Report AE9/77

Asymptotic Properties of Maximum Likelihood Estimators in a Nonlinear Regression Model with Unknown Parameters in the Disturbance Covariance Matrix

by R.D.H. Heijmans
J.R. Magnus

Instituut voor Actuariaat & Econometrie

Classification: AMS(MOS) subject classification scheme (1970): 62E20, 62F10, 62J05, 62P20

Keywords: Asymptotic properties, information matrix, maximum likelihood estimation, nonlinear regression

Abstract: For the nonlinear regression model $y_t = X_t(\beta) + \varepsilon_t$ where the vector ε is distributed $N(0,\Omega(\theta))$ it is shown that under fairly general condition the maximum likelihood estimator of θ and β are consistent and asymptotically normal distributed.

October 1977

Free copies are obtainable from:
Bibliotheek - Instituut voor Actuariaat en Econometrie
Jodenbreestraat 23 - Amsterdam 'phone: 020-5254287/17

Universiteit van Amsterdam

Asymptotic Properties of Maximum Likelihood Estimators in a Nonlinear Regression Model with Unknown Parameters in the Disturbance Covariance Matrix

By Risto D.H. Heijmans and Jan R. Magnus
University of Amsterdam, Amsterdam, The Netherlands

1. *Introduction and summary*

Almost all results on the asymptotic properties of maximum likelihood (ML) estimators relate to the standard cases where the observations are independent and identically distributed random variables. Two recent exceptions are papers by Weiss [8] and Crowder [1]. Both authors study general cases and give conditions for consistency and asymptotic normality. Weiss also discusses asymptotic efficiency in some sense. It appears, however, that their conditions are quite gruesome and hard, if at all, to verify[1].

In the present paper we shall focus on the ML estimation of the parameters in a nonlinear model

$$y_t = \chi_t(\beta) + \varepsilon_t \qquad (t=1...n),$$

where the disturbance vector $\varepsilon = (\varepsilon_1 ... \varepsilon_n)'$ is distributed $N(0,\Omega)$. The positive definite matrix Ω may depend on a fixed number of parameters $\theta = (\theta_1 ... \theta_m)'$. In this case the observations y_t are neither independent nor identically distributed. Therefore new theorems had to be developed. Earlier work on ML estimation of linear regression models with unknown parameters in the disturbance covariance matrix has been done by Hildreth [3] and Magnus [4]. The first author studied asymptotic properties of ML estimators in an autoregressive model. The second derived the ML equations and the information matrix for the general linear model, and studied finite properties of the ML estimators, but little was said about asymptotic properties. The present paper, generalizing both [3] and [4], fills this gap. Throughout there is some stress on precision and verifiability of assumptions.

The plan of this paper is as follows: In section two we present the model and derive the ML equations and the information matrix. The loglikelihood appears to be regular with respect to its first and second derivatives. This is proved in section three. Next, we establish strong consistency of the ML estimators $\hat{\beta}$ and $\hat{\theta}$. At that point we allow ourselves a digression into the study of quadratic forms which may prove of independent interest. Sufficient conditions are derived for the asymptotic normality of $\varepsilon' A \varepsilon$, where A is symmetric and $\varepsilon \approx N(0,\Omega)$. Also a vector generalization is presented. In section 6 this theory is applied to prove the asymptotic normality of $\hat{\beta}$ and $\hat{\theta}$. We conclude the paper with a short discussion of our findings.

---

1) Recently, Vickers [7], by strengthening Weiss' conditions, obtained more tractable results.

2. *The maximum likelihood equations and the information matrix*

We shall consider models of the following structure

(1) $$y_t = \chi_t(\beta) + \varepsilon_t \qquad (t=1...n)$$

or, in vector notation,

(2) $$y = g(\beta) + \varepsilon ,$$

where y contains n observations on the dependent variable, $g(\beta)$ contains the regressors as (nonlinear) functions of k parameters $\beta_1...\beta_k$, and ε is the disturbance vector.

We shall not assume that the errors ε_t are independently or identically distributed. Instead we shall make the following three assumptions:

A1: ε is normally distributed, $E\varepsilon=0$, $E\varepsilon\varepsilon'=\Omega$, where Ω is a positive definite matrix whose elements are twice differentiable functions of a finite and fixed number of parameters $\theta_1,\theta_2,...,\theta_m$, i.e. $\Omega=\Omega(\theta)$, $\theta \in \Theta$.

A2: The χ_t are known twice differentiable functions of the k parameters $\beta_1...\beta_k$. The (n,k) matrix of first derivatives $H=(h_{tj})$ with $h_{tj} = \partial\chi_t/\partial\beta_j$ has full rank. n>k.

A3: The parameters in β are independent from those in θ.

The probability density of y takes the form

(3) $$(2\pi)^{-n/2} |\Omega|^{-\frac{1}{2}} \exp - \tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon.$$

The loglikelihood is

(4) $$\Lambda = \gamma + \tfrac{1}{2} \log|\Omega^{-1}| - \tfrac{1}{2}\varepsilon'\Omega^{-1}\varepsilon ,$$

where γ is a constant.

THEOREM 1

The nonlinear regression model (2) under the assumptions A1, A2, and A3 has the following first-order ML conditions:

(5) $$\begin{cases} \hat{H}'\hat{\Omega}^{-1} e = 0 \\ \operatorname{tr}\left(\dfrac{\partial\Omega^{-1}}{\partial\theta_h}\Omega\right)_{\theta=\hat{\theta}} = e'\left(\dfrac{\partial\Omega^{-1}}{\partial\theta_h}\right)_{\theta=\hat{\theta}} e \qquad (h=1...m) . \end{cases}$$

Here $\hat{\beta}$ and $\hat{\theta}$ denote the ML values of β and θ, $\hat{\Omega}=\Omega(\hat{\theta})$, $\hat{H}=H(\hat{\beta})$, and $e=y-g(\hat{\beta})$.

The information matrix of the loglikelihood function (4) is

$$\Psi = \begin{bmatrix} H'\Omega^{-1}H & 0 \\ 0 & \tfrac{1}{2}\Psi_\theta \end{bmatrix}, \tag{6}$$

where Ψ_θ is a symmetric (m,m) matrix with typical element

$$(\Psi_\theta)_{ij} = \operatorname{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta_i}\,\Omega\right)\left(\frac{\partial\Omega^{-1}}{\partial\theta_j}\,\Omega\right) \qquad (i,j=1\ldots m). \tag{7}$$

PROOF

The proof is basically similar to the proofs of theorems 1, 2 and 3 in Magnus [4], but it is much shorter since we do not need an explicit expression for the Hessian matrix.

Let $V=\Omega^{-1}$, then upon differentiating the loglikelihood (4)

$$\begin{aligned} d\Lambda &= \tfrac{1}{2}\operatorname{tr}V^{-1}(dV) - \varepsilon'V(d\varepsilon) - \tfrac{1}{2}\varepsilon'(dV)\varepsilon \\ &= \tfrac{1}{2}\operatorname{tr}(V^{-1}-\varepsilon\varepsilon')(dV) + \varepsilon'V(dg) \\ &= \tfrac{1}{2}\operatorname{tr}(V^{-1}-\varepsilon\varepsilon')(dV) + \varepsilon'VHd\beta\ . \end{aligned} \tag{8}$$

Necessary for a maximum is that $d\Lambda=0$ for all $d\beta\neq0$ and $d\theta\neq0$. This gives the ML equations (5). The differential of Λ can be explicitly expressed in terms of $(d\theta)$ and $(d\beta)$:

$$d\Lambda = \tfrac{1}{2}(d\theta)'\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right)\operatorname{vec}(V^{-1}-\varepsilon\varepsilon') + (d\beta)'H'V\varepsilon\ . \tag{9}$$

Differentiating (9) yields

$$\begin{aligned} d^2\Lambda = {}& \tfrac{1}{2}(d\theta)'\, d\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right)\operatorname{vec}(V^{-1}-\varepsilon\varepsilon') + \tfrac{1}{2}(d\theta)'\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right) d(\operatorname{vec}V^{-1}) \\ & - \tfrac{1}{2}(d\theta)'\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right) d(\operatorname{vec}\varepsilon\varepsilon') + (d\beta)'\, d(H'V)\varepsilon + (d\beta)'H'V(d\varepsilon)\ . \end{aligned}$$

From here we could proceed as in [4] to derive the Hessian matrix. However, since we only need the information matrix, we can take a considerable shortcut. Taking expectations it is easily seen that the first, third, and fourth term in the above expression vanish. This gives

$$\begin{aligned} -Ed^2\Lambda &= -\tfrac{1}{2}(d\theta)'\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right)d(\operatorname{vec}V^{-1}) + (d\beta)'H'V(dg) \\ &= \tfrac{1}{2}(d\theta)'\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right)(V^{-1}\otimes V^{-1})\left(\frac{\partial\operatorname{vec}V}{\partial\theta}\right)'d\theta + (d\beta)'H'VH(d\beta)\ , \end{aligned}$$

and hence

$$\Psi = \begin{bmatrix} H'VH & 0 \\ 0 & \tfrac{1}{2}\Psi_\theta \end{bmatrix},$$

where

$$\Psi_\theta = \left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)(V^{-1}\otimes V^{-1})\left(\frac{\partial \mathrm{vec}V}{\partial\theta}\right)'$$

with typical element $(\Psi_\theta)_{ij} = \mathrm{tr}\left(\frac{\partial V}{\partial\theta_i}V^{-1}\right)\left(\frac{\partial V}{\partial\theta_j}V^{-1}\right)$. (Q.E.D.)

Remark 1

The linear regression model $y=X\beta+\varepsilon$, of course, is a special case of the structure (2). The relevant formulae are found by putting $H=X$ everywhere.

The ML estimates for β and θ are those values which satisfy (5). If more than one solution of (5) is found, we choose those values which maximize the loglikelihood (4). In this paper we shall not be concerned with how to solve the ML equations (5). Several methods are feasible, e.g. the Newton-Raphson iterative procedure.

3. *The regularity of* Λ

In this section we shall prove the following

lemma 1

The loglikelihood Λ is regular with respect to its first and second derivatives, i.e.

$$Ed\Lambda = 0 \quad\text{and}\quad -Ed^2\Lambda = E(d\Lambda)^2 .$$

proof [2)]

Starting from (8) we have

$$d\Lambda = \tfrac{1}{2}\mathrm{tr}(V^{-1}-\varepsilon\varepsilon')(dV) + \varepsilon'VH(d\beta)$$
$$= \tfrac{1}{2}(\mathrm{vec}dV)'\mathrm{vec}(V^{-1}-\varepsilon\varepsilon') + (d\beta)'H'V\varepsilon .$$

Now, $Ed\Lambda=0$, since $E\varepsilon\varepsilon'=V^{-1}$ and $E\varepsilon=0$.

Further,

$$(d\Lambda)^2 = \tfrac{1}{4}(\mathrm{vec}dV)'\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')[\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')]'\,\mathrm{vec}dV$$
$$+ (d\beta)'H'V\varepsilon\varepsilon'VHd\beta$$
$$+ \tfrac{1}{2}(\mathrm{vec}dV)'[\mathrm{vec}(V^{-1}-\varepsilon\varepsilon')]\varepsilon'VHd\beta .$$

2) We provide an alternative (and hopefully a simplification) to the corresponding proof in [4]. This proof was proposed to us by H. Neudecker.

It is easy to see that $E[\text{vec}(V^{-1}-\varepsilon\varepsilon')]\varepsilon' = 0$, since $E\varepsilon_i\varepsilon_j\varepsilon_k=0$ for all i,j,k. Further,

$$E\text{vec}(V^{-1}-\varepsilon\varepsilon')[\text{vec}(V^{-1}-\varepsilon\varepsilon')]' = \text{var}[\text{vec}(V^{-1}-\varepsilon\varepsilon')] = \text{var}(\text{vec}\varepsilon\varepsilon') = \text{var}(\varepsilon\otimes\varepsilon).$$

This leads to

$$E(d\Lambda)^2 = \frac{1}{n}(\text{vec}dV)' \text{var}(\varepsilon\otimes\varepsilon) \text{vec}dV + (d\beta)'H'VH(d\beta) .$$

At this point we need two results from [5] concerning the symmetric (n^2,n^2) commutation matrix K_n:

$$K_n \text{vec } A = \text{vec } A' \text{ , where A is some (n,n) matrix}$$

$$\text{var}(\varepsilon\otimes\varepsilon) = (I+K_n)(V^{-1}\otimes V^{-1}) .$$

Since dV is a symmetric matrix, we find that

$$\begin{aligned}\frac{1}{4}(\text{vec}dV)' \text{var}(\varepsilon\otimes\varepsilon) \text{vec}dV &= \frac{1}{4}(\text{vec}dV)'(I+K_n)(V^{-1}\otimes V^{-1})\text{vec}dV \\ &= \tfrac{1}{2}(\text{vec}dV)'(V^{-1}\otimes V^{-1})\text{vec}dV .\end{aligned}$$

Thus,

$$\begin{aligned}E(d\Lambda)^2 &= \tfrac{1}{2}(\text{vec}dV)'(V^{-1}\otimes V^{-1})\text{vec}dV + (d\beta)'H'VH(d\beta) \\ &= \tfrac{1}{2}(d\theta)'\left(\frac{\partial\text{vec}V}{\partial\theta}\right)(V^{-1}\otimes V^{-1})\left(\frac{\partial\text{vec}V}{\partial\theta}\right)'d\theta + (d\beta)'H'VH(d\beta) \\ &= -Ed^2\Lambda . \qquad \text{(Q.E.D.)}\end{aligned}$$

4. *Strong consistency of the ML estimators*

When $y_1 \ldots y_n$ are <u>independent</u> observations and ζ a parameter to be estimated, then it is well known that the ML equation has a root with probability 1 as $n\to\infty$, which is consistent for ζ, if the loglikelihood Λ_n is differentiable in an interval including the true value [3]. Rao's proof can be summarized as follows: Let ζ_0 be the true value and consider two values $\zeta_0 \pm \delta$. Since the y_i (i=1...n) are independent we have, as $n\to\infty$

$$\frac{1}{n}\Lambda_n - \frac{1}{n}E\Lambda_n \to 0 \qquad \text{with probability 1.}$$

Thus, as $n\to\infty$

$$\frac{1}{n}[\Lambda_n(\zeta_0\pm\delta) - \Lambda_n(\zeta_0)] < 0 \quad \text{with probability 1.}$$

If $\Lambda_n(\zeta)$ is differentiable in $(\zeta_0\pm\delta)$, then $\Lambda_n(\zeta)$ attains a (local) maximum within $(\zeta_0\pm\delta)$ and its derivative vanishes at that point. A root $\hat{\zeta}$ so located is consistent for ζ_0.

---

3) Rao [6, pp. 364-5].

In the present case, however, we estimate the parameters from a <u>single</u> (vector) observation on y. Therefore we must show that, even when the y_i are <u>not</u> independent, $\frac{1}{n}\Lambda_n$ converges to its expected value with probability 1.

From now on, a subscript will denote the number of observations. Thus, Λ_n Ω_n, etc. denote the loglikelihood and the covariance matrix based on n observations.

The loglikelihood of the first k observations is

$$\Lambda_k = -\frac{k}{2}\log 2\pi + \tfrac{1}{2}\log|\Omega_k^{-1}| - \tfrac{1}{2}\varepsilon_k'\Omega_k^{-1}\varepsilon_k \qquad (k=1\ldots n),$$

where ε_k contains the first k elements of ε_n, and Ω_k is the north-west (k,k) submatrix of Ω_n. We partition Ω_k as follows

$$\Omega_k = \begin{bmatrix} \Omega_{k-1} & d_k \\ d_k' & \omega_{kk} \end{bmatrix}.$$

It is easy to verify that

$$|\Omega_k| = \alpha_k|\Omega_{k-1}| \qquad (k=2\ldots n)$$

and

$$\Omega_k^{-1} = \begin{bmatrix} \Omega_{k-1}^{-1} & 0 \\ 0' & 0 \end{bmatrix} + \frac{1}{\alpha_k} z_k z_k' \qquad (k=2\ldots n),$$

where

$$\alpha_k = \omega_{kk} - d_k'\Omega_{k-1}^{-1}d_k \qquad (k=2\ldots n)$$

and

$$z_k' = \begin{bmatrix} -\Omega_{k-1}^{-1} d_k \\ 1 \end{bmatrix} \qquad (k=2\ldots n).$$

Define

$$\lambda_1 \equiv \Lambda_1 = -\tfrac{1}{2}\log 2\pi\omega_{11} - \frac{1}{2\omega_{11}}\varepsilon_1^2,$$

$$\lambda_k \equiv \Lambda_k - \Lambda_{k-1} = -\tfrac{1}{2}\log 2\pi\alpha_k - \frac{1}{2\alpha_k}(z_k'\varepsilon_k)^2 \qquad (k=2\ldots n).$$

Then

$$\Lambda_n = \sum_i \lambda_i .$$

We shall prove that the λ_i are stochastically independent. Let p_k be some nonstochastic real-valued k-vector and π_k its last element. Then

$$\mathrm{cov}(p_k'\varepsilon_k, z_k'\varepsilon_k) = E(p_k'\varepsilon_k)(\varepsilon_k' z_k) = p_k'\Omega_k z_k =$$

$$= p_k' \begin{bmatrix} \Omega_{k-1} & d_k \\ & \\ d_k' & \omega_{kk} \end{bmatrix} \begin{bmatrix} -\Omega_{k-1}^{-1} d_k \\ \\ 1 \end{bmatrix} = p_k' \begin{bmatrix} 0 \\ \\ \alpha_k \end{bmatrix} = \pi_k \alpha_k \ .$$

The following properties of $z_k'\varepsilon_k$ are now straightforward

(i) $z_k'\varepsilon_k$ is normally distributed, $Ez_k'\varepsilon_k=0$, $\mathrm{var}(z_k'\varepsilon_k)=\alpha_k$,

(ii) $z_h'\varepsilon_h$ and $z_k'\varepsilon_k$ ($h\neq k$) are stochastically independent,

(iii) $z_k'\varepsilon_k$ and ε_1 are stochastically independent.

Thus $\{\lambda_i\}$, $i=1,2,\ldots$ is a sequence of independent random variables. Further

$$\mathrm{var}\,(\lambda_i) = \tfrac{1}{2} \qquad (i=1\ldots n) \ ,$$

$$\sum_i \frac{\mathrm{var}(\lambda_i)}{i^2} = \frac{\pi^2}{12} < \infty \ ,$$

and

$$E\lambda_i = -\tfrac{1}{2} - \tfrac{1}{2}\log 2\pi\alpha_i \ .$$

Hence, by Kolmogorov's theorem (Rao [6, p. 114]), the sequence λ_i obeys the law of large numbers, that is, as $n\to\infty$

$$\frac{1}{n}\sum_{i=1}^{n} \lambda_i - \frac{1}{n}\sum_{i=1}^{n} E(\lambda_i) \to 0 \qquad \text{with probability 1.}$$

Now, since $\sum_i \lambda_i = \Lambda_n$, we have, as $n\to\infty$

$$\frac{1}{n}\Lambda_n - \frac{1}{n}E\Lambda_n \to 0 \qquad \text{with probability 1,}$$

which proves the desired result.

Before turning to the asymptotic normality of $\hat{\beta}$ and $\hat{\theta}$ we shall study the asymptotic behavior of quadratic forms in normal variables. This theory will be applied in section 6.

5. *The asymptotic normality of quadratic forms in normal variables*

Let A_n be some symmetric (n,n) matrix with rank r_n. The elements of A_n may depend upon n which implies that A_{n-1} may not be the north-west submatrix of A_n. As before, ε_n is a normally distributed n-vector, $E\varepsilon_n=0$, $E\varepsilon_n\varepsilon_n'=\Omega_n$, and rank $(\Omega_n)=n$. The elements of Ω_n may also depend upon n.

Consider the quadratic form $\varepsilon_n'A_n\varepsilon_n$. Its expectation and variance are given by

$$E(\varepsilon_n'A_n\varepsilon_n) = tr(A_n\Omega_n)$$

and

$$var(\varepsilon_n'A_n\varepsilon_n) = 2\ tr(A_n\Omega_n)^2 \ .$$

We shall first derive sufficient condition for the asymptotic normality of $\varepsilon_n'A_n\varepsilon_n$.

THEOREM 2

Assume that, as $n \to \infty$

(10)
(i) $r_n \to \infty$,
(ii) $(1/r_n)\ tr(A_n\Omega_n)^2 \to \psi$, some finite positive number,
(iii) $(1/\sqrt{r_n})A_n\Omega_n \to 0$,

then $\varepsilon_n'A_n\varepsilon_n$ is asymptotically normally distributed, that is, as $n \to \infty$

$$(1/\sqrt{r_n})\ \left[\varepsilon_n'A_n\varepsilon_n - tr(A_n\Omega_n)\right] \longrightarrow N(0,2\psi)\ .$$

Remark 2

Condition (i) ensures that $\varepsilon_n'A_n\varepsilon_n$ becomes an infinite sum. The second condition states that $(1/r_n)\ var(\varepsilon_n'A_n\varepsilon_n)$ has a finite limit. The last condition implies that $\lambda(A_n\Omega_n)/\sqrt{r_n} \to 0$ (and vice versa), where $\lambda(.)$ stands for any eigenvalue.

Remark 3

It is well known (Rao [6, p. 188]) that $\varepsilon_n'A_n\varepsilon_n$ forllows a $\chi^2(r_n)$ distribution if and only if $A_n\Omega_nA_n = A_n$. In that case $A_n\Omega_n$ is idempotent and thus

$$r_n = rank(A_n) = rank(A_n\Omega_n) = tr(A_n\Omega_n)\ .$$

Therefore, the conditions (10) are fulfilled, provided only that $r_n\to\infty$. This, of course, is as expected, since a χ^2 distribution converges to a normal distribution.

<u>proof</u>

Since A_n is symmetric and Ω_n^{-1} is positive definite, there exists a non-singular matrix F_n such that

$$F_n' A_n F_n = \Lambda_n \qquad \text{and} \qquad F_n' \Omega_n^{-1} F_n = I_n ,$$

where Λ_n is a diagonal matrix containing the roots of $|A_n - \lambda\Omega_n^{-1}| = 0$.[4)] Note that a root λ so obtained is also a root of $A_n\Omega_n$.

Let $v_n = F_n^{-1}\varepsilon_n$, then $v_n \approx N(0,I_n)$ and

$$\varepsilon_n' A_n \varepsilon_n = v_n' \Lambda_n v_n = \sum_{j=1}^{r_n} \lambda_{nj} v_{nj}^2 ,$$

where λ_{nj} denotes the j-th nonzero root of $A_n\Omega_n$ and v_{nj} the corresponding component of v_n. Obviously

$$(11) \qquad \frac{\varepsilon_n' A_n \varepsilon_n - \mathrm{tr}A_n\Omega_n}{\sqrt{2\mathrm{tr}(A_n\Omega_n)^2}} = \frac{\sum_{j=1}^{r_n} \lambda_{nj}(v_{nj}^2 - 1)}{\sqrt{2\mathrm{tr}(A_n\Omega_n)^2}} = \sum_{j=1}^{r_n} \xi_{nj} ,$$

where

$$(12) \qquad \xi_{nj} = \frac{\lambda_{nj}(v_{nj}^2 - 1)}{\sqrt{2\mathrm{tr}(A_n\Omega_n)^2}} .$$

The ξ_{nj} are stochastically independent variables, and,as v_{nj}^2 follows a $\chi^2(1)$ distribution, it is easy to verify that

$$\mathrm{var}(\xi_{nj}) = \lambda_{nj}^2 \,/\, \mathrm{tr}(A_n\Omega_n)^2 < 1 ,$$

$$\mathrm{var}(\sum_j \xi_{nj}) = \sum_j \mathrm{var}(\xi_{nj}) = 1 ,$$

$$\lim_{n\to\infty} \max_{1\leq j\leq r_n} \mathrm{var}(\xi_{nj}) = \lim_{n\to\infty} \frac{\max_j \lambda_{nj}^2/r_n}{\mathrm{tr}(A_n\Omega_n)^2/r_n} = 0 .$$

The last equality flows from conditions (ii) and (iii). In other words, the stochastic variables ξ_{nj} form an <u>elementary system</u> [5)].

---

4) See Rao [6, p. 41].

5) See Gnedenko [2, p. 332].

In the present case the elementary system is normalized by

$$E\, \xi_{nj} = 0$$

and

$$\sum_j E\, \xi_{nj}^2 = 1\ .$$

We are now in a position to apply a theorem in Gnedenko [2, p. 338] which states that the sequence of distribution functions of $\sum_j \xi_{nj}$ converges to a standard normal distribution if (and only if)

$$\lim_{n\to\infty} \sum_{j=1}^{r_n} \int_{|x|>\tau} x^2\, d\, F_{nj}(x) = 0 \qquad \text{for all } \tau>0\ ,$$

where $F_{nj}(.)$ denotes the distribution function of ξ_{nj}. Now, in view of (11) and condition (ii),

$$\sum_j \xi_{nj} \to N(0,1)$$

is equivalent with

$$(1/\sqrt{r_n})\,\left(\varepsilon_n' A_n \varepsilon_n - tr(A_n \Omega_n)\right) \to N(0,2\psi)\ .$$

The only thing to be shown, then, is that

$$\lim_{n\to\infty} \sum_{j=1}^{r_n} \int_{|x|>\tau} x^2\, f_{nj}(x)\, dx = 0 \qquad \text{for all } \tau>0\ . \tag{13}$$

Here $f_{nj}(.)$ is the density of ξ_{nj}. Let $\sigma_n^2 = 2\, tr(A_n\Omega_n)^2$, then

$$f_{nj}(x) = \begin{cases} \dfrac{\sigma_n \exp - \frac{1}{2}\left(\frac{\sigma_n x+\lambda_{nj}}{\lambda_{nj}}\right)}{\sqrt{2\pi}\,.\,\sqrt{|\lambda_{nj}|\ |\sigma_n x+\lambda_{nj}|}} & \text{for } \dfrac{\sigma_n x+\lambda_{nj}}{\lambda_{nj}} > 0 \\ 0 & \text{elsewhere}\ , \end{cases} \tag{14}$$

which follows from (12) and the fact that v_{nj}^2 is $\chi^2(1)$ distributed. Suppose $\lambda_{nj}<0$ (The case $\lambda_{nj}>0$ can be treated in a similar fashion). Then,

$$\int_{|x|>\tau} x^2 f_{nj}(x)dx = \int_{-\infty}^{-\tau} x^2\, f_{nj}(x)dx + \int_{\tau}^{-\lambda_{nj}/\sigma_n} x^2\, f_{nj}(x)dx\ .$$

For n sufficiently large, the expression $-\lambda_{nj}/\sigma_n$ will be smaller than τ and the second integral on the right hand side will equal zero. So we need only investigate the behavior of the first integral on the right hand side.

Now,

$$\int_{-\infty}^{-\tau} x^2 f_{nj}(x)dx \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\tau} \frac{\sigma_n x^2 \exp - \frac{1}{2}(\frac{\sigma_n x}{\lambda_{nj}} + 1)}{\sqrt{|\lambda_{nj}| \, |\sigma_n x + \lambda_{nj}|}} dx$$

$$= \frac{\sigma_n(\exp-\frac{1}{2})}{\sqrt{2\pi|\lambda_{nj}|}} \int_{-\infty}^{-\tau} \frac{x^2 \exp-\frac{1}{2}(\sigma_n x/\lambda_{nj})}{\sqrt{|\sigma_n x+\lambda_{nj}|}} dx$$

$$= \frac{\sigma_n(\exp-\frac{1}{2})}{\sqrt{2\pi|\lambda_{nj}|}} \int_{\tau}^{\infty} \frac{x^2 \exp-\frac{1}{2}(\sigma_n x/|\lambda_{nj}|)}{\sqrt{\sigma_n x+|\lambda_{nj}|}} dx$$

$$\leqq \frac{(\exp-\frac{1}{2})}{\sqrt{2\pi}} \cdot \frac{\sigma_n}{\sqrt{|\lambda_{nj}|(\sigma_n\tau+|\lambda_{nj}|)}} \int_{\tau}^{\infty} x^2 \exp - \frac{1}{2}(\frac{\sigma_n x}{|\lambda_{nj}|}) dx$$

$$= \frac{(\exp-\frac{1}{2})}{\sqrt{2\pi}} \cdot \frac{\sigma_n}{\sqrt{|\lambda_{nj}|(\sigma_n\tau+|\lambda_{nj}|)}} \cdot \left(\exp\frac{-\sigma_n\tau}{2|\lambda_{nj}|}\right) \cdot \left(\frac{2|\lambda_{nj}|\tau^2}{\sigma_n} + \frac{8\tau\lambda_{nj}^2}{\sigma_n^2} + \frac{16|\lambda_{nj}|^3}{\sigma_n^3}\right)$$

$$= \frac{(\exp-\frac{1}{2})}{\sqrt{2\pi}} \left(\frac{2\tau^2+8\tau|\lambda_{nj}|/\sigma_n+16\lambda_{nj}^2/\sigma_n^2}{\sqrt{1+\sigma_n\tau/|\lambda_{nj}|}}\right) \exp - \frac{\sigma_n\tau}{2|\lambda_{nj}|} .$$

We may then write

$$\lim_{n\to\infty} \sum_j \int_{-\infty}^{\tau} x^2 f_{nj}(x)dx$$

$$\leqslant \frac{(\exp-\frac{1}{2})}{\sqrt{2\pi}} \lim_{n\to\infty} \sum_j \left[\left(\frac{2\tau^2+8\tau|\lambda_{nj}|/\sigma_n+16\lambda_{nj}^2/\sigma_n^2}{\sqrt{1+\sigma_n\tau/|\lambda_{nj}|}}\right) \exp - \frac{\sigma_n\tau}{2|\lambda_{nj}|}\right]$$

$$\leqslant \frac{(\exp-\frac{1}{2})}{\sqrt{2\pi}} \lim_{n\to\infty} \left[\max_j \left(\frac{2\tau^2+8\tau|\lambda_{nj}|/\sigma_n+16\lambda_{nj}^2/\sigma_n^2}{\sqrt{1+\sigma_n\tau/|\lambda_{nj}|}}\right)\right] \left[\sum_j \exp \frac{-\sigma_n\tau}{2|\lambda_{nj}|}\right] .$$

Now,

$$\sum_j \exp - \frac{\sigma_n \tau}{2|\lambda_{nj}|} = \sum_j \left[\sum_{k=0}^{\infty} \frac{(\sigma_n \tau/2|\lambda_{nj}|)^k}{k!} \right]^{-1}$$

$$\leq \sum_j \left[\frac{(\sigma_n \tau/2|\lambda_{nj}|)^2}{2} \right]^{-1} = \frac{8}{\tau^2 \sigma_n^2} \sum_j \lambda_{nj}^2 = \frac{4}{\tau^2} .$$

Further, we know from conditions (ii) and (iii) that $\lim_{n\to\infty} |\lambda_{nj}|/\sigma_n = 0$ for all j. Thus,

$$\lim_{n\to\infty} \sum_j \int_{-\infty}^{\tau} x^2 f_{nj}(x)dx$$

$$\leq \frac{(\exp - \frac{1}{2})}{\sqrt{2\pi}} \cdot \frac{4}{\tau^2} \cdot \lim_{n\to\infty} \max_j \left[\frac{2\tau^2 + 8\tau|\lambda_{nj}|/\sigma_n + 16\lambda_{nj}^2/\sigma_n^2}{\sqrt{1+\sigma_n \tau/|\lambda_{nj}|}} \right] = 0.$$

This proves (13). (Q.E.D.)

The vector generalization of theorem 2, which we shall need in the next section, can be stated as follows

<u>THEOREM 3</u>

Suppose we are given m symmetric (n,n) matrices A_{ni} (i=1...m). Define r_n=rank(A_{n1})[6]. Again, ε_n is a normally distributed n-vector, $E\varepsilon_n = 0$, $E\varepsilon_n \varepsilon_n' = \Omega_n$, rank($\Omega_n$) = n.
Assume that, as $n\to\infty$

(i) $r_n \to \infty$

(ii) let $b = (b_1 \ldots b_m)'$ be some real m-vector, and $r_n(b) = \text{rank}(\sum_{i=1}^{m} b_i A_{ni})$.
Then there exists a finite function $\phi(b)$ such that $r_n(b)/r_n \to \phi(b)$ for all $b \neq 0$.

---

6) Of course, r_n may be defined as rank (A_{ni}) for any i ($1 \leq i \leq m$).

(15) (iii) The matrix $(1/r_n)\Psi_n$ converges to a positive definite matrix Ψ, where Ψ_n is a symmetric (m,m) matrix with typical element $(\Psi_n)_{ij} = \text{tr } A_{ni}\Omega_n A_{nj}\Omega_n$.

(iv) The matrices $(1/\sqrt{r_n})\, A_{ni}\Omega_n$ (i=1...m) converge to the null matrix.

Under the conditions (i) - (iv) we have, as $n\to\infty$

$$(16)\ \frac{1}{\sqrt{r_n}}\begin{pmatrix}\varepsilon_n' A_{n1}\varepsilon_n - \text{tr } A_{n1}\Omega_n\\ \vdots \\ \varepsilon_n' A_{nm}\varepsilon_n - \text{tr } A_{nm}\Omega_n\end{pmatrix} \longrightarrow N(0,2\Psi).$$

remark 4

Of course, each stochastic variable $(1/\sqrt{r_{ni}})(\varepsilon_n' A_{ni}\varepsilon_n - \text{tr } A_{ni}\Omega_n)$, i=1...m converges to a normal distribution if condition (10) holds for i=1...m. In theorem 3, however, we demand that the joint distribution of these variables converges to a normal distribution.

remark 5

Conditions (1),(iii) and (iv) are easy to verify. Condition (ii), however, is a nasty one. It says that the rank of any linear combination of the matrices A_{ni} goes to infinity with the same speed. In lemma 2 we give sufficient and verifiable conditions under which (ii) holds.

PROOF

Let $b = (b_1 ... b_m)'$ be some real m-vector. Define $A_n = \sum_{i=1}^{m} b_i A_{ni}$, and $r_n(b) = \text{rank } (A_n)$. Then, as $n\to\infty$

(a) $r_n(b) \to \infty$,

(b) $\text{tr}(A_n\Omega_n)^2/r_n(b) = \text{tr}(\sum_i b_i A_{ni}\Omega_n)(\sum_j b_j A_{nj}\Omega_n)/r_n(b)$

$= \sum_{ij} b_i b_j \text{tr}(A_{ni}\Omega_n A_{nj}\Omega_n)/r_n(b) = b'\Psi_n b/r_n(b)$

$= (r_n/r_n(b)).b'(1/r_n)\Psi_n b \to \phi^{-1}(b).b'\Psi b,$

(c) $A_n\Omega_n/\sqrt{r_n(b)} \to 0.$

Thus, A_n satisfies conditions (10) of theorem 2. This implies that

$$(1/\sqrt{r_n(b)})(\varepsilon_n' A_n \varepsilon_n - \mathrm{tr}\, A_n \Omega_n) \to N(0, 2\,\phi^{-1}(b).b'\Psi b),$$

that is

$$(1/\sqrt{r_n})\ (\varepsilon_n' A_n \varepsilon_n - \mathrm{tr}\, A_n \Omega_n) \to N(0, 2b'\Psi b),$$

or

$$(1/\sqrt{r_n})\ \sum_{i=1}^{m} b_i (\varepsilon_n' A_{ni} \varepsilon_n - \mathrm{tr}\, A_{ni} \Omega_n) \to N(0, 2b'\Psi b).$$

Since this holds for every $b \neq 0$, it follows that[7)]

$$\frac{1}{\sqrt{r_n}} \begin{pmatrix} \varepsilon_n' A_{n1} \varepsilon_n - \mathrm{tr}\, A_{n1} \Omega_n \\ \\ \varepsilon_n' A_{nm} \varepsilon_n - \mathrm{tr}\, A_{nm} \Omega_n \end{pmatrix} \to N(0, 2\Psi) \qquad \text{(Q.E.D.)}$$

As noted before, condition (ii) of theorem 3 is a troublesome one, since it implies an uncountable number of conditions. The following lemma shows that condition (ii) can be strengthened in such a way that is becomes verifiable.

<u>lemma 2</u>

Theorem 2 remains true when conditions (ii) and (iv) are replaced by

(ii*) There exist finite positive numbers α_i (i=1...m), such that $(1/r_n)$ rank $(A_{ni}) \to \alpha_i$, as $n \to \infty$. (Of course, $\alpha_1 = 1$)

(iv*) The eigenvalues of $A_{ni}\Omega_n$ (i=1...m) are uniformly <u>bounded</u>.[8)] (This is the case when the eigenvalues of Ω_n and A_{ni} (i=1...m) are uniformly bounded).

<u>PROOF</u>

It is clear that (iv*) implies (iv). Now, suppose that conditions (i), (ii*), (iii), and (iv*) hold. We shall prove that (ii) holds.
Again, $A_n = \sum_{i=1}^{m} b_i A_{ni}$ and $r_n(b) = \mathrm{rank}(A_n)$. We know that

$$\mathrm{var}(\varepsilon_n' A_n \varepsilon_n) = 2\ \mathrm{tr}(A_n \Omega_n)^2 = 2b'\Psi_n b\cdot$$

Therefore,

---

7) See the multivariate central limit theorem in Rao [6, p.128].

8) In this context, uniform means that the bound does not depend on n.

$$b'(1/r_n)\Psi_n b = (1/r_n(b))\ \mathrm{tr}(A_n\Omega_n)^2\ .\ r_n(b)/r_n\ .$$

Now, $b'(1/r_n)\Psi_n b$ converges to a finite limit unequal to zero (condition (iii)). Further, $r_n(b) \leq \sum_{i=1}^{m} \mathrm{rank}(A_{ni})$[9]. It then follows from condition (ii*) that $r_n(b)/r_n$ is uniformly bounded. Therefore, if for all $b \neq 0$ there exists a positive number $M(b)$ such that

$$(17)\quad (1/r_n(b))\ \mathrm{tr}(A_n\Omega_n)^2 \leq M(b),$$

then $r_n(b)/r_n$ must converge to a finite positive limit.[10]
Since $r_n(b)$ equals the number of nonzero eigenvalues of $A_n\Omega_n$, we have

$$(1/r_n(b))\ \mathrm{tr}(A_n\Omega_n)^2 \leq \mu(A_n\Omega_n)^2,$$

where $\mu(B)$ denotes the spectral radius of B.[11] Now,

$$\begin{aligned}
\mu(A_n\Omega_n) &= \max_x \left|\frac{x'\Omega_n^{\frac{1}{2}}A_n\Omega_n^{\frac{1}{2}}x}{x'x}\right| \\
&= \max_x \left|\sum_{i=1}^{m} b_i \frac{x'\Omega_n^{\frac{1}{2}}A_{ni}\Omega_n^{\frac{1}{2}}x}{x'x}\right| \\
&\leq \max_x \sum_i |b_i| \left|\frac{x'\Omega_n^{\frac{1}{2}}A_{ni}\Omega_n^{\frac{1}{2}}x}{x'x}\right| \\
&\leq \sum_i |b_i| \max_x \left|\frac{x'\Omega_n^{\frac{1}{2}}A_{ni}\Omega_n^{\frac{1}{2}}x}{x'x}\right| \\
&= \sum_i |b_i|\ \mu(A_{ni}\Omega_n)\ .
\end{aligned}$$

---

9) For any two matrices A and B such that A+B is defined, rank (A+B) $\leq$ rank(A) + rank(B).

10) There cannot be two accumulation points. Suppose $r_n(b)/r_n$ has two accumulation points γ_1 and γ_2. Then $(1/r_n(b))\ \mathrm{tr}(A_n\Omega_n)^2$ also has two accumulation points. Their product then has four accumulation points which must be all equal. This implies $\gamma_1 = \gamma_2$.

11) The spectral radius $\mu(B)$ of a square matrix B is the greatest of the absolute values of its eigenvalues.

Since we have assumed that the eigenvalues of $A_{ni}\Omega_n$ are uniformly bounded, $\mu(A_{ni}\Omega_n)$ is uniformly bounded, and $\mu(A_n\Omega_n)$ is bounded by a function of b. Therefore, $\mu(A_n\Omega_n)^2$ is bounded by a function of b. This establishes (17). Thus, $r_n(b)/r_n$ converges to a finite positive limit. Finally we note that

$$\mu(A_{ni}\Omega_n) = \max_x \left| \frac{x'\Omega_n^{\frac{1}{2}}A_{ni}\Omega_n^{\frac{1}{2}}x}{x'\Omega_n x} \right| \frac{x'\Omega_n x}{x'x}$$

$$\leq \max_y \left| \frac{y'A_{ni}y}{y'y} \right| . \max_x \frac{x'\Omega_n x}{x'x} = \mu(A_{ni})\mu(\Omega_n).$$

Thus, if the eigenvalues of Ω_n and A_{ni} (i=1...m) are uniformly bounded, then the eigenvalues of $A_{ni}\Omega_n$ are uniformly bounded. (Q.E.D.)

6. *Asymptotic normality of the ML estimators*

Four preliminary lemmas are needed to prove the asymptotic normality of $\hat{\beta}$ and $\hat{\theta}$. First, the following assumption is made:

A4: The matrix $(1/n)H'\Omega_n^{-1}H$ converges to a positive definite matrix Q as $n \to \infty$.

<u>lemma 3</u>

The vector $(1/\sqrt{n})\partial\Lambda_n/\partial\beta$ is distributed $N(0,(1/n)H'\Omega_n^{-1}H)$. Further, if assumption A4 is satisfied,

$$(1/\sqrt{n})\partial\Lambda_n/\partial\beta \to N(0,Q), \text{ as } n \to \infty.$$

<u>proof</u>

The lemma follows from the fact that $\partial\Lambda_n/\partial\beta=H'\Omega_n^{-1}\varepsilon_n$ (See (9)). (Q.E.D.)

Let us now introduce some definitions

$$A_{ni} \equiv \partial\Omega_n^{-1}/\partial\theta_i \quad (i=1...m),$$

$$r_n = \text{rank } (A_{n1}),$$

$\Psi_{n\theta}$ is the symmetric (m,m) matrix defined earlier in (7), with typical element $(\Psi_{n\theta})_{ij} = \text{tr } A_{ni}\Omega_n A_{nj}\Omega_n$.

Some further assumptions will be needed:

A5: The ranks of the m matrices A_{ni}[12] all go to infinity with the same speed, that is

$r_n \to \infty$, as $n \to \infty$

and

$(1/r_n)\text{rank}(A_{ni}) \to \alpha_i$ (i=1...m), as $n \to \infty$,

where the α_i are finite positive numbers, $\alpha_1=1$.

A6: The eigenvalues of $A_{ni}\Omega_n$ (i=1...m) are uniformly bounded.[13]

A7: The matrix $(1/\sqrt{r_n})\Psi_{n\theta}$ converges to a positive definite matrix Ψ_θ as $n \to \infty$.

<u>lemma 4</u>

Under the assumptions (A5) - (A7)[14],

$$(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta \to N(0,\tfrac{1}{2}\Psi_\theta), \text{ as } n \to \infty.$$

<u>proof</u>

From (9) we know that

$$(18)\quad \partial\Lambda_n/\partial\theta = \tfrac{1}{2}\left(\frac{\partial \text{vec}\Omega_n^{-1}}{\partial\theta}\right) \text{vec}(\Omega_n - \varepsilon_n\varepsilon_n') = -\tfrac{1}{2}\begin{bmatrix} \varepsilon_n' A_{n1}\varepsilon_n - \text{tr } A_{n1}\Omega_n \\ \vdots \\ \varepsilon_n' A_{nm}\varepsilon_n - \text{tr } A_{nm}\Omega_n \end{bmatrix}.$$

---

12) Note that rank (A_{ni}) = rank $(\partial\Omega^{-1}/\partial\theta_i)$ = rank $(\partial\Omega/\partial\theta_i)$, since $(\partial\Omega^{-1}/\partial\theta_i)\Omega = -\Omega^{-1}(\partial\Omega/\partial\theta_i)$.

13) Sufficient for the eigenvalues of $A_{ni}\Omega_n$ (i=1...m) to be uniformly bounded is that the eigenvalues of Ω_n and A_{ni} (i=1...m) are uniformly bounded (see lemma 2).

14) It should be noted that assumptions A5 and A6 are stronger than necessary, and may be replaced by conditions (15i),(15ii) and (15iv). The reason why we prefer A5 and A6 is that they allow verification, whereas condition (15ii) usually doesn't.

Applying theorem 3, we find the desired result. (Q.E.D.)

Thus, in lemmas 3 and 4, we have proved that $(1/\sqrt{n})\partial\Lambda_n/\partial\beta$ and $(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta$ are asymptotically normal. We shall now prove the asymptotic normality of the joint distribution of these vectors.

lemma 5

Under the assumptions (A4) - (A7),

$$\begin{pmatrix}(1/\sqrt{n})\partial\Lambda_n/\partial\beta \\ (1/\sqrt{r_n})\partial\Lambda_n/\partial\theta\end{pmatrix} \to N\left[0, \begin{pmatrix} Q & 0 \\ 0 & \frac{1}{2}\Psi_\theta\end{pmatrix}\right] \quad \text{, as } n \to \infty.$$

proof

For any positive definite (n,n) matrix P, the following equalities hold:

$$(19) \int \ldots \int \exp(-\tfrac{1}{2}x'P^{-1}x)dx_1 \ldots dx_n = (2\pi)^{n/2}|P|^{\frac{1}{2}} \text{ ,}$$

and

$$(20) \int \ldots \int \exp(-\tfrac{1}{2}x'P^{-1}x+t'x)dx_1 \ldots dx_n = (2\pi)^{n/2}|P|^{\frac{1}{2}} \exp \tfrac{1}{2}t'Pt.$$

The first equality simply states that the multivariate normal density with zero mean integrates to unity. The second equality reflects the fact that, if $x \simeq N(0,P)$, the moment generating function of $t'x$ is $\exp \frac{1}{2}t'Pt$.

Let $M_{n,\theta}(t) = E \exp t'(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta$ be the moment generating function of $(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta$. Then,

$$M_{n,\theta}(t) = (2\pi)^{-n/2}|\Omega_n|^{-\frac{1}{2}} \int \ldots \int \exp(t'(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta - \tfrac{1}{2}\varepsilon_n'\Omega_n^{-1}\varepsilon_n)d\varepsilon_1 \ldots d\varepsilon_n.$$

Substituting for $\partial\Lambda_n/\partial\theta$ the expression in (18), we find

$$M_{n,\theta}(t)=(2\pi)^{-n/2}|\Omega_n|^{-\frac{1}{2}}\exp\left[\tfrac{1}{2}\sum_{j=1}^{m} t_j(1/\sqrt{r_n})\mathrm{tr}(A_{nj}\Omega_n)\right] \cdot$$

$$\cdot \int \ldots \int \exp\left[-\tfrac{1}{2}\varepsilon_n'\Omega_n^{-\frac{1}{2}}\{I_n+\sum_j t_j(1/\sqrt{r_n})\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}}\}\Omega_n^{-\frac{1}{2}}\varepsilon_n\right]d\varepsilon_1 \ldots d\varepsilon_n \text{ .}$$

For n sufficiently large, the matrix

$$I_n + \sum_j t_j (1/\sqrt{r_n}) \Omega_n^{\frac{1}{2}} A_{nj} \Omega_n^{\frac{1}{2}}$$

will be positive definite (Assumption A6)[15]. Therefore, by (19), for large n

$$M_{n,\theta}(t) = |\Omega_n|^{-\frac{1}{2}} \left| \Omega_n^{-\frac{1}{2}} (I_n + \sum_j t_j (1/\sqrt{r_n}) \Omega_n^{\frac{1}{2}} A_{nj} \Omega_n^{\frac{1}{2}}) \Omega_n^{-\frac{1}{2}} \right|^{-\frac{1}{2}} \bullet$$

$$\bullet \exp\left(\tfrac{1}{2} \sum_j t_j (1/\sqrt{r_n}) \operatorname{tr}(A_{nj} \Omega_n)\right)$$

$$= |I_n + \sum_j t_j (1/\sqrt{r_n}) \Omega_n^{\frac{1}{2}} A_{nj} \Omega_n^{\frac{1}{2}}|^{-\frac{1}{2}} \,.\, \exp\left(\tfrac{1}{2} \sum_j t_j (1/\sqrt{r_n}) \operatorname{tr}(A_{nj} \Omega_n)\right).$$

Let $\phi_{n,\theta}(t)$ be the characteristic function of $(1/\sqrt{r_n}) \partial \Lambda_n / \partial \theta$, and $i = \sqrt{-1}$, then

$$\lim_{n\to\infty} \phi_{n,\theta}(t) = \lim_{n\to\infty} \exp\left(\tfrac{1}{2} i \sum_j t_j (1/\sqrt{r_n}) \operatorname{tr}(A_{nj} \Omega_n)\right),$$

since

$$\lim_{n\to\infty} |I_n + i \sum_j t_j (1/\sqrt{r_n}) \Omega_n^{\frac{1}{2}} A_{nj} \Omega_n^{\frac{1}{2}}| = 1 \ .$$

Also, by lemma 4, we know that $(1/\sqrt{r_n}) \partial \Lambda_n / \partial \theta$ is asymptotically distributed as $N(0, \frac{1}{2}\Psi_\theta)$. Hence,

$$\lim_{n\to\infty} \phi_{n,\theta}(t) = \exp -\tfrac{1}{4} t' \Psi_\theta t,$$

and thus

$$(21) \quad \lim_{n\to\infty} \exp \tfrac{1}{2} i \sum_j t_j (1/\sqrt{r_n}) \operatorname{tr}(A_{nj} \Omega_n) = \exp -\tfrac{1}{4} t' \Psi_\theta t.$$

Let us now consider the moment generating function $M_n(s,t)$ of $[(1/\sqrt{n}) \partial \Lambda_n / \partial \beta', (1/\sqrt{r_n}) \partial \Lambda_n / \partial \theta']'$. In the same way as before, we can write

---

15) In fact, the moment generating function (instead of the characteristic function) was used to ensure the positive definiteness of this matrix.

$$M_n(s,t) = E \exp\left[s'(1/\sqrt{n})\partial\Lambda_n/\partial\beta + t'(1/\sqrt{r_n})\partial\Lambda_n/\partial\theta\right]$$

$$= (2\pi)^{-n/2}|\Omega_n|^{-\frac{1}{2}} \exp\left[\tfrac{1}{2}\sum_{j=1}^{m} t_j(1/\sqrt{r_n})\mathrm{tr}(A_{nj}\Omega_n)\right] \cdot$$

$$\cdot \int\ldots\int \exp\left[s'(1/\sqrt{n})H'\Omega_n^{-1}\varepsilon_n - \tfrac{1}{2}\varepsilon_n'\Omega_n^{-\frac{1}{2}}\{I_n + \sum_j t_j(1/\sqrt{r_n})\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}}\}\Omega_n^{-\frac{1}{2}}\varepsilon_n\right]$$

$$d\varepsilon_1\ldots d\varepsilon_n .$$

By (20), we may evaluate the integral for large n. This gives for large n

$$M_n(s,t) = |\Omega_n|^{-\frac{1}{2}}|\Omega_n^{-\frac{1}{2}}(I_n + \sum_j t_j(1/\sqrt{r_n})\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}})\Omega_n^{-\frac{1}{2}}|^{-\frac{1}{2}} \cdot$$

$$\cdot \exp\left[\tfrac{1}{2}\sum_j t_j(1/\sqrt{r_n})\mathrm{tr}\,(A_{nj}\Omega_n)\right]$$

$$\cdot \exp\left[\tfrac{1}{2}s'(1/n)H'\Omega_n^{-1}\Omega_n^{\frac{1}{2}}\{I_n + \sum_j t_j(1/\sqrt{r_n})\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}}\}^{-1}\Omega_n^{\frac{1}{2}}\Omega_n^{-1}Hs\right] .$$

Let $\phi_n(s,t)$ be the characteristic function corresponding to $M_n(s,t)$. Then,

$$\lim_{n\to\infty} \phi_n(s,t) = \lim_{n\to\infty} \exp\left[\tfrac{1}{2}i\sum_j t_j(1/\sqrt{r_n})\mathrm{tr}\,(A_{nj}\Omega_n)\right] . \exp\left[-\tfrac{1}{2}s'(1/n)H'\Omega_n^{-1}Hs\right]$$

$$= \exp(-\tfrac{1}{4}t'\Psi_\theta t) . \exp(-\tfrac{1}{2}s'Qs),$$

by virtue of (21) and (A4). Thus,

$$\lim_{n\to\infty} \phi_n(s,t) = \exp -\tfrac{1}{2}(t'(\tfrac{1}{2}\Psi_\theta)t + s'Qs).$$

This implies that, as $n \to \infty$

$$\begin{pmatrix}(1/\sqrt{n})\partial\Lambda_n/\partial\beta \\ (1/\sqrt{r_n})\partial\Lambda_n/\partial\theta\end{pmatrix} \to N\left(0, \begin{pmatrix}Q & 0 \\ 0 & \tfrac{1}{2}\Psi_\theta\end{pmatrix}\right) . \qquad \text{(Q.E.D.)}$$

Let us now make two final assumptions:

A8: $(1/r_n^2)\mathrm{tr}\left(\frac{\partial^2\Omega_n^{-1}}{\partial\theta_i\partial\theta_j}\Omega_n\right)^2 \to 0$ (i,j=1...m), as $n \to \infty$.

A9: $(1/n^2)\left(\frac{\partial^2 g(\beta)}{\partial\beta_i\partial\beta_j}\right)' \Omega_n^{-1}\left(\frac{\partial^2 g(\beta)}{\partial\beta_i\partial\beta_j}\right) \to 0$ (i,j=1...k), as $n \to \infty$.

remark 6

Again, assumptions A8 and A9 are stronger than necessary. As will be clear from lemma 6, it is sufficient to assume that, as $n\to\infty$

$$(1/r_n)\left[\varepsilon_n' \frac{\partial^2 \Omega_n^{-1}}{\partial\theta_i \partial\theta_j}\varepsilon_n - \mathrm{tr}\left(\frac{\partial^2 \Omega_n^{-1}}{\partial\theta_i \partial\theta_j}\Omega_n\right)\right] \to 0 \quad \text{in probability } (i,j=1\ldots m),$$

and

$$(1/n)\left(\frac{\partial^2 g(\beta)}{\partial\beta_i \partial\beta_j}\right)' \Omega_n^{-1}\varepsilon_n \to 0 \quad \text{in probability } (i,j=1\ldots k)\ .$$

Assumptions A8 and A9 imply even convergence in quadratic mean. However, A8 and A9 allow straightforward verification, which we find important.

remark 7

It should be noted that A9 is the only assumption which arises typically in the nonlinear case. The linear regression model $y = X\beta + \varepsilon$ implies $\partial^2 g(\beta)/\partial\beta_i\partial\beta_j=0$, so that A9 is trivially true. As to the verification of A9, it is easy to see that, if the eigenvalues of Ω_n are uniformly bounded away from zero[16], sufficient for A9 is that

$$(1/n^2)\ (\partial^2 g(\beta)/\partial\beta_i\ \partial\beta_j)'(\partial^2 g(\beta)/\partial\beta_i\partial\beta_j) \to 0 \quad (i,j=1\ldots k),\ \text{as } n\to\infty.$$

lemma 6

Let B_n be the Hessian matrix of the loglikelihood (4) divided by appropriate factors n and r_n:

$$B_n = \begin{bmatrix} (1/n)\partial^2\Lambda_n/\partial\beta\partial\beta' & (1/\sqrt{nr_n})\partial^2\Lambda_n/\partial\theta\partial\beta' \\ (1/\sqrt{nr_n})\partial^2\Lambda_n/\partial\beta\partial\theta' & (1/r_n)\partial^2\Lambda_n/\partial\theta\partial\theta' \end{bmatrix}.$$

Then, under assumptions A4, A6, A8 and A9,

$B_n - EB_n \to 0$ in probability, as $n \to \infty$.

---

16) This means the following: There exists a $\delta > 0$ such that for all n, $\lambda(\Omega_n) > \delta$, where $\lambda(.)$ stands for any eigenvalue.

proof

From (6) we see that

$$EB_n = -\begin{bmatrix} (1/n)H'\Omega_n^{-1}H & 0 \\ 0 & (1/2r_n)\Psi_{n\theta} \end{bmatrix} .$$

Further, from (9),

$$\partial\Lambda_n/\partial\beta_i = (\partial g(\beta)/\partial\beta_i)'\Omega_n^{-1}\varepsilon_n \quad \text{and} \quad \partial\Lambda_n/\partial\theta_j = -\tfrac{1}{2}(\varepsilon_n' A_{nj}\varepsilon_n - \operatorname{tr} A_{nj}\Omega_n).$$

Hence,

$$\partial^2\Lambda_n/\partial\beta_i\partial\beta_j = (\partial^2 g(\beta)/\partial\beta_i\partial\beta_j)'\Omega_n^{-1}\varepsilon_n - (\partial g(\beta)/\partial\beta_j)'\Omega_n^{-1}(\partial g(\beta)/\partial\beta_i),$$

$$\partial^2\Lambda_n/\partial\theta_j\partial\beta_i = (\partial g(\beta)/\partial\beta_i)'A_{nj}\varepsilon_n ,$$

and

$$\partial^2\Lambda_n/\partial\theta_i\partial\theta_j = -\tfrac{1}{2}\left[\varepsilon_n'(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\varepsilon_n - \operatorname{tr}(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\Omega_n\right] + \tfrac{1}{2}\operatorname{tr}(A_{ni}\partial\Omega_n/\partial\theta_j)$$

$$= -\tfrac{1}{2}\left[\varepsilon_n'(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\varepsilon_n - \operatorname{tr}(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\Omega_n\right] - \tfrac{1}{2}\operatorname{tr}A_{ni}\Omega_n A_{nj}\Omega_n .$$

Thus, we have to show that, as $n\to\infty$

(a) $(1/n)\,(\partial^2 g(\beta)/\partial\beta_i\partial\beta_j)'\Omega_n^{-1}\varepsilon_n \to 0$ in probability,

(b) $(1/\sqrt{nr_n})\,(\partial g(\beta)/\partial\beta_i)'A_{nj}\varepsilon_n \to 0$ in probability,

and

(c) $(1/r_n)\,\left[\varepsilon_n'(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\varepsilon_n - \operatorname{tr}(\partial^2\Omega_n^{-1}/\partial\theta_i\partial\theta_j)\Omega_n\right] \to 0$ in probability.

It is sufficient, though not necessary, to show that the variances of (a), (b), and (c) vanish, as $n\to\infty$. Now, assumptions A8 and A9 say that the variances of (c) and (a) vanish, as $n\to\infty$. As to (b),

$$\begin{aligned} &\operatorname{var}\left[(1/\sqrt{nr_n})(\partial g(\beta)/\partial\beta_i)' A_{nj}\varepsilon_n\right] \\ &= (1/nr_n)(\partial g(\beta)/\partial\beta_i)' A_{nj}\Omega_n A_{nj}(\partial g(\beta)/\partial\beta_i) \\ &= (1/nr_n)(\Omega_n^{-\frac{1}{2}}\partial g(\beta)/\partial\beta_i)'(\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}})^2(\Omega_n^{-\frac{1}{2}}\partial g(\beta)/\partial\beta_i) \\ &\leqq (1/nr_n)\mu(\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}})^2 \,.\, (\partial g(\beta)/\partial\beta_i)'\Omega_n^{-1}(\partial g(\beta)/\partial\beta_i) \ , \end{aligned}$$

where, as before, $\mu(.)$ denotes the spectral radius. From A6 it follows that $\mu(\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}})^2$ is uniformly bounded. Hence, $(1/r_n)\mu(\Omega_n^{\frac{1}{2}}A_{nj}\Omega_n^{\frac{1}{2}})^2 \to 0$, as $n\to\infty$. Further, from A4, $(1/n)(\partial g(\beta)/\partial\beta_i)'\Omega_n^{-1}(\partial g(\beta)/\partial\beta_i)$ is uniformly bounded. This establishes (b). (Q.E.D.)

We are now in a position to consider the asymptotic distribution of the ML estimators $\hat\beta$ and $\hat\theta$. From the foregoing discussion it follows that all conditions for the traditional proof of asymptotic normality are fulfilled[17]: $\hat\beta$ and $\hat\theta$ are consistent, Λ_n is regular, and, as $n\to\infty$

$$B_n - EB_n \to 0 \qquad \text{in probability}$$

and

$$\begin{pmatrix}(1/\sqrt{n})\partial\Lambda_n/\partial\beta \\ (1/\sqrt{r_n})\partial\Lambda_n/\partial\theta\end{pmatrix} \to N\left(0, \begin{pmatrix}Q & 0 \\ 0 & \frac{1}{2}\Psi_\theta\end{pmatrix}\right) .$$

Hence, we may state the following result

THEOREM 4

Under the assumptions A4 - A9, the ML estimators $\hat\beta$ and $\hat\theta$ are asymptotically normal, that is, as $n\to\infty$

$$\begin{pmatrix}\sqrt{n}\ (\hat\beta-\beta) \\ \sqrt{r_n}\ (\hat\theta-\theta)\end{pmatrix} \to N\left(0, \begin{pmatrix}Q^{-1} & 0 \\ 0 & 2\Psi_\theta^{-1}\end{pmatrix}\right) .$$

---

17) See e.g. Zacks [9, pp. 246-247].

7. *Discussion and conclusion*

In this paper we have sought to establish verifiable conditions under which the ML estimators $\hat{\beta}$ and $\hat{\theta}$ of the nonlinear regression model (1) are strongly consistent and asymptotically normal. The first three conditions (A1) - (A3) describe the model. Of the remaining six assumptions, A5 (the rank condition) deserves some special attention. In our opinion A5 is necessary, since without it, we may not have an infinite sum as $n \to \infty$ (see theorem 2). However, no study is known to us where the rank condition is explicitly formulated. All conditions can be straightforwardly applied to the linear model $y = X\beta + \varepsilon$, by replacing H by X. Condition A9, being typical for the nonlinear case, then disappears.

Our conditions (A4) - (A8) should be confronted with the corresponding conditions (9) - (12) in [4]. These appear to be remarkably similar, although their derivations are very different. However, two differences are apparent: First, the rank condition does not appear in [4], and secondly, convergence should be uniform in that paper.

Of course, future research should establish conditions under which $\hat{\beta}$ and $\hat{\theta}$ are BAN estimators. Let us end with a conjecture: In theorem 2, the three conditions (10) are sufficient for the asymptotic normality of $\varepsilon' A\varepsilon$. The first condition is certainly necessary. Our conjecture is that the other two conditions are necessary too, i.e. (10) is sufficient <u>and</u> necessary for the asymptotic normality of $\varepsilon' A\varepsilon$.

REFERENCES

[1] Crowder, M.J.: "Maximum Likelihood Estimation for Dependent Observations," Journal of the Royal Statistical Society, series A, 139 (1976), 45 - 53.

[2] Gnedenko, B.V.: The Theory of Probability. Fourth Edition, New York: Chelsea Publishing Company, 1968.

[3] Hildreth, C.: "Asymptotic Distribution of Maximum Likelihood Estimators in a Linear Model with Autoregressive Disturbances," The Annals of Mathematical Statistics, 40 (1969), 583 - 594.

[4] Magnus, J.R.: "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," Journal of Econometrics, forthcoming.

[5] Magnus, J.R., and H. Neudecker: "The Commutation Matrix: Some Theorems and Applications," submitted for publication, 1977.

[6] Rao, C.R.: Linear Statistical Inference and its Applications. Second Edition. New York: John Wiley, 1973.

[7] Vickers, M.K.: "Optimal Asymptotic Properties of Maximum Likelihood Estimators for Parameters of Some Econometric Models," technical report, Cornell University, 1977.

[8] Weiss, L.: "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases,II," Journal of the American Statistical Association, 68 (1973), 428 - 430.

[9] Zacks, S.: The Theory of Statistical Inference. New York: John Wiley, 1971.