n^{1/2}(\beta)$ are lower triangular matrices with positive diagonal entries, the matrix $\mathbf{F}_n^{-1/2} \mathbf{F}_n^{1/2}(\beta)$ shares the same properties. Hence this matrix is the Cholesky square root of $\mathbf{F}_n^{-1/2} \mathbf{F}_n(\beta) \mathbf{F}_n^{-T/2}$, and the assertion follows from the continuity of the Cholesky square root. Note that we cannot argue in this way, if $\mathbf{A}^{1/2}$ is the symmetric positive definite square root, since a product of symmetric matrices is generally not symmetric. Thus, use of the Cholesky square root is advantageous not only from the computational but from the theoretical viewpoint, too.

For an arbitrary continuous square root, condition (N) implies condition (Q)

if with a constant c

$$\lambda_{\max}\mathbf{F}_n/\lambda_{\min}\mathbf{F}_n \leq c < \infty, \quad n \geq n_0. \tag{3.6}$$

This can be demonstrated with similar arguments to those of Gourieroux and Monfort (1981, page 85).

Remarks (i), (iii) apply to all our results on asymptotic normality and will not be mentioned in the sequel.

3.2 *Proofs.* PROOF OF THEOREM 1. Due to the positive definiteness of $\mathbf{F}_n(\beta)$ and the convexity of B, there is at most one zero of the score function. This zero gives a (local and global) maximum of the likelihood if it exists. Define the MLE $\hat{\beta}_n$ as the only zero of the score function if it exists, and as an arbitrary constant in B otherwise.

For any n, $\delta > 0$ the event

$$l_n(\beta) - l_n(\beta_0) < 0 \quad \text{for all } \beta \in \partial N_n(\delta),$$

implies that there is a local maximum inside of $N_n(\delta)$. From the uniqueness properties discussed above, this maximum must be located at $\hat{\beta}_n$. It is shown below that for any $\eta > 0$ there exist $\delta > 0$ and n_1 such that

$$P(l_n(\beta) - l_n(\beta_0) < 0 \quad \text{for all} \quad \beta \in \partial N_n(\delta)) \geq 1 - \eta, \tag{3.7}$$

for all $n \geq n_1$, inducing (i) and, in view of (D), (ii) of Theorem 1. Moreover, (3.7) implies that for any $\eta > 0$ there exist $\delta > 0$ and n_1 such that

$$P(\| \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \| \leq \delta) \geq 1 - \eta \quad \text{for all} \quad n \geq n_1. \tag{3.8}$$

This result will be used in the proof of Theorem 3.

Now, with $\lambda = \mathbf{F}_n^{T/2}(\beta - \beta_0)/\delta$, the Taylor expansion of the log likelihood becomes

$$l_n(\beta) - l_n(\beta_0) = \delta\lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n - \delta^2\lambda'\mathbf{V}_n(\tilde{\beta})\lambda/2, \quad \lambda'\lambda = 1,$$

where $\tilde{\beta}$ lies between β and β_0. Using $\max \lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n = \| \mathbf{F}_n^{-1/2}\mathbf{s}_n \|$ for $\lambda'\lambda = 1$, we recognize that it suffices to show for any $\eta > 0$

$$P(\| \mathbf{F}_n^{-1/2}\mathbf{s}_n \|^2 < \delta^2\lambda_{\min}^2\mathbf{V}_n(\tilde{\beta})/4) \geq 1 - \eta \tag{3.9}$$

for some $\delta > 0$ and sufficiently large n. From (C) and the Markov inequality—note that $E \| \mathbf{F}_n^{-1/2}\mathbf{s}_n \|^2 = p$—we get

$$P(\| \mathbf{F}_n^{-1/2}\mathbf{s}_n \|^2 < (\delta c)^2/4) \geq 1 - 4p/(\delta c)^2 = 1 - \eta$$

for $\delta^2 = 4p/(c^2\eta)$ and sufficiently large n, implying (3.9), (3.7) and (3.8).

PROOF OF THEOREM 2. Theorem 2 is proved in a similar way as Theorem 1. Given an arbitrary $\varepsilon > 0$ with $K_\varepsilon(\beta_0) = \{\beta: \| \beta - \beta_0 \| \leq \varepsilon\}$ contained in the neighborhood N of condition (S_δ), we now demonstrate that, with a random

number n_2, the event

$$(3.10)\quad l_n(\beta) - l_n(\beta_0) < 0 \quad \text{for all} \quad \beta \text{ with } \|\beta - \beta_0\| = \varepsilon, \quad \text{for all} \quad n \geq n_2,$$

has probability one. This supplies the stronger form of asymptotic existence as well as the strong consistency of the MLE.

Now, with $\lambda = (\beta - \beta_0)/\varepsilon$, the Taylor expansion of the log likelihood becomes

$$l_n(\beta) - l_n(\beta_0) = \varepsilon\lambda'\mathbf{s}_n - \varepsilon^2\lambda'\mathbf{F}_n(\tilde{\beta})\lambda/2,$$

where $\tilde{\beta}$ lies between β and β_0. Dividing by $(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta}$, we see that the event

$$(3.11)\qquad \frac{\lambda'\mathbf{s}_n}{(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta}} < \frac{\varepsilon}{2}\,\frac{\lambda'\mathbf{F}_n(\tilde{\beta})\lambda}{(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta'}} \quad \lambda'\lambda = 1, \quad n \geq n_2$$

is equivalent to the event (3.10). Each component of $\mathbf{s}_n$ has a variance less than $\lambda_{\max}\mathbf{F}_n$. A componentwise application of Kolmogorov's strong law of large numbers (e.g. Wu, 1981, Lemma 2) shows that $\mathbf{s}_n/(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta} \to_{\text{a.s.}} \mathbf{0}$. By the Cauchy-Schwarz inequality, the left side in (3.11) converges to zero a.s., uniformly for all λ with $\lambda'\lambda = 1$. From condition (S_δ), the right side in (3.11) is bounded from below by $c\varepsilon/2$ if $n \geq n_1$. Hence the event (3.11) resp. (3.10) has probability one.

It should be remarked that n_2 can be chosen measurable.

PROOF OF LEMMA 1. This lemma is proven by showing that the moment generating function $E\exp(\delta\lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n)$ of the linear combination $\lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n$ with $\lambda'\lambda = 1$ converges to the moment generating function of the standard normal distribution.

Fix the scalar δ and the vector λ with $\lambda'\lambda = 1$. For the sequence $\beta_n = \beta_0 + \delta\mathbf{F}_n^{-T/2}\lambda$, $n = 1, 2, \cdots$, we have $\beta_n \in N_n(\delta)$. The Taylor expansion of the log likelihood is

$$l_n(\beta_n) = l_n(\beta_0) + (\beta_n - \beta_0)'\mathbf{s}_n - (\beta_n - \beta_0)'\mathbf{F}_n(\tilde{\beta}_n)(\beta_n - \beta_0)/2,$$

with $\tilde{\beta}_n$ on the line segment between β_n and β_0. Taking exponentials and rearranging yields

$$\exp(\lambda'\mathbf{V}_n(\tilde{\beta}_n)\lambda\delta^2/2)L_n(\beta_n) = \exp(\delta\lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n)L_n(\beta_0),$$

where $L_n(\beta)$ denotes the likelihood. The left side is integrable since $\exp(\lambda'\mathbf{V}_n(\beta)\lambda\delta^2/2)$ is a continuous function of β, and therefore is bounded on the compact line segment between β_n and β_0. Integrating both sides with respect to the dominating measure, we obtain

$$(3.12)\qquad E_{\beta_n}\exp(\lambda'\mathbf{V}_n(\tilde{\beta}_n)\lambda\ \delta^2/2) = E\ \exp(\delta\lambda'\mathbf{F}_n^{-1/2}\mathbf{s}_n).$$

Condition (N), $\tilde{\beta}_n \in N_n(\delta)$ and the continuity of the exponential function together imply that, for any $\varepsilon > 0$, there exists a nonrandom number n_1 with

$$|\exp(\lambda'\mathbf{V}_n(\tilde{\beta}_n)\lambda\ \delta^2/2) - \exp(\delta^2/2)| \leq \varepsilon, \quad n \geq n_1.$$

Integration of this inequality together with $|\int \cdot| \leq \int|\cdot|$ shows that the left side of equality (3.12), and therefore the right side converges to the moment

generating function $\exp(\delta^2/2)$ of the standard normal. From the continuity theorem for moment generating functions, it follows that $\lambda' \mathbf{F}_n^{-1/2} \mathbf{s}_n$ is asymptotically a standard normal variable. Since λ is an arbitrary unit vector, Lemma 1 holds.

PROOF OF THEOREM 3. As usual, $\mathbf{s}_n = \mathbf{s}_n(\beta_0)$ is expanded about $\hat{\beta}_n$. From the mean value theorem for vector valued functions (e.g. Heuser, 1981, page 278) we obtain

$$\mathbf{s}_n = \left[\int_0^1 \mathbf{F}_n(\beta_0 + t(\hat{\beta}_n - \beta_0))\, dt \right] (\hat{\beta}_n - \beta_0)$$

where integration is to be read elementwise, and

$$\mathbf{F}_n^{-1/2} \mathbf{s}_n = \left[\int_0^1 \mathbf{V}_n(\beta_0 + t(\hat{\beta}_n - \beta_0))\, dt \right] \mathbf{F}_n^{T/2} (\hat{\beta}_n - \beta_0). \tag{3.13}$$

From (N) and $\| \int \cdot \| \leq \int \| \cdot \|$ we have, for any $\varepsilon > 0$,

$$\left\| \int_0^1 \mathbf{V}_n(\beta_0 + t(\hat{\beta}_n - \beta_0))\, dt - \mathbf{I} \right\| \leq \int_0^1 \varepsilon\, dt = \varepsilon$$

if n is large enough and if, with some $\delta > 0$, $\hat{\beta}_n$ is in $N_n(\delta)$. In view of (3.8), δ can be chosen so that the probability of this event is arbitrarily close to 1. Hence,

$$\int_0^1 \mathbf{V}_n(\beta_0 + t(\hat{\beta}_n - \beta_0))\, dt \to_p \mathbf{I},$$

and with the continuity theorem applied to formula (3.13) the conclusion of Theorem 3 follows from Lemma 1.

3.3 *Verification of the conditions and some examples.* First we treat some examples to show how the conditions reduce for special exponential families. Subsequently we give some corollaries of general interest. Proofs are deferred to Subsection 3.4.

EXAMPLES. (i) The Poisson model. Let the independent counts $\{y_n\}$ have densities

$$P(y_n = y) = \exp(\theta_n y - e^{\theta_n})/y!, \quad y = 0, 1, 2, \cdots, \; n = 1, 2, \cdots,$$

where $\theta_n = \mathbf{z}_n' \beta_0$. The information matrix for this model is

$$\mathbf{F}_n(\beta) = \textstyle\sum_1^n \mathbf{z}_i \mathbf{z}_i' \exp(\mathbf{z}_i' \beta).$$

Assume the admissible set B to be $\mathbb{R}^p$.

Condition (D) allows no general reduction. However, it can be shown (Subsection 3.4) that, under (D), condition (N) is implied by

$$\mathbf{z}_n' \mathbf{F}_n^{-1} \mathbf{z}_n \to 0. \tag{3.14}$$

Since condition (C) is implied by condition (N), weak consistency and asymptotic

normality of the MLE hold under (D) and (3.14). For fixed dimension p, these conditions are equivalent to Condition 2 of Haberman (1977b, page 1154).

Concerning the growth of $\{\mathbf{z}_n\}$, the critical bound is $\{\ln n\}$, at least in the simple case of a scalar sequence $\{z_n\}$. If $\{z_n\}$ is bounded away from zero, and $z_n = o(\ln n)$, (D) and (3.14) hold (Subsection 3.4). Hence, sublogarithmic growth is admissible. On the other hand, if $z_n \geq c \ln n$, $n \geq n_1$, with some constants $c > 0$, n_1, then condition (D) fails to hold for all $\beta_0 \in \mathbb{R}$ (Subsection 3.4). Hence, logarithmic or superlogarithmic growth is not admissible.

(ii) Gamma distributed variates. Consider the family of gamma densities

$$f(y \mid \theta, r) = \Gamma(r)^{-1}(-\theta)^r y^{r-1}\exp(\theta y), \quad y \geq 0,$$

for a fixed shape parameter $r > 0$. This is a natural exponential family in θ with natural parameter space $\Theta = (-\infty, 0)$ and $b(\theta) = -r \ln(-\theta)$, $b'(\theta) = -r\,\theta^{-1}$, $b''(\theta) = r\,\theta^{-2}$. Let the independent observations $\{y_n\}$ be gamma distributed with natural parameter $\theta_n = \mathbf{z}_n'\beta_0$ for the nth observation. The information matrix for this model is $\mathbf{F}_n(\beta) = r \sum_1^n \mathbf{z}_i\mathbf{z}_i'(\mathbf{z}_i'\beta)^{-2}$, $n = 1, 2, \cdots$. Let the admissible set B be some nonempty, open, convex subset of $B_* = \{\beta: \mathbf{z}_n'\beta < 0, n = 1, 2, \cdots\}$.

The set B_* must be nonempty. This requirement already restricts the possible sequences of regressors: all vectors $\mathbf{z}_n$ have to lie on the same side of some hyperplane through the origin. Beyond that, condition (G_1) below requires the unit vectors $\mathbf{z}_n/\|\mathbf{z}_n\|$ to be bounded away from this hyperplane. This forces B_* to contain interior points, see Subsection 3.4. There it will also be shown that this sharpening is without loss of generality, since otherwise the dimension of β may be reduced. Hence, condition (G_1) merely demands the regression problem to be posed properly.

(G_1) There exists a vector $\beta \in \mathbb{R}^p$ and a constant γ with $\mathbf{z}_n'\beta/\|\mathbf{z}_n\| \leq \gamma < 0$, $n = 1, 2, \cdots$.

Under (G_1), the following condition (G_2) is equivalent to condition (D) and does even imply condition (C) and condition (N) (Subsection 3.4). Hence, weak consistency and asymptotic normality hold under the minimal requirements (G_1) and (D) resp. (G_1) and (G_2). Note that (G_2) does not involve the unknown parameters.

(G_2) $$\lambda_{\min} \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i'/\|\mathbf{z}_i\|^2 \to \infty.$$

Finally, under (G_1), condition (G_2) and the following condition $(G_{3,\delta})$ are equivalent to (D) and (S_δ). Hence, strong consistency holds under (G_1), (G_2), and $(G_{3,\delta})$.

$(G_{3,\delta})$ For some constants $c > 0$, $\delta > 0$, n_1,

$$\lambda_{\min} \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i'/\|\mathbf{z}_i\|^2 \geq c(\lambda_{\max} \sum_{i=1}^n \mathbf{z}_i\mathbf{z}_i'/\|\mathbf{z}_i\|^2)^{1/2+\delta}, \quad n \geq n_1.$$

(iii) This example shows that (C) is weaker than (N). In the exponential family density, we take $c(y)$ as a mixture of normal densities,

$$c(y) = \varepsilon(2\pi)^{-1/2}\exp(-y^2/2) + (1-\varepsilon)(2\pi\sigma^2)^{-1/2}\exp(-y^2/(2\sigma^2)), \quad \sigma^2 > 1.$$

This choice may be motivated as in robust statistics: for $\varepsilon > 0$ heavier tails than

for the normal density can be achieved. From the norming condition

$$\int f(y \mid \theta)\, dy = 1,$$

we obtain

$$b(\theta) = \ln[\varepsilon \exp(\theta^2/2) + (1 - \varepsilon)\exp(\sigma^2\theta^2/2)],$$

with $\Theta = (-\infty, +\infty)$ as the natural parameter space. Differentiation yields

$$\mu(\theta) = b'(\theta) = \theta(1 + (1 - \varepsilon)(\sigma^2 - 1)a(\theta)),$$

$$\sigma^2(\theta) = b''(\theta) = 1 + (1 - \varepsilon)(\sigma^2 - 1)a(\theta) + \theta(1 - \varepsilon)(\sigma^2 - 1)a'(\theta),$$

with

$$a(\theta) = (\varepsilon \exp(-\theta^2(\sigma^2 - 1)/2) + 1 - \varepsilon)^{-1}.$$

It is easily seen that $b'(\theta)/\theta \to \sigma^2$ and $b''(\theta) \to \sigma^2$ as $\theta \to \pm\infty$. For simplicity, we consider scalar sequences $\{z_n\}$. The parameter β may vary in $\mathbb{R}$. Conditions (D) and (C) are fulfilled for all $\beta_0 \in \mathbb{R}$ if

$$\textstyle\sum_{i=1}^n z_i^2 \to \infty.$$

Condition (N) holds for all $\beta_0 \in \mathbb{R}$ if, additionally,

$$z_n^2/\textstyle\sum_{i=1}^n z_i^2 \to 0.$$

However, if $\sum_1^n z_i^2 \to \infty$ but $z_n^2/\sum_1^n z_i^2 \geq k > 0$, then (N) is violated for $\beta_0 = 0$, whereas (C) remains true.

After these examples we consider three situations of practical importance. Sometimes sufficient conditions are of interest which do not involve the parameter β. Such assumptions are given by

(R_c) (i) the sequence $\{\mathbf{Z}_n\}$ lies in a compact set $\mathcal{Z}$ with $\mathbf{Z}'\beta \in \Theta^0$ for all $\mathbf{Z} \in \mathcal{Z}, \beta \in B$,

(R_c) (ii) $\lambda_{\min} \sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i' \to \infty$.

COROLLARY 1 (Regressors with a compact range). *If* (R_c) *holds, then the conditions* (D) *and* (N) *are fulfilled. Hence, the MLE asymptotically exists, is weakly consistent and asymptotically normal. If, in addition,*

$$\lambda_{\min} \textstyle\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i'/(\lambda_{\max} \sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i')^{1/2+\delta} \geq c > 0, \quad n \geq n_1, \tag{3.15}$$

with some constants $c, \delta > 0, n_1$, *then condition* (S_δ) *is satisfied and the MLE is strongly consistent.*

REMARK. The first part of the corollary could also be derived from Condition (3b) of Haberman (1977a). As the proof of Corollary 1 shows, conditions in terms of the higher derivatives can be stated, which are weaker than (R_c) but still sufficient for (N). Furthermore, conditions sufficient for (N), respectively (C), may be given, which resemble (N) respectively (C) in terms of the variances of the observations, see e.g. (3.18). Thus, (D) and (3.18) ensure asymptotic normality

(a related but somewhat weaker result is provided by Condition (3c) of Haberman). By the same reasoning, (D) and

$$\lambda'\Sigma_n(\beta)\lambda \geq c\lambda'\Sigma_n\lambda \quad \text{for all} \quad \lambda \in \mathbb{R}^p, \quad \beta \in N_n(\delta), \quad n \geq n_1,$$

with c, δ, n_1 as in (C), ensure weak consistency.

The next corollary treats the case where the range of $\mathbf{y}$ is bounded, e.g. in the logit model. We require

(R_b) (i) the range of $\mathbf{y}$ is bounded,

(ii) $\lambda_{\min}\mathbf{F}_n \to \infty$,

(iii) $\operatorname{tr} \mathbf{Z}_n'\mathbf{F}_n^{-1}\mathbf{Z}_n \to 0$.

COROLLARY 2 (Responses with a bounded range). *Under* (R_b), *condition* (N) *is fulfilled. Hence, the MLE asymptotically exists, is weakly consistent and asymptotically normal.*

REMARK. (R_b) (iii) is a negligibility condition of the type of Feller's condition in the central limit theorem. The same condition already implied (N) in the Poisson example. (R_b) (ii) and (R_b) (iii) are equivalent to

$$\max_{1\leq i\leq n}\operatorname{tr} \mathbf{Z}_i'\mathbf{F}_n^{-1}\mathbf{Z}_i \to 0.$$

This again can be shown to be equivalent to $b_n^2 \to 0$ in Haberman's Condition 3, implying (N') and (N). Therefore, we omit a verification of our conditions via (3.18).

If the regressors are random, let the pairs $\{(\mathbf{y}_n, \mathbf{Z}_n)\}$ be i.i.d. as the pair $(\mathbf{y}, \mathbf{Z})$, say. Let $\mathbf{F}(\beta) = \mathbf{Z}\Sigma\mathbf{Z}'$, with $\Sigma = \Sigma(\mathbf{Z}'\beta)$, denote the conditional information of the single observation $\mathbf{y}$ given $\mathbf{Z}$. We require the following mild conditions on the marginal distribution of the matrix $\mathbf{Z}$:

(R_s) (i) $E\mathbf{F}$ exists and is positive definite,

(ii) $E \max_{\beta\in N} \| \mathbf{F}(\beta) \|$ exists for a compact neighborhood N of β_0.

REMARKS. (i) If $\mathbf{Z} \in \mathcal{Z}$ is compact, (R_s) reduces to the positive definiteness of $E\mathbf{Z}\mathbf{Z}'$. If the range of $\mathbf{y}$ is bounded, e.g. in the logit model, (R_s) reduces to the existence and positive definiteness of $E\mathbf{Z}\mathbf{Z}'$.

(ii) Condition (R_s) (ii) allows the application of the strong law of large numbers for Banach space valued random variables (e.g. Padgett and Taylor, 1973) to obtain uniform convergence of certain random functions. This is useful in the proof of the following corollary.

COROLLARY 3 (Stochastic regressors). *If* (R_s) *holds, then the conditions* (D), (N) *and* ($S_{1/2}$) *are fulfilled a.s. Hence, the MLE asymptotically exists and is strongly consistent. It is also asymptotically normal:* $\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$.

REMARK. The corollary is first shown to be valid conditionally, given the sequence $\{\mathbf{Z}_n\}$. Then it is demonstrated that the corollary is valid unconditionally as well. Note that the norming matrix $\mathbf{F}_n^{T/2}$ is random. From the law of large numbers and the continuity theorem, we can obtain an unconditional version with nonrandom normalisation:

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, (E\mathbf{F})^{-1}). \tag{3.16}$$

3.4 *Proofs.* PROOFS TO EXAMPLE (i). At first, condition (N) is implied by (D) and (3.14). By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |\, \mathbf{z}_n'(\beta - \beta_0)\,|^2 &= |\, \mathbf{z}_n'\mathbf{F}_n^{-T/2}\mathbf{F}_n^{T/2}(\beta - \beta_0)\,|^2 \\ &\le \mathbf{z}_n'\mathbf{F}_n^{-1}\mathbf{z}_n\delta^2, \quad \text{if } \beta \in N_n(\delta). \end{aligned}$$

From (3.14), it follows that

$$\max_{\beta \in N_n(\delta)} |\exp(\mathbf{z}_n'(\beta - \beta_0)) - 1| \to 0,$$

for any $\delta > 0$. With $\sigma_n^2(\beta) = \exp(\mathbf{z}_n'\beta)$ for the Poisson model, this is equivalent to the statement that for any $\delta, \varepsilon > 0$

$$|\, \sigma_n^2(\beta) - \sigma_n^2(\beta_0)\,| \le \varepsilon\sigma_n^2(\beta_0), \quad \beta \in N_n(\delta), \quad n \ge n_1.$$

For $q = 1$, this is formula (3.18) in the proof of Corollary 1, and condition (N) follows with the same arguments.

Secondly, condition (D) and (3.14) hold, if $\{z_n\}$ is bounded away from zero and $z_n = o(\ln n)$: The statement $z_n = o(\ln n)$ is equivalent to the assertion that, for any $\beta \in \mathbb{R}$, $\gamma > 0$ and $n \ge n_1 = n_1(\beta, \gamma)$, the inequality $\exp(z_n\beta) \ge n^{-\gamma}$ holds. Information may be estimated from below,

$$\textstyle\sum_1^n z_i^2 \exp(z_i\beta) \ge \sum_{n_1}^n z_i^2 i^{-\gamma} \ge c \sum_{n_1}^n i^{-\gamma},$$

with some constant $c > 0$. The last inequality holds since $\{z_n\}$ is bounded away from zero. Condition (D) follows from the divergence of $\sum i^{-\gamma}$ for $\gamma \le 1$. More exactly, if $\gamma < 1$, we have $\sum_{n_1}^n i^{-\gamma} = n^{1-\gamma}(c_* + o(1))$, with some constant $c_* > 0$. This implies that in z_n^2/F_n the denominator diverges at least at a polynomial rate, whereas the numerator, if at all, diverges slower than $(\ln n)^2$. Hence, (3.14) holds.

Finally, if $z_n/\ln n \ge c > 0$, information may be estimated from above. For $\beta < 0$, with constants c_0, n_1, the inequalities

$$\textstyle\sum_1^n z_i^2 \exp(z_i\beta) \le c_0 + c^2 \sum_{n_1}^n (\ln i)^2 i^{c\beta} \le c_0 + c^2 \sum_{n_1}^n i^{2+c\beta}$$

are valid (note that $z^2\exp(z\beta)$ is a monotone decreasing function of z, if $\beta < 0$ and z is large enough). The right side converges for $\beta < -3/c$. Hence (D) is not fulfilled for all $\beta_0 \in \mathbb{R}$.

PROOFS TO EXAMPLE (ii). Condition (G_1) implies that B_* contains interior points: the vector β of (G_1) is an element of B_*. Moreover, since

$$\mathbf{z}_n'\tilde{\beta} = \mathbf{z}_n'(\tilde{\beta} - \beta) + \mathbf{z}_n'\beta \le \|\, \mathbf{z}_n\| \,(\|\, \tilde{\beta} - \beta\,\| + \gamma) < 0,$$

if $\|\, \tilde{\beta} - \beta\,\| < -\gamma$, β is an interior point of B_*.

If (G_1) does not hold, the dimension of β may be reduced: first note that B_* is a convex cone in $\mathbb{R}^p$, i.e. $\alpha B_* \subset B_*$, $\alpha > 0$, and B_* convex. A convex cone in $\mathbb{R}^p$ has empty interior if and only if it is contained in a subspace with dimension less than p. If (G_1) is violated, for any $\beta \in B_*$ there exists a subsequence $\{\mathbf{z}_{n(j)}\}$ with $\mathbf{z}'_{n(j)}\beta / \| \mathbf{z}_{n(j)} \| \to 0$. For the sequence $\beta_j = \beta - \mathbf{z}_{n(j)}(\mathbf{z}'_{n(j)}\beta)/\| \mathbf{z}_{n(j)} \|^2$, $j = 1, 2, \cdots$, we have $\mathbf{z}'_{n(j)}\beta_j = 0$, which implies $\beta_j \notin B_*$, $j = 1, 2, \cdots$, and $\beta_j \to \beta$. Hence, any $\beta \in B_*$ must be a boundary point of B_*. By an application of the theorem mentioned above, the conclusion follows.

Condition (N) and, a fortiori, condition (C) are fulfilled under (G_1) and (G_2): first note that, if condition (G_1) holds for $\beta = \beta_0$, it holds, with some other constant, in a neighborhood of β_0, too. From this fact and

$$| (\beta_0'\mathbf{v})^2 - (\beta'\mathbf{v})^2 | \leq \| \beta_0 + \beta \| \, \| \beta_0 - \beta \| \, \| \mathbf{v} \|^2$$

it follows that, for any $\varepsilon > 0$, there exists a neighborhood N_ε of β_0 where for all β

$$| (\beta'\mathbf{z}_i)^{-2} - (\beta_0'\mathbf{z}_i)^{-2} | \leq \varepsilon (\beta_0'\mathbf{z}_i)^{-2},$$

uniformly for all vectors $\mathbf{z}_i$, $i = 1, 2, \cdots$. Multiplying by $(\lambda'\mathbf{z}_i)^2$ and summing up, we obtain, in the neighborhood N_ε of β_0,

$$| \lambda'\mathbf{F}_n(\beta)\lambda - \lambda'\mathbf{F}_n\lambda | \leq \varepsilon \lambda'\mathbf{F}_n\lambda.$$

Since asymptotically, for any $\delta > 0$, the neighborhood $N_n(\delta)$ is contained in N_ε, condition (N) follows in the form (3.3).

PROOFS TO EXAMPLE (iii). Since $\sigma^2(\theta)$ is bounded from above and bounded away from zero, uniformly for all θ, conditions (D) and (C) hold, if $\sum_1^n z_i^2 \to \infty$, and $z_n^2/F_n \to 0$ follows from $z_n^2/\sum_1^n z_i^2 \to 0$. Furthermore, the derivative of $\sigma^2(\theta)$ is bounded. Therefore $| \sigma_n^2(\beta) - \sigma_n^2(\beta_0) | \leq K | z_n | \, | \beta - \beta_0 |$, with $\beta \in N_n(\delta)$. In fact if $z_n^2/F_n \to 0$, $| \sigma^2(\beta) - \sigma^2(\beta_0) |$ will be arbitrarily small for sufficiently large n, implying (3.18) and (N), by the same arguments as in the proof of Corollary 1. On the other side, if $z_n^2/\sum_1^n z_i^2 \geq k > 0$ and $\beta_0 = 0$, it is easily seen that $| \sigma^2(\beta_n) - \sigma^2(\beta_0) | \geq c_\delta > 0$, with $\beta_n = \beta_0 + \delta/F_n^{1/2}$, for any $\delta > 0$. With similar arguments as in the proof of Corollary 1, we conclude that (N) is violated.

PROOF OF COROLLARY 1. From the compactness assumption (R_c) (i), we have

$$0 < c_1 \leq \lambda_{\min}\Sigma_n \leq \lambda_{\max}\Sigma_n \leq c_2 < \infty, \quad n = 1, 2, \cdots, \tag{3.17}$$

with constants c_1, c_2. From

$$\lambda'\mathbf{F}_n\lambda = \sum_{i=1}^n \lambda'\mathbf{Z}_i\Sigma_i\mathbf{Z}_i'\lambda \geq c_1 \sum_{i=1}^n \lambda'\mathbf{Z}_i\mathbf{Z}_i'\lambda$$

it is seen that (R_c) (ii) implies (D). Using both inequalities in (3.17), it is similarly demonstrated that, with appropriate constants, (3.15) implies (S_δ).

It remains to show that condition (N) holds. From compactness, it follows that the derivatives of $\lambda'\Sigma_n(\beta)\lambda$, $n = 1, 2, \cdots$, with respect to β are bounded, uniformly in n and λ with $\lambda'\lambda = 1$. By Taylor expansion we conclude that for

any $\delta > 0$ and $\varepsilon > 0$ there is a n_1 such that

$$(3.18) \quad |\lambda'\Sigma_n(\beta)\lambda - \lambda'\Sigma_n\lambda| \leq \varepsilon\lambda'\Sigma_n\lambda \quad \text{for all} \quad \lambda \in \mathbb{R}^p, \quad \beta \in N_n(\delta), \quad n \geq n_1.$$

This condition resembles condition (N) in terms of the variances of the observations; see expression (3.3). Continuity of $\Sigma_n(\beta)$ and divergence and monotony of $\mathbf{F}_n$ imply

$$(3.19) \qquad |\lambda'\Sigma_i(\beta)\lambda - \lambda'\Sigma_i\lambda| \leq \varepsilon\lambda'\Sigma_i\lambda, \quad i = 1, 2, \cdots, n,$$

λ, β and n as above. Since (3.19) holds uniformly for all λ, we can substitute $\mathbf{Z}_i\lambda$ for all λ. Summation and the triangle inequality supply condition (N) in the form (3.3).

PROOF OF COROLLARY 3. As functions of β, the matrices $\mathbf{F}_n(\beta)$ are elements of the Banach space of continuous mappings from N into the space of $p \times p$-matrices. Under condition (R_s) (ii), we obtain from the strong law of large numbers for i.i.d. Banach space valued random variables that $\mathbf{F}_n(\beta)/n$ converges a.s. to $E\mathbf{F}(\beta)$, uniformly in $\beta \in N$. Condition (D) follows immediately, a.s. The mean $E\mathbf{F}(\beta)$ is a continuous function of β. With the usual decomposition we obtain that, for any $\varepsilon > 0$, a.s.,

$$\begin{aligned}\|\mathbf{F}_n(\beta)/n - \mathbf{F}_n/n\| &\leq \|\mathbf{F}_n(\beta)/n - E\mathbf{F}(\beta)\| \\ &\quad + \|E\mathbf{F}(\beta) - E\mathbf{F}\| + \|E\mathbf{F} - \mathbf{F}_n/n\| \leq \varepsilon,\end{aligned}$$

if β is sufficiently near to β_0 and $n \geq n_1$, say. From this inequality, condition (N) may be inferred since, under convergence, $\mathbf{F}_n^{-1/2}$ and $\mathbf{F}_n^{-T/2}$ may be substituted by $n^{-1/2}$.

For symmetric matrices $\mathbf{A}$, $\mathbf{B}$ it generally holds that $|\lambda_{\min}\mathbf{A} - \lambda_{\min}\mathbf{B}|$, $|\lambda_{\max}\mathbf{A} - \lambda_{\max}\mathbf{B}| \leq c\|\mathbf{A} - \mathbf{B}\|$ with a constant $c > 0$ depending only on the used norm. Hence, the eigenvalues $\lambda_{\min}\mathbf{F}_n(\beta)/n$, $\lambda_{\max}\mathbf{F}_n(\beta)/n$ converge a.s. to $\lambda_{\min}E\mathbf{F}(\beta)$ resp. $\lambda_{\max}E\mathbf{F}(\beta)$, uniformly in $\beta \in N$, too. Thus, for any $\varepsilon > 0$, we have

$$(3.20) \qquad \lambda_{\min}\mathbf{F}_n(\beta)/\lambda_{\max}\mathbf{F}_n \geq (\lambda_{\min}E\mathbf{F}(\beta) - \varepsilon)/(\lambda_{\max}E\mathbf{F} + \varepsilon)$$

for sufficiently large n, uniformly in $\beta \in N$, a.s. From continuity, the eigenvalue $\lambda_{\min}E\mathbf{F}(\beta)$, $\beta \in N$, is bounded away from zero. If $\varepsilon > 0$ is chosen small enough, the right side of (3.20) is bounded away from zero, and ($S_{1/2}$) follows a.s.

Since we estimate β_0 from the conditional likelihood of $\mathbf{y}_1, \cdots, \mathbf{y}_n$, the conclusions of the corollary at first refer to the corresponding conditional probability measure. However, they remain valid unconditionally too. For strong consistency, this is seen by the law of total probability. The a.s. conditional convergence of $\mathbf{t}_n = \mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0)$ to a standard normal variate $\mathbf{t}$ is equivalent to $E(g(\mathbf{t}_n) \mid \{\mathbf{Z}_n\}) \to_{\text{a.s.}} E\,g(\mathbf{t})$ for any bounded continuous functional g. Integration yields $E\,g(\mathbf{t}_n) \to E\,g(\mathbf{t})$, by dominated convergence. Hence, the convergence in distribution holds unconditionally too. For weak consistency we could argue similarly.

4. Nonnatural link functions.

4.1 *Discussion of additional problems.* Nonnatural link functions enlarge the class of generalized linear models, but they cause additional difficulties in establishing consistency and asymptotic normality, arising mainly from the fact that in general $\mathbf{H}_n(\beta) \neq \mathbf{F}_n(\beta)$. Consequently, the uniqueness of the MLE cannot be guaranteed except in special cases (for a number of important examples see Wedderburn, 1976), since the log likelihood is generally not concave in B. Therefore, a local maximum in a neighborhood of the true β_0 does not necessarily define a global maximum, and we can only prove asymptotic existence and consistency of a sequence $\{\hat{\beta}_n\}$ of solutions of the maximum likelihood equations $\mathbf{s}_n(\beta) = \mathbf{0}$. Yet we can still follow the line of arguments in Section 3 to prove consistency and asymptotic normality, if we replace $\mathbf{F}_n(\beta)$ by $\mathbf{H}_n(\beta)$ and formulate appropriate modifications of the assumptions (K), (S_δ) and (N) (recall that $\mathbf{F}_n(\beta) = \sum_1^n \mathbf{Z}_i\mathbf{U}_i(\beta)\mathbf{\Sigma}_i(\beta)\mathbf{U}_i'(\beta)\mathbf{Z}_i'$ in the case of nonnatural link functions). Weak and strong consistency conditions are:

(C*) for all $\delta > 0$,

$$P(\mathbf{H}_n(\beta) - c\mathbf{F}_n \text{ positive semidefinite for all } \beta \in N_n(\delta)) \to 1,$$

with some constant $c > 0$, c independent of δ.

(S_δ^*) there is a neighborhood $N \subset B$ of β_0 such that a.s.

$$\lambda_{\min}\mathbf{H}_n(\beta) \geq c(\lambda_{\max}\mathbf{F}_n)^{1/2+\delta}, \quad \beta \in N, \quad n \geq n_1,$$

with some constants c, $\delta > 0$ and a random number n_1.

It is easily verified that Theorems 1 (resp. 2) of Section 3 remain true under $\lambda_{\min}\mathbf{F}_n \to \infty$ and (C*) (resp. (S_δ^*)) for a sequence $\{\hat{\beta}_n\}$, as discussed above. The form of $\mathbf{R}_n(\beta)$ in the decomposition $\mathbf{H}_n(\beta) = \mathbf{F}_n(\beta) - \mathbf{R}_n(\beta)$ suggests that it may be shown that $\mathbf{F}_n(\beta)$ dominates $\mathbf{R}_n(\beta)$ for large n, applying some law of large numbers. We will pursue this idea in Section 4.2.

Modifying the normality condition (N) in the analogous way we arrive at

(N*) for all $\delta > 0$,

$$\max_{\beta \in N_n(\delta)} \| \mathbf{V}_n(\beta) - \mathbf{I} \| \to_p 0,$$

where $\mathbf{V}_n(\beta) = \mathbf{F}_n^{-1/2}\mathbf{H}_n(\beta)\mathbf{F}_n^{-T/2}$.

As in Section 3, (N*) implies (C*).

Now a reexamination of the proofs of Lemma 1 and Theorem 3 exhibits the following: If the normed score function $\mathbf{F}_n^{-1/2}\mathbf{s}_n$ is asymptotically normally distributed, then the method of proof for Theorem 3 works under (N*) again. However, to extend the proof of the asymptotic normality of $\mathbf{F}_n^{-1/2}\mathbf{s}_n$ in Lemma 1, a strengthened version of (N*) is required,

(N**) for all $\delta > 0$ and $\lambda \in \mathbb{R}^p$, $\lambda'\lambda = 1$,

$$\max_{\beta \in N_n(\delta)} \| \mathbf{V}_n(\beta) - \mathbf{I} \| \to 0 \text{ in probability under } P \text{ and } P_{\beta_n},$$

with $\beta_n = \beta_0 + \delta\mathbf{F}_n^{-T/2}\lambda$.

Under (D) and (N**) Lemma 1 and, hence, Theorem 3 remain true. A detailed proof of this statement would follow the line of arguments in Kaufmann (1983, page 54). In the next subsection we shall not demonstrate asymptotic normality of the normed MLE via (N**), but give conditions which assure that (N*) is fulfilled and establish $\mathbf{F}_n^{-1/2}\mathbf{s}_n \to_d N(\mathbf{0}, \mathbf{I})$ via the Lindeberg-Feller central limit theorem. However, it should be remarked that, under (D), this is equivalent to a direct verification of (N**): If (D), (N*) and $\mathbf{F}_n^{-1/2}\mathbf{s}_n \to_d N(\mathbf{0}, \mathbf{I})$ are true, then it can be deduced by the same arguments as in Basawa and Scott (1983, page 26) that P and the sequence $\{P_{\beta_n}\}$ are contiguous. Thus, (N**) holds. Conversely, (N**) implies (N*) and, together with (D), $\mathbf{F}_n^{-1/2}\mathbf{s}_n \to_d N(\mathbf{0}, \mathbf{I})$.

For an application of Haberman's fixed point approach in the different, but related, context of survival models we refer to Friedman (1982).

4.2 *Verification of the conditions.* We treat the same cases of general interest as in Subsection 3.3. The assumptions (R_c), (R_b) and (R_s) have to be modified:

(R_c^*) (i) the sequence $\{\mathbf{Z}_n\}$ lies in a compact set $\mathcal{Z}$ with $\mathbf{u}(\mathbf{Z}'\beta) \in \Theta^0$ for all $\mathbf{Z} \in \mathcal{Z}, \beta \in B$,

(ii) $\lambda_{\min} \sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i' \to \infty$.

(R_b^*) (i) the range of $\mathbf{y}$ is bounded,

(ii) $\lambda_{\min}\mathbf{F}_n \to \infty$,

(iii) $\operatorname{tr} \mathbf{Z}_n'\mathbf{F}_n^{-1}\mathbf{Z}_n \to 0$,

(iv) $\sum_{i=1}^n \operatorname{tr} \mathbf{Z}_i'\mathbf{F}_n^{-1}\mathbf{Z}_i \le c < \infty$,

(v) the first and second derivatives of $\mathbf{u}$ are bounded, the second derivative is uniformly continuous.

For stochastic regressors, let the pairs $(\mathbf{y}_n, \mathbf{Z}_n)$ be i.i.d. as the pair $(\mathbf{y}, \mathbf{Z})$, say. Define the random matrices $\mathbf{F}(\beta) = \mathbf{F}_1(\beta)$ and $\mathbf{R}(\beta) = \mathbf{R}_1(\beta)$ where $\mathbf{Z}_1$ is replaced by $\mathbf{Z}$. Then we require the following condition:

(R_s^*) (i) $E\mathbf{F}$ exists and is positive definite,

(ii) $E \max_{\beta \in N} \| \mathbf{F}(\beta) \|$ and $E \max_{\beta \in N} \| \mathbf{R}(\beta) \|$ exist for a compact neighborhood $N \subset B$ of β_0.

The remark (i) after condition (R_s) in Subsection 3.3 remains valid if, additionally, the first and second derivatives of $\mathbf{u}$ are bounded.

LEMMA 2. *Under* (R_c^*), *or* (R_b^*) *without* (iv), *or* (R_s^*) (i), *the normed score function is asymptotically normal*:

$$\mathbf{F}_n^{-1/2}\mathbf{s}_n \to_d N(\mathbf{0}, \mathbf{I}).$$

The following theorem refers to regressors with a compact range or to responses with a bounded range.

THEOREM 4. *Under* (R_c^*) *or* (R_b^*) *there exists a sequence* $\{\hat{\beta}_n\}$ *of random*

variables with

(i) $P(\mathbf{s}_n(\hat{\beta}_n) = \mathbf{0}) \to 1$ *(asymptotic existence),*

(ii) $\hat{\beta}_n \to_p \beta_0$ *(weak consistency),*

(iii) $\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$ *(asymptotic normality).*

If the regressors are stochastic, or have a compact range and additionally obey an eigenvalue condition, then even strong consistency can be achieved.

THEOREM 5. *Suppose* (R_c^*) *and*

$$\lambda_{\min} \sum_1^n \mathbf{Z}_i\mathbf{Z}_i' \geq c\, \lambda_{\max} \sum_1^n \mathbf{Z}_i\mathbf{Z}_i', \quad n \geq n_1 \tag{4.1}$$

to hold, with some constants $c > 0$, n_1. *Alternatively, let* (R_s^*) *hold. Then there exists a sequence* $\{\hat{\beta}_n\}$ *of random variables and a random number* n_2 *with*

(i) $P(\mathbf{s}_n(\hat{\beta}_n) = \mathbf{0}$ *for all* $n \geq n_2) = 1$ *(asymptotic existence),*

(ii) $\hat{\beta}_n \to_{\text{a.s.}} \beta_0$ *(strong consistency),*

(iii) $\mathbf{F}_n^{T/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{I})$ *(asymptotic normality).*

REMARK. As in Subsection 3.3, an unconditional version of Theorem 5 holds under (R_s^*):

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to_d \mathrm{N}(\mathbf{0}, (E\mathbf{F})^{-1}).$$

EXAMPLES. (i) A special example, where (R_c^*) holds, are grouped data. The compact set $\mathcal{Z}$ consists of a finite number I of regressors $\mathbf{Z}_1, \cdots, \mathbf{Z}_I$. There are n_i observations for each $\mathbf{Z}_i$. If $n_i/n \to \lambda_i > 0$ and $\sum_1^I \mathbf{Z}_i\mathbf{Z}_i'$ has full rank, it follows that

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to_d N(\mathbf{0}, \mathbf{V}^{-1}), \quad \mathbf{V} = \sum_1^I \lambda_i \mathbf{Z}_i\mathbf{U}_i\mathbf{\Sigma}_i\mathbf{U}_i'\mathbf{Z}_i'.$$

A similar result can be obtained from Bradley and Gart (1962, part I) in a paper on associated populations. Part II of this paper would also give results for $I \to \infty$. However, the conditions are far stronger and more difficult to check than ours.

(ii) Binomial models. If we slightly strengthen the assumptions on the derivatives of u in (R_b^*), then (iv) of (R_b^*) can be dropped in the case of binary responses: assume u to be three times continuously differentiable, u', u'', u''' to be bounded and u' to be bounded away from zero. Since $u' = 1$, $u'' = 0$ for natural link functions, these assumptions mean that g must not depart too much from the natural link function; they are fulfilled e.g. in the probit model. If, additionally,

$$\max_{1 \leq i \leq n} \mathbf{z}_i'\mathbf{F}_n^{-1}\mathbf{z}_i \to 0$$

holds, then the MLE is weakly consistent and asymptotically normal as in Theorem 4.

4.3 *Proofs.* PROOF OF LEMMA 2. We use the Lindeberg-Feller theorem for triangular arrays. Fix λ with $\lambda'\lambda = 1$. For the triangular array

$$z_{ni} = \lambda' \mathbf{F}_n^{-1/2} \mathbf{Z}_i \mathbf{U}_i (\mathbf{y}_i - \boldsymbol{\mu}_i)$$

we have

$$Ez_{ni} = 0, \quad \text{var} \sum_{i=1}^n z_{ni} = \text{var}\, \lambda' \mathbf{F}_n^{-1/2} \mathbf{s}_n = 1.$$

We shall show that the Lindeberg condition is satisfied, i.e. for any $\delta > 0$

$$g_n(\delta) = \sum_{i=1}^n \int_{\{z^2 > \delta^2\}} z^2 \, dF_{ni} \to 0, \tag{4.2}$$

where F_{ni} is the distribution function of z_{ni}.

(i) Regressors with a compact range or responses with a bounded range: Set $\alpha_{ni}' = \lambda' \mathbf{F}_n^{-1/2} \mathbf{Z}_i \mathbf{U}_i$. Insertion into (4.2) and use of $z_{ni}^2 \le \alpha_{ni}' \alpha_{ni} \| \mathbf{y}_i - \boldsymbol{\mu}_i \|^2$ (Cauchy-Schwarz inequality) provides

$$g_n(\delta) \le \sum_{i=1}^n \alpha_{ni}' \alpha_{ni} \int_{B(n,i)} y^2 dG_{\mathbf{Z}_i}. \tag{4.3}$$

Here $G_{\mathbf{Z}}$ is the distribution function of $\| \mathbf{y} - \boldsymbol{\mu}(\mathbf{Z}'\beta_0) \|$ for a given $\mathbf{Z}$, and $B(n, i)$ is the set $\{y^2 > \delta^2 / \alpha_{ni}' \alpha_{ni}\}$. From the compactness of $\mathcal{Z}$ and (R_c^*) (ii) it follows that $\lambda_{\min} \mathbf{F}_n \to \infty$. Using the compactness of $\mathcal{Z}$ again, respectively (R_b^*) (ii), (iii), (v), we get

$$\max_{i=1,\cdots,n} \alpha_{ni}' \alpha_{ni} \to 0. \tag{4.4}$$

Thus, defining

$$h_c(\mathbf{Z}) = \int_{\{y^2 > \delta_c^2\}} y^2 \, dG_{\mathbf{Z}},$$

we have, for any (large) $c > 0$,

$$\int_{B(n,i)} y^2 \, dG_{\mathbf{Z}_i} \le h_c(\mathbf{Z}_i), \quad i = 1, \cdots, n, \; n \ge n_2(c),$$

and, inserting into (4.3),

$$g_n(\delta) \le \sum_{i=1}^n \alpha_{ni}' \alpha_{ni} h_c(\mathbf{Z}_i), \quad n \ge n_2(c). \tag{4.5}$$

Under (R_b^*) (i), $h_c(\mathbf{Z}_i)$ vanishes identically for c large enough, since $\| \mathbf{y}_i - \boldsymbol{\mu}_i \|$ is bounded, whereas $\delta^2 c \to \infty$. Thus $g_n(\delta) \to 0$.

Under (R_c^*) (i), $\sum \alpha_{ni}' \alpha_{ni} \le K < \infty$ with a constant K. Therefore,

$$g_n(\delta) \le K \max_{\mathbf{Z} \in \mathbf{z}} h_c(\mathbf{Z}), \quad n \ge n_2(c).$$

It remains to show that $\max h_c(\mathbf{Z}) \to 0$ as $c \to \infty$. The function $h_c(\mathbf{Z})$ has the following properties: $h_c(\mathbf{Z})$ is continuous in $\mathbf{Z}$ (by the Helly-Bray Lemma and continuity properties of exponential families), $h_c(\mathbf{Z}) \to 0$ (pointwise for any $\mathbf{Z} \in \mathcal{Z}$), and $h_c(\mathbf{Z})$ is monotonously decreasing in c. Due to the last property and

the compactness of $\mathcal{Z}$, pointwise convergence implies uniform convergence, i.e. $\max h_c(\mathbf{Z}) \to 0$ as $c \to \infty$.

(ii) Stochastic regressors: Redefine α'_{ni} as $\alpha'_{ni} = \lambda' \mathbf{F}_n^{-1/2} \mathbf{Z}_i \mathbf{U}_i \boldsymbol{\Sigma}_i^{1/2}$, and $G_{\mathbf{Z}}$ in (4.3) as the distribution function of $\| \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}'\beta_0)\,(\mathbf{y} - \mu(\mathbf{Z}'\beta_0)) \|$. From $\mathbf{F}_n/n \to \mathbf{F}$ a.s., it can be deduced that (4.4) holds a.s. Thus (4.5) remains valid a.s., providing

$$g_n(\delta) \leq \| n\mathbf{F}_n^{-1} \| \, \| 1/n \textstyle\sum_{i=1}^n \mathbf{Z}_i \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{U}_i' \mathbf{Z}_i' h_c(\mathbf{Z}_i) \|, \quad n \geq n_2(c), \quad \text{a.s.}$$

The first term $\| n\mathbf{F}_n^{-1} \|$ is uniformly bounded, a.s. For any fixed $\delta > 0$, the strong law of large numbers yields (note that $h_c(\mathbf{Z}_i) \leq p$)

$$\| 1/n \textstyle\sum_{i=1}^n \mathbf{Z}_i \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{U}_i' \mathbf{Z}_i' h_c(\mathbf{Z}_i) \| \to_{\text{a.s.}} \| E\mathbf{F} h_c(\mathbf{Z}) \|.$$

Since $h_c(\mathbf{Z}) \to 0$ as $c \to \infty$, an application of the dominated convergence theorem proves that $\| E\mathbf{F} h_c(\mathbf{Z}) \| \to 0$ as $c \to \infty$. Choosing first c and then n large enough, it follows that $g_n(\delta)$ will be arbitrarily small for sufficiently large n, a.s.

PROOF OF THEOREM 4. Under both conditions, divergence of the information and, from Lemma 2, $\mathbf{F}_n^{-1/2}\mathbf{s}_n \to N(\mathbf{0}, \mathbf{I})$ hold. Hence, following the discussion in Subsection 4.1, it is sufficient to establish condition (N*). To keep notation simple, we consider only univariate responses ($q = 1$). Multivariate responses can be treated similarly. At first, assume (R_c^*) to hold.

For $q = 1$, let $u_n'(\beta)$, $u_n''(\beta)$ be the first respectively second derivative of $u(\gamma)$, evaluated at $\mathbf{z}_n'\beta$. We have $\mathbf{F}_n(\beta) = \sum_1^n \mathbf{z}_i\mathbf{z}_i' (u_i'(\beta))^2 \sigma_i^2(\beta)$ and $\mathbf{H}_n(\beta) = \mathbf{F}_n(\beta) - \mathbf{R}_n(\beta)$ with $\mathbf{R}_n(\beta) = \sum_1^n \mathbf{z}_i\mathbf{z}_i' u_i''(\beta)(y_i - \mu_i(\beta))$. Condition (N*) will be verified via the decomposition

$$\mathbf{V}_n(\beta) - \mathbf{I} = \mathbf{A}_n(\beta) - \mathbf{B}_n - \mathbf{C}_n(\beta) - \mathbf{D}_n(\beta),$$

with

$$\begin{aligned}
\mathbf{A}_n(\beta) &= \textstyle\sum_1^n \alpha_{ni}\alpha_{ni}'[(u_i'(\beta))^2\sigma_i^2(\beta) - (u_i')^2\sigma_i^2], \\
\mathbf{B}_n &= \textstyle\sum_1^n \alpha_{ni}\alpha_{ni}' u_i''(y_i - \mu_i), \\
\mathbf{C}_n(\beta) &= \textstyle\sum_1^n \alpha_{ni}\alpha_{ni}'(u_i''(\beta) - u_i'')(y_i - \mu_i), \\
\mathbf{D}_n(\beta) &= \textstyle\sum_1^n \alpha_{ni}\alpha_{ni}' u_i''(\beta)(\mu_i - \mu_i(\beta))
\end{aligned}$$

where $\alpha_{ni} = \mathbf{F}_n^{-1/2}\mathbf{z}_i$, $i = 1, \cdots, n$, $n \geq n_0$. Since $(u_n')^2\sigma_n^2$, $n = 1, 2, \cdots$, is bounded away from zero, it holds, with a constant c, that

$$\textstyle\sum_1^n \| \alpha_{ni}\alpha_{ni}' \| = \sum_1^n \alpha_{ni}'\alpha_{ni} \leq c < \infty, \quad n \geq n_0. \tag{4.6}$$

Fix $\delta > 0$. Using (4.6), we have

$$\max_{\beta \in N_n(\delta)} \| \mathbf{A}_n(\beta) \| \leq c \max_{\beta \in N_n(\delta)} \max_{i=1,\cdots,n} | (u_i'(\beta))^2\sigma_i^2(\beta) - (u_i')^2\sigma_i^2 |,$$

$n \geq n_0$, Since

$$\begin{aligned}
&\max_i | (u_i'(\beta))^2\sigma_i^2(\beta) - (u_i')^2\sigma_i^2 | \\
&\quad \leq \max_{\mathbf{z} \in \mathcal{Z}} | (u'(\mathbf{z}'\beta))^2\sigma^2(u(\mathbf{z}'\beta)) - (u'(\mathbf{z}'\beta_0))^2\sigma^2(u(\mathbf{z}'\beta_0)) |,
\end{aligned} \tag{4.7}$$

and the right side of (4.7) is a continuous function of β with a zero at $\beta = \beta_0$, we obtain

$$\max_{\beta \in N_n(\delta)} \| \mathbf{A}_n(\beta) \| \to 0. \tag{4.8}$$

Next look at $\mathbf{B}_n \to_p \mathbf{0}$. The variance of the (s, t)-element of $\mathbf{B}_n$ is (recall $E\mathbf{B}_n = \mathbf{0}$)

$$\operatorname{var}(\mathbf{B}_n)_{st} = \sum_1^n \alpha_{ni,s}^2 \alpha_{ni,t}^2 (u_i'')^2 \sigma_i^2,$$

where $\alpha_{ni,s}$ denotes the sth component of α_{ni}, and can be estimated from above:

$$\begin{aligned} \operatorname{var}(\mathbf{B}_n)_{st} &\le \sum_1^n (\alpha_{ni}' \alpha_{ni})^2 (u_i'')^2 \sigma_i^2 \\ &\le c_* \max_{i=1,\cdots,n} \alpha_{ni}' \alpha_{ni}, \quad n \ge n_0, \end{aligned}$$

with a constant $c_* < \infty$. The rightmost inequality follows from (R_c^*) (i), which gives an upper bound to $(u_i'')^2 \sigma_i^2$, $i = 1, 2, \cdots$, and (4.6). Furthermore, the compactness assumption implies $\max_{i=1,\cdots,n} \alpha_{ni}' \alpha_{ni} \to 0$. This yields $\mathbf{B}_n \to \mathbf{0}$ in quadratic mean and in probability.

For $\max_{\beta \in N_n(\delta)} \| \mathbf{C}_n(\beta) \|$, convergence to zero can be demonstrated in the first mean:

$$E \max_{\beta \in N_n(\delta)} \| \mathbf{C}_n(\beta) \| \le c \max_{\beta \in N_n(\delta), i=1,\cdots,n} | u_i''(\beta) - u_i'' | \max_{i=1,\cdots,n} E | y_i - \mu_i |.$$

With analogous arguments as in the proof of (4.8) it can be shown that the right side of this inequality, and hence the left, converges to zero.

Finally, $\max \| \mathbf{D}_n(\beta) \| \to 0$ can be demonstrated similarly as (4.8). Condition (N*) is established under (R_c^*).

If (R_b^*) holds, we obtain condition (N*) using the same decomposition of $\mathbf{V}_n(\beta) - \mathbf{I}$. The inequality (4.6) is equivalent with (R_b^*) (iv). The function $(u'(\gamma))^2 \sigma^2(u(\gamma))$ is uniformly continuous, since its derivative is uniformly bounded. Using this fact instead of (4.7), assertion (4.8) follows from (R_b^*) (ii), (iii). The other terms in the decomposition may be handled by the parallel arguments, too.

PROOF OF THEOREM 5. (i) Regressors with a compact range: Assertion (iii) of the theorem is included in Theorem 4. To obtain the assertions (i) and (ii) of the theorem, we demonstrate that $(\mathrm{S}_{1/2}^*)$ holds under (R_c^*) and (4.1). Set $\lambda_n = \lambda_{\max} \mathbf{F}_n$, $n = 1, 2, \cdots$. Due to $\mathbf{H}_n(\beta) = \mathbf{F}_n(\beta) - \mathbf{R}_n(\beta)$ we have

$$\lambda_{\min} \mathbf{H}_n(\beta)/\lambda_n \ge \lambda_{\min} \mathbf{F}_n(\beta)/\lambda_n - \| \mathbf{R}_n(\beta) \|/\lambda_n.$$

Similarly as in the proof of Corollary 1, it can be shown that $\lambda_{\min} \mathbf{F}_n(\beta)/\lambda_n$ is bounded away from zero, uniformly in n, in a neighborhood of β_0. Condition $(\mathrm{S}_{1/2}^*)$ holds if $\| \mathbf{R}_n(\beta) \|/\lambda_n$ is arbitrarily small for sufficiently large n in a (sufficiently small) neighborhood N of β_0. This can be verified via the decomposition $\mathbf{R}_n(\beta) = \mathbf{B}_n + \mathbf{C}_n(\beta) + \mathbf{D}_n(\beta)$ with

$$\begin{aligned} \mathbf{B}_n &= \sum_1^n \mathbf{z}_i \mathbf{z}_i' u_i''(y_i - \mu_i), \\ \mathbf{C}_n(\beta) &= \sum_1^n \mathbf{z}_i \mathbf{z}_i' (u_i''(\beta) - u_i'')(y_i - \mu_i), \\ \mathbf{D}_n(\beta) &= \sum_1^n \mathbf{z}_i \mathbf{z}_i' u_i''(\beta)(\mu_i - \mu_i(\beta)). \end{aligned}$$

(For simplicity, we consider only univariate responses, as in the proof of Theorem 4).

For any λ with $\lambda'\lambda = 1$, we have

$$\text{var } \lambda'\mathbf{B}_n\lambda = \sum_1^n (\lambda'\mathbf{z}_i)^2(u_i'')^2\sigma_i^2 \leq c \sum_1^n (\lambda'\mathbf{z}_i)^2, \quad n = 1, 2, \cdots,$$

with some constant c. Using (R_c^*) (i) again, it follows that $\sum_1^n (\lambda'\mathbf{z}_i)^2 \leq c_*\lambda_n$, $n = 1, 2, \cdots$, with some constant c_*. Hence, $\lambda'\mathbf{B}_n\lambda/\lambda_n$ converges a.s. to zero, by an application of Lemma 2 of Wu (1981). Since λ with $\lambda'\lambda = 1$ was arbitrary and pointwise convergence of quadratic forms on the unit ball implies uniform convergence, $\| \mathbf{B}_n \|/\lambda_n$ converges to zero, a.s. Now, consider the second term in the decomposition of $\mathbf{R}_n(\beta)$:

$$\| \mathbf{C}_n(\beta) \|/\lambda_n \leq \max_{i=1,\cdots,n} | u_i''(\beta) - u_i'' | \sum_1^n \mathbf{z}_i'\mathbf{z}_i | y_i - \mu_i | /\lambda_n.$$

Using again Lemma 2 of Wu (1981), it follows that

$$\sum_1^n \mathbf{z}_i'\mathbf{z}_i(| y_i - \mu_i | - E | y_i - \mu_i |)/\lambda_n \to 0, \text{ a.s.}$$

Since $\sum_1^n \mathbf{z}_i'\mathbf{z}_i E | y_i - \mu_i | /\lambda_n$ is bounded, uniformly in n, the term $\sum_1^n \mathbf{z}_i'\mathbf{z}_i | y_i - \mu_i | /\lambda_n$ is a.s. bounded, uniformly in n. Since $\max_{i=1,\cdots,n,\beta\in N} | u_i''(\beta) - u_i'' |$ can be made arbitrarily small in a neighborhood N of β_0, the same result holds for

$$\max_{n=1,2,\cdots,\beta\in N} \| \mathbf{C}_n(\beta) \|/\lambda_n, \text{ a.s.}$$

The third term in the decomposition of $\mathbf{R}_n(\beta)$ can be handled with similar, but simpler arguments, since $\mathbf{D}_n(\beta)/\lambda_n$ is nonrandom. Collecting together, we obtain the desired result on $\| \mathbf{R}_n(\beta) \|/\lambda_n$, and $(S_{1/2}^*)$ follows.

(ii) Stochastic regressors: With the same arguments as in the proof of Corollary 3, we obtain

$$\mathbf{F}_n(\beta)/n \to E\mathbf{F}(\beta), \quad \mathbf{R}_n(\beta)/n \to \mathbf{0}$$

uniformly in $\beta \in N$, a.s. Condition (D) follows immediately a.s., (N*) and $(S_{1/2}^*)$ similarly as in the proof of Corollary 3.

PROOF TO EXAMPLE (ii). We use the same decomposition of $\mathbf{V}_n(\beta) - \mathbf{I}$ as in the proof of Theorem 4. It may suffice to consider the term $\mathbf{C}_n(\beta)$. For any λ with $\lambda'\lambda = 1$, we have

$$\begin{aligned} &E \max_{\beta\in N_n(\delta)} \lambda'\mathbf{C}_n(\beta)\lambda \\ &\leq \max_{i,\beta} \frac{| u_i''(\beta) - u_i'' |}{| u_i' |^2} \max_i \frac{E | y_i - \mu_i |}{\sigma_i^2} \sum_{i=1}^n (\lambda'\mathbf{F}_n^{-1/2}\mathbf{z}_i)^2 | u_i' |^2\sigma_i^2 \\ &\leq \max_{i,\beta} \frac{| u_i'''(\tilde{\beta}) |}{| u_i' |^2} \frac{\delta}{\lambda_{\min}\mathbf{F}_n^{1/2}} 2 \cdot 1 \to 0, \end{aligned}$$

since $E | y_i - \mu_i | = 2\sigma_i^2$ for binomial responses, and

$$\sum_1^n (\lambda'\mathbf{F}_n^{-1/2}\mathbf{z}_i)^2(u_i')^2\sigma_i^2 = 1.$$

Acknowledgement. We thank a referee for his valuable comments and suggestions and Dr. Friedemann Ost for his assistance in preparing the English manuscript.

REFERENCES

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* North Holland, Amsterdam.

BASAWA, J. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models.* Springer, Berlin.

BRADLEY, R. A. and GART, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika* **49** 205–214.

DRYGAS, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrsch. verw. Gebiete* **34** 119–127.

EICKER, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Ann. Math. Statist.* **34** 447–456.

FAHRMEIR, L. and KREDLER, CH. (1984). Verallgemeinerte lineare Modelle, in: Fahrmeir, L. and Hamerle, A. (ed.), *Multivariate Statistische Verfahren.* De Gruyter, Berlin.

FRIEDMAN, M. (1982). Piecewise exponential models for survival data with covariates. *Ann. Statist.* **10** 101–113.

GOURIEROUX, C. and MONFORT, A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *J. Econometrics* **17** 83–97.

HABERMAN, S. J. (1974). *The Analysis of Frequency Data.* Univ. of Chicago Press, Chicago.

HABERMAN, S. J. (1977a). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841.

HABERMAN, S. J. (1977b). Log-linear models and frequency tables with small expected cell counts. *Ann. Statist.* **5** 1148–1169.

HABERMAN, S. J. (1984). *Log-concave Likelihoods and Maximum Likelihood Estimation.* Preprint.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and its Application.* Academic, New York.

HEUSER, H. (1981). *Lehrbuch der Analysis, Teil 2.* Teubner, Stuttgart.

KAUFMANN, H. (1983). Mehrdimensionale Maximum Likelihood Schätzung bei stochastischen Prozessen: Asymptotische Theorie. Dissertation, Universität Regensburg.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9** 343–361.

MATHIEU, J. R. (1981). Tests of χ^2 in the generalized linear model. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 509–527.

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.

MCFADDEN, D. (1974). Conditional logit analysis of qualitative choice behaviour, in: Zarembka, P., *Frontiers in Econometrics.* Academic, New York.

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

NORDBERG, L. (1980). Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models. *Scand. J. Statist.* **7** 27–32.

PADGETT, W. Y. and TAYLOR, R. L. (1973). *Laws of Large Numbers for Normed Linear Spaces and Certain Fréchet Spaces.* Springer, Berlin.

STOER, J. (1976). *Einführung in die Numerische Mathematik I,* 2nd ed. Springer, Berlin.

SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1375–1381.

WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for generalized linear models. *Biometrika* **63** 27–32.

WITTING, H. and NÖLLE, G. (1970). *Angewandte Mathematische Statistik.* Teubner, Stuttgart.
WU, C. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

UNIVERSITÄT REGENSBURG
FACHBEREICH WIRTSCHAFTSWISSENSCHAFT
LEHRSTUHL FUR STATISTIK
8400 REGENSBURG
UNIVERSITÄTSSTRASSE 31 POSTFACH
WEST GERMANY