1989

Asymptotic Expansions and Curvature Measures in a Nonlinear Regression Model

Koichi Maekawa

19116

ISSN:0318-725X
ISBN:0-7714-1129-4

RESEARCH REPORT 8907

ASYMPTOTIC EXPANSIONS AND CURVATURE MEASURES IN A NONLINEAR REGRESSION MODEL

Koichi Maekawa

Department of Economics

University of Western Ontario & Hiroshima University

April 1989

*This paper was written while the author visited the Department of Economics, University of Western Ontario (UWO), during Fall 1988-Spring 1989. I am very grateful to the department for its hospitality and, especially for secretaries at the word processing centre. I thank seminar participants at UWO, Yale, and Hiroshima University for their helpful comments. (All correspondence is to be addressed to Department of Economics, Hiroshima University, Higashi Senda, Naka-ku, Hiroshima, Japan 730.)

Asymptotic Expansions and Curvature Measures in a Nonlinear Regression Model

by

Koichi Maekawa*

Department of Economics
University of Western Ontario and Hiroshima University

April 1989

Summary

This paper derives the asymptotic expansions of the distribution function of the maximum likelihood estimator (MLE) and the log likelihood ratio (LR) test in a nonlinear regression model. We investigate the effects of nonlinearity of our model on the asymptotic expansions by making use of two kinds of curvature measures: intrinsic curvature and parameter effect curvature defined by Bates and Watts (1980). It is shown that, after suitable transformation, the distribution function of the MLE up to $O(T^{-\frac{1}{2}})$ is shown to be related to only the parameter effect curvature. The intrinsic curvature appears only in a term of $O(T^{-1})$ in the distribution of LR. Furthermore, we briefly discuss the relationship between the intrinsic curvature and Efron's statistical curvature.

Keywords: Asymptotic expansion, Curvature measures, Log likelihood ratio test, Maximum likelihood estimator, Nonlinear regression.

*This paper was written while the author visited the Department of Economics, University of Western Ontario (UWO), during Fall 1988 – Spring 1989. I am very grateful to the department for its hospitality and, especially for secretaries at the word processing centre. I thank seminar participants at UWO, Yale, and Hiroshima University for their helpful comments. (All correspondence is to be addressed to Department of Economics, Hiroshima University, Higashi Senda, Naka–ku, Hiroshima, Japan 730.)

1. Introduction

We investigate the effects of nonlinearity in a nonlinear regression model on the asymptotic expansions of distribution functions of the maximum likelihood estimator (MLE) and the log likelihood ratio (LR) test by making use of Edgeworth type asymptotic expansions and relative curvature measures of nonlinearity defined by Bates and Watts (1980) (abbreviated as BW hereafter). We deal the following nonlinear regression model:

$$(1) \qquad y_t = f(x_t,\theta) + \varepsilon_t, \qquad t = 1,2,\ldots,T$$

where x_t is (k×1) vector of independent variables, θ is a (p×1) parameter vector, y_t the dependent variable, and f( ) the nonlinear response function. We make the following assumptions on ε_t and f( ) to develop asymptotic expansions.

<u>Assumption 1.</u> $\varepsilon_t \sim$ i.i.d. $N(0,\sigma^2)$, $t = 1,\ldots,T$.

<u>Assumption 2.</u> $f(x_t,\theta)$ is continuous and differentiable in θ up to the necessary order.

Regularity conditions for valid asymptotic expansions in nonlinear regression are discussed in detail by Ivanov (1976). However the proof of validity for asymptotic expansion is beyond the scope of this paper, so we will not discuss it here. Under the normality assumption on ε_t, the log likelihood function $\ell(\theta)$ for (1) is written as

$$(2) \qquad \ell(\theta) = -\frac{T}{2}\ln 2\pi - T\ln\sigma - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[y_t - f(x_t,\theta)]^2.$$

The MLE $\hat{\theta}$ is obtained as a solution of the first order condition

(3) $$\frac{\partial \ell(\theta)}{\partial \theta_i} = \frac{1}{\sigma^2} \sum_{t=1}^{T} [y_t - f(x_t,\theta)] f_t^{(i)} = 0,$$

where $f_t^{(i)} = \dfrac{\partial f(x_t,\theta)}{\partial \theta_i}$. Furthermore we define

$$f_t^{(i,j)} = \frac{\partial^2 f(x_t,\theta)}{\partial \theta_i \partial \theta_j},\ f_t^{(i,j,k)} = \frac{\partial^3 f(x_t,\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k},$$

and we assume

Assumption 3. The moments of derivatives such as

$$\text{(a) } \frac{1}{T}\sum_{t=1}^{T} f_t^{(i)2},\ \text{(b) } \frac{1}{T}\sum_{t=1}^{T} f_t^{(i)} f_t^{(j,t)},\ \text{(c) } \frac{1}{T}\sum_{t=1}^{T} f_t^{(i)} f_t^{(i,j,k)},\ \text{(d) } \frac{1}{T}\sum_{t=1}^{T} f_t^{(i)^2} f_t^{(j,k)},$$

and all sum of products associated 5 times differentiation such

$$\text{(e) } \frac{1}{T}\sum_t f_t^{(i,j,k,\ell,m)},\ \frac{1}{T}\sum_t f^{(j,k,\ell,m)},\ \frac{1}{T}\sum_t f_t^{(i,j)} f_t^{(k,\ell,m)}$$

have the order 0(1), where

$$f_t^{(j_1,\ldots j_s)} = \frac{\partial^s}{\partial \theta_{j_1} \ldots \partial \theta_{j_s}} f(x_t,\theta).$$ □

Note that these assumptions (a)~(e) are for the third order asymptotic expansions. If we are concerned with the second order expansions, (e) can be omitted. (e) assures that the remainder term in stochastic expansion of $\sqrt{T}(\hat{\theta}-\theta)$ up to $O(T^{-1})$ has $o(T^{-1})$.

As is easily seen, the MLE $\hat{\theta}$ is the least square estimator (LSE) in the sense that

$$(4) \qquad \sum_{t=1}^{T} [y_t - f(x_t,\theta)]^2$$

is minimized with respect to θ.

Although the MLE $\hat{\theta}$ has desirable asymptotic properties, it is biased estimator. Box (1971) derived the approximate bias of $\hat{\theta}$ in a more general case where there are more than two response functions but the normality of the error terms are assumed. On the other hand, BW proposed two kinds of measures of nonlinearity: intrinsic curvature and parameter effect curvature in order to investigate the effects of nonlinearity in nonlinear regression models. To have a clear physical and geometrical interpretation of these curvatures, the reader should refer to BW's original paper and/or an excellent exposition by Ratkowsky (1983, p.p. 7–8). BW showed that after a suitable reparameterization, Box's approximation of bias up to $O(T^{-\frac{1}{2}})$ in the LSE could be expressed only in terms of parameter effect curvature. As a result, they emphasized that "to the extent of Box's approximation, bias is strictly a property of the parameterization". This result naturally leads us to the following questions:

To the extent of Edgeworth type asymptotic expansion, are distributional properties of MLE and LR still related only to the parameter effect curvature? How could the intrinsic curvature affect the distributional properties of them? These questions are the motivations of this paper.

Section 2 of the paper derives the formal asymptotic expansion of distribution function of MLE for a multi–parameter case. Since it was derived by Maekawa and Lu Xianzi (1988), we only show the results omitting the detailed derivations. A supplement for details is available on request. In Section 3, we rewrite the asymptotic expansion formulas in terms of the curvature measures. Section 4 derives asymptotic expansions for MLE and LR, up to $O(T^{-1})$ in terms of curvatures. Section 5 has summary and conclusions.

2. Asymptotic expansion of MLE

Asymptotic expansions of MLE were studied by a considerable number of authors based on different techniques and in different contexts. Among others, the most relevant techniques to us are those in Takeuchi (1974) and Tankiguichi (1986), although the latter author dealt with Gaussian ARMA processes. We shall formally follow their derivation of asymptotic expansion.

Now we introduce the following definitions. Set

$$Z_1(\theta)' = [Z_{11}(\theta),\ldots,Z_{1p}(\theta)],$$

$$Z_{1i} = \frac{1}{\sqrt{T}}\frac{\partial \ell(\theta)}{\partial \theta_i},$$

$$Z_{2i} = \frac{1}{\sqrt{T}}\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right]\right] \equiv \{Z_{2k\ell}\}\ (p\times p),$$

$$I(\theta) = E[Z_1(\theta)Z_1(\theta)'] = -\frac{1}{T}E\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right]$$

$$J_{ijk}(\theta) = \mathrm{cov}[Z_{1i}(\theta), Z_{2jk}(\theta)]$$

$$\frac{1}{\sqrt{T}}K_{ik\ell}(\theta) = \mathrm{cum}\{Z_{1i}(\theta), Z_{1k}(\theta), Z_{1\ell}(\theta)\}$$

$$R_{ik\ell}(\theta) = \frac{1}{T}E\left[\frac{\partial^3 \ell(\theta)}{\partial\theta_i\partial\theta_k\partial\theta_\ell}\right]$$

where cum(x,y,z) denotes the third order cumulant of random variables x,y,z. Furthermore we need:

<u>Assumption 4</u>. The matrix $I(\theta)$ is positive definite for all θ.

Then it can be shown that under Assumptions 1–4, the stochastic expansion of the j^{th} element of standardized MLE $u_j = \sqrt{T}(\hat{\theta}_j - \theta_j)$ is given by

Theorem 1

$$(4)\qquad u_j = \sqrt{T}(\hat{\theta}_j - \theta_j)$$
$$= \sum_{k=1}^{p} I^{jk} Z_{1k} + \frac{1}{\sqrt{T}}\Big[\sum_{u,v,w=1}^{p} I^{ju} I^{vw} Z_{2uv} Z_{1w}$$
$$+ \frac{1}{2} \sum_{u,v,w,b,c=1}^{p} I^{ju} R_{uvw} I^{vb} I^{wc} Z_{1b} Z_{1c}\Big] + o_p\left[T^{-\frac{1}{2}}\right],$$

where I^{ij} is the $(i,j)^{th}$ element of $I(\theta)^{-1}$.

Proof. See Taniguchi and Maekawa (1988, Theorem 1).

Taking expectation of (3), it is not difficult to obtain:

Lemma 1

$$(5)\qquad E(u_j) = \frac{1}{\sqrt{T}} \sum_{t,w,v=1}^{p} I^{jt} I^{wv} (J_{vtw} + \tfrac{1}{2} R_{vtw}) + o\left[T^{-\frac{1}{2}}\right]$$
$$= \frac{1}{\sqrt{T}} c_j + o\left[\frac{1}{\sqrt{T}}\right],$$

$$(6)\qquad \mathrm{cum}\,(u_j, u_{j'}) = I^{jj'} + o\left[\frac{1}{\sqrt{T}}\right] = c_{jj'} + o\left[T^{-\frac{1}{2}}\right]$$

$$(7)\qquad \mathrm{cum}(u_j, u_{j'}, u_{j''}) = -\frac{1}{\sqrt{T}} \sum_{j_1 j_2 j_3=1}^{p} I^{jj_1} I^{j'j_2} I^{j''j_3} \times$$
$$\left\{2K_{j_1 j_2 j_3} + J_{j_1 j_2 j_3} + J_{j_2 j_3 j_1} + J_{j_3 j_1 j_2}\right\} + o\left[T^{-\frac{1}{2}}\right]$$
$$= \frac{1}{\sqrt{T}} c_{jj'j''} + o,\left[T^{-\frac{1}{2}}\right]$$

where the second equality signs in (5), (6), and (7) define c_j, $c_{jj'}$, and $c_{jj'j''}$.

Using c_j, $c_{jj'}$, $c_{jj'j''}$, the asymptotic expansion of distribution function of $u_1, u_2, \ldots, u_p$, up to $o\left[T^{-\frac{1}{2}}\right]$, is given by

Theorem 2

(8) $$p(u_1 \le x_1, u_2 \le x_2, \ldots, u_p \le x_p)$$
$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} g(u,\Omega)[1 + \frac{1}{\sqrt{T}} c_i H_i(u)$$
$$+ \frac{1}{6\sqrt{T}} \sum_{i,j,k=1} c_{ijk} H_{ijk}(u)]du + o\left[T^{-\frac{1}{2}}\right],$$

where $u' = (u_1, u_2, \ldots, u_p)$,

(9) $$g(u,\Omega) = (2\pi)^{-\frac{p}{2}} |\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}u'\Omega u\right], \; \Omega = \{c_{ij}\}$$

and H_i, H_{ijk} are Hermite polynomial defined by

(10) $$H_{j_1 j_2 \cdots j_s}(u) = \frac{(-1)^s}{g(u,\Omega)} \frac{\partial^s}{\partial u_{j_i} \cdots \partial u_{j_s}} g(u,\Omega).$$

Proof. The proof of (8) is entirely analogous to that of Theorem 3 in Taniguchi and Maekawa (1988).

If we calculate c_i, c_{ij}, c_{ijk} from our log likelihood function (2). We have the explicit formula in terms of derivatives of $f(x_t,\theta)$ with respect to θ_i. But since the result would be extremely complicated, we only show the explicit formulas of components of c_j, $c_{jj'}$, $c_{jj'j''}$, i.e., I_{ij}, $J_{ik\ell}$, $K_{ik\ell}$, and $R_{ik\ell}$.

Theorem 3

$$I_{ij}(\theta) = \frac{1}{T\sigma^2}\sum_{t=1}^{T} f_t^{(i)} f_t^{(j)}$$

$$J_{k\ell}(\theta) = \frac{1}{T\sigma^2}\sum_{t=1}^{T} f_t^{(i)} f_t^{(k,\ell)}$$

$$K_{ik\ell}(\theta) = 0$$

$$R_{ik\ell}(\theta) = -\frac{1}{T\sigma^2}\sum_{t=1}^{T}\left[f_t^{(i,\ell)} f_t^{(k)} + f_t^{(\ell)} f_t^{(k,\ell)} + f_t^{(i)} f_t^{(k,\ell)}\right].$$

Proof. By direct calculation, we immediately obtain

$$I_{ij}(\theta) = \frac{1}{T} E\left(\frac{\partial \ell(\theta)}{\partial \theta_i}\right)\left(\frac{\partial \ell(\theta)}{\partial \theta_i}\right)$$

$$= \frac{1}{T\sigma^2} E\left(\sum_{t=1}^{T} u_t f_t^{(i)}\right)\left(\sum_{t=1}^{T} u_t f_t^{(j)}\right)$$

$$= \frac{1}{T\sigma^2}\sum_{t=1}^{T} f_t^{(i)} f_t^{(j)}.$$

Similarly, $J_{ik\ell}(\theta)$, $K_{ik\ell}(\theta)$, and $R_{ik\ell}(\theta)$ are easily obtained. □

We note that if the model (1) is linear, then $f_t^{(i,j)} = 0$, $J_{ik\ell}(\theta) = 0$, $K_{ik\ell}(\theta) = 0$ and hence $c_j = 0$. This agrees with the fact that in the standard linear regression model LSE and MLE is unbiased. We also note that if p = 1 (θ is a single parameter), it is easy to see that the bias c_1 is

$$c_1 = -\frac{\sigma^2 \sum_{t=1}^{T} f_t^{(1)} f_t^{(1,1)}}{2[\Sigma f_t^{(1)^2}]^2}$$

which agrees with Box's (1971, p. 177) bias formula (2.25).

3. Representation in terms of Curvature Measure

This section derives the alternative representation of the asymptotic expansion of distribution function of MLE in Eq.(8) in terms of the curvature measures defined by BW.

To do so we first introduce their curvature measures. Define $(T \times p)$ matrix $V = [v_1, v_2, \ldots, v_p]$ whose ith column is v_i:

$$v_i = \left[\frac{\partial f(x_1,\theta)}{\partial \theta_i}, \frac{\partial f(x_2,\theta)}{\partial \theta_i}, \ldots, \frac{\partial f(x_T,\theta)}{\partial \theta_i}\right]' \Bigg|_{\theta = \theta_0} \qquad i = 1,2,\ldots p$$

and define the second partial derivative (p×p) matrix $\{v_{tij}\}$:

$$\{v_{tij}\} = \{\frac{\partial f(x_t,\theta)}{\partial \theta_i \partial \theta_j}\}\Bigg|_{\theta = \theta_0}.$$

and collect them into $p \times p$ matrix of $T-$ vectors V.. such as

$$V.. = \begin{bmatrix} \{v_{1ij}\} \\ \{v_{2ij}\} \\ \vdots \\ \{v_{Tij}\} \end{bmatrix}.$$

It can be shown that V. has the QR decomposition such that

$$V. = QR = Q\begin{bmatrix} \tilde{R} \\ \cdots \\ 0 \end{bmatrix} \begin{matrix} \}\ p \times p \\ \}\ (T - p) \times p \end{matrix}$$

where $\tilde{R}$ is upper triangular and Q is an orthogonal matrix, where the first p column of Q is v_i, i.e., $Q = [V \ldots N]$ (see BW (1980, eq.(2.10))

Let $\eta(\theta)$ be a $(T \times 1)$ vector $\eta(\theta) = (f(x_1,\theta), f(x_2,\theta), \ldots, f(x_T,\theta))'$ then Taylor expansion of $\eta(\theta)$ at $\theta = \theta_0$, up to the second order, is written as

(11) $\eta(\theta) \doteq \eta(\theta_0) + V.(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'V..(\theta - \theta_0)$.

Now following BW, we reparameterize θ to ϕ by

$$\theta = \tilde{R}^{-1}\phi = L\phi$$

and rotate the coordinator of the sample space by orthogonal matrix Q'. A geometrical interpretation of this rotation is given in BW article in p. 6. By these reparameterization and rotation, Taylor expansion (11) is written

$$Q'\eta(L\phi) \doteq Q'\eta(L\phi_0) + Q'V.L(\phi - \phi_0) + \frac{1}{2}Q'(\phi - \phi_0)L'V..L(\phi - \phi_0)$$

where $\phi_0 = L^{-1}\theta_0$.

The last term in rhs of the above formula can be written as a $(T \times 1)$ vector

$$\frac{1}{2}Q'(\phi - \phi_0)'L'V..L(\phi - \phi_0) = \frac{1}{2}\begin{bmatrix}(\phi - \phi_0)'A_1..(\phi - \phi_0)\\ (\phi - \phi_0)'A_2..(\phi - \phi_0)\\ \vdots \\ (\phi - \phi_0)'A_T..(\phi - \phi_0)\end{bmatrix} = \frac{1}{2}(\phi - \phi_0)'A..(\phi - \phi_0)$$

where A.. is T–vector of $p \times p$ matrix:

$$A.. = \begin{bmatrix}A_1..\\ A_2..\\ \vdots \\ A_T..\end{bmatrix} = \begin{bmatrix}A^T..\\ A^N..\end{bmatrix}$$

where $A_i..$ is called ith face of A.., and $A^T..$, $A^N..$ represent the first p faces and the last $(T - P)$ faces respectively. BW showed that the parameter effect curvature is defined by a linear combination of elements of $A^T..$ and the intrinsic curvature by a linear combination of those of $A^N..$.

Now let W. be a $(T \times p)$ matrix of first order derivative in the ϕ–co–ordinate whose ith column is w_i:

$$w_i = \left[\frac{\partial\eta_1(\phi)}{\partial\phi_i}, \frac{\partial\eta_2(\phi)}{\partial\phi_i}, \ldots, \frac{\partial\eta_T(\phi)}{\partial\phi_i}\right]'$$

and

$W. = (w_1, w_2, \dots, w_p)$.

Similarly we define a $(p \times p)$ matrix $\{w_{tjk}\}$ of second order derivative where

$$w_{tjk} = \frac{\partial^2 \eta_t(\phi)}{\partial \phi_j \partial \phi_k}.$$

We also define $(p \times p) \times T$ matrix

$$W = \begin{bmatrix} \{w_{1ij}\} \\ \{w_{2ij}\} \\ \vdots \\ \{w_{Tij}\} \end{bmatrix}.$$

It is easy to see that

$$W. = V.L \quad \text{and} \quad W.. = L'V..L\,,$$

where W. and W.. are counterparts of U. and U.. in BW (p.p. 6–7). Using these notations and the definition of $A^T_{..}$, it can be shown that

(11) a typical element of $A^T_{..} = \sum_{t=1}^{T} w_{ti} w_{tjk}$.

We are now ready to rewrite the asymptotic expansion (8) in terms of elements of A.. . It is not difficult to show the following results:

Lemma 2. Let $S = \sqrt{T}\, L^{-1}(\hat\theta - \theta) = \sqrt{T}\,(\hat\phi - \phi)$, then cumulants of s_i, s_j, s_k, which are element of S, are written as

$$E(s_i) = \sum_{t=1}^{T} \sum_{\alpha=1}^{p} w_{t\alpha} w_{ti\alpha} + o\left[T^{-\frac{1}{2}}\right] = \bar c_i + o\left[T^{-\frac{1}{2}}\right]$$

$$\text{cum}\,(s_i, s_j) = \bar c_{ij} + o\left[T^{-\frac{1}{2}}\right] \begin{cases} = 1 + o\left[T^{-\frac{1}{2}}\right] & \text{for } i = j \\ = 0 + o\left[T^{-\frac{1}{2}}\right] & \text{for } i \neq j \end{cases}$$

$$\mathrm{cum}\,(s_i,s_j,s_k) = \sum_{t=1}^{T} (w_{ti}w_{tjk} + w_{tj}w_{tik} + w_{tk}w_{tij}) + o\left[T^{-\frac{1}{2}}\right]$$

$$= \bar{c}_{ijk} + o\left[T^{-\frac{1}{2}}\right]$$

Proof. See Appendix.

We note that $\bar{c}_i$, $\bar{c}_{ijk}$ are expressed in terms of elements of $A^T_{\cdot\cdot}$, and $\bar{c}_{ij}$ is a constant. Therefore BW's transformation matrix L^{-1} makes the covariance matrix of $\sqrt{T}(\hat{\phi} - \phi)$ the asymptotically identity matrix. Applying the general formula of asymptotic expansion, we obtain:

Theorem 4. The asymptotic expansion of distribution function of $S = \sqrt{T}\,(\hat{\phi} - \phi)$ is given by

(12) $P(s_1 \le z_1, s_2 \le z_2, \dots, s_p \le z_p)$

$$= \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_p} g(s,\bar{\Omega})\,[1 + \frac{1}{\sqrt{T}} \sum_{t=1}^{P} \bar{c}_i H_i(s)$$

$$+ \frac{1}{6\sqrt{T}} \sum_{i,j,k} \bar{c}_{ijk}\, H_{ijk}(s)]ds_1 \dots ds_p + o(\frac{1}{\sqrt{T}})$$

where $s = (s_1,\dots,s_p)'$, $\bar{\Omega} = \{\bar{c}_{ij}\}$ and $g(s,\bar{\Omega})$, $H_i(s)$ and $H_{ijk}(S)$ are defined by (9) and (10).

We notice that, in reparameterizing θ to ϕ, the distribution function of $S = \sqrt{T}(\hat{\phi} - \phi)$ is expressed only in terms of elements of $A^T_{\cdot\cdot}$; components of parameter effect. This answers our question and allow us to extend BW's statement in saying that not only bias but also distribution up to $O\left[T^{-\frac{1}{2}}\right]$ is strictly a property of the parameterization. More precisely we can say that discrepancy from normality of the approximate distribution of $\sqrt{T}(\hat{\phi} - \phi)$ up to $O\left[T^{-\frac{1}{2}}\right]$ is strictly a property of the parameterization.

4. Third Order Asymptotic Expansions of MLE and LR for p = 1

4.1 Expansion for LR

We consider the problem of testing a null hypothesis:

$$H : \theta = \theta_0 + {}^{x}/\sqrt{T} \ (x > 0)$$

and against an alternative

$$K : \theta = \theta_0.$$

Likelihood ratio test statistic is defined by

$$LR = \ell(\theta_0) - \ell(\theta_1), \text{ where } \theta_1 = \theta_0 + \frac{x}{\sqrt{T}} .$$

Following Taniguchi (1986), we derive asymptotic expansion of LR. To do so we introduce the following notations:

$$Z_1 = \frac{1}{\sqrt{T}} \frac{\partial \ell(\theta)}{\partial \theta}$$

$$Z_2 = \frac{1}{\sqrt{T}} \left\{ \frac{\partial^2 \ell(\theta)}{\partial \theta^2} - E\left[\frac{\partial^2 \ell(\theta}{\partial \theta^2}\right] \right\}$$

$$Z_3 = \frac{1}{\sqrt{T}} \left\{ \frac{\partial^3 \ell(\theta)}{\partial \theta^3} - E\left[\frac{\partial^3 \ell(\theta)}{\partial \theta^3}\right] \right\}$$

$$E\left[Z_1^{\ 2}\right] = I + O(T^{-1})$$

$$E\left[Z_1 Z_2\right] = J + O(T^{-1})$$

$$E\left[Z_1^{\ 3}\right] = \frac{1}{\sqrt{T}} K + O\left[T^{-\frac{3}{2}}\right]$$

$$E\left[Z_1 Z_3\right] = P + O(T^{-1})$$

$$\mathrm{Var}\left[Z_2^2\right] = M + O(T^{-1})$$

$$E\left[Z_1^2 Z_2\right] = \frac{1}{\sqrt{T}} N + O\left[T^{-\frac{3}{2}}\right]$$

$$\mathrm{cum}(Z_1, Z_1, Z_1, Z_1) = \frac{1}{T} H + O\left[T^{-2}\right]$$

$$E\left\{\frac{1}{\sqrt{T}} \frac{\partial \ell(\theta)}{\partial \theta}\right\}^2 = I(\theta) + \Delta(\theta)/T + o(T^{-1}).$$

We consider a standardized LR as

$$\left\{LR - E_{\theta_1}(LR)\right\} / \left\{x/E\left[Z_1{}^2\right]\right\}$$

To obtain asymptotic expansion, up to $O(T^{-1})$, we have only to consider an approximation

$$Y_T = \left\{LR^* - E_{\theta_1}(LR^*)\right\}/(x\sqrt{I_T})$$

where $I_T = I(\theta) + \Delta(\theta)/\sqrt{T}$ and LR* is Taylor approximation:

$$LR^* = -\frac{x}{T}\left\{\frac{\partial \ell(\theta)}{\partial \theta}\right\}_{\theta_o} - \frac{x^2}{2T}\left\{\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right\}_{\theta_o} - \frac{x^3}{6T\sqrt{T}}\left\{\frac{\partial^3 \ell(\theta)}{6\theta^3}\right\}_{\theta_o}$$

$$-\frac{x^4}{24T^2}\left\{\frac{\partial^4 \ell(\theta)}{\partial \theta^4}\right\}_{\theta_o}.$$

It is shown that

$$\mathrm{Var}_{\theta_1}(Y_T) = 1 + \frac{1}{\sqrt{T}} b_1 + \frac{1}{T} b_2 + o(T^{-1}),$$

$$\mathrm{cum}_{\theta_1}(Y_T, Y_T, Y_T) = \frac{1}{\sqrt{T}} c_1 + \frac{1}{T} c_2 + o(T^{-1}),$$

$$\mathrm{cum}_{\theta_1}(Y_T, Y_T, Y_T, Y_T) = \frac{1}{T} d_1 + o(T^{-1}),$$

where

$$b_1 = \frac{x}{I}(J+K), \quad b_2 = \left[\frac{x^2}{12I}\right](4L + 3M + 18N + 6H),$$

$$c_1 = -\frac{K}{I^{3/2}}, \quad c_2 = -\left[\frac{x}{2I^{3/2}}\right](3N + 2H), \quad d_1 = \frac{H}{I^2}.$$

Taniguchi (1986, p. 15) derives the asymptotic expansions of Y_T under H and K, up to $O(T^{-1})$, under fairly general regularity conditions discussed in detail therein. The results are:

$$P_{\theta_1}(Y_T \leq a) = \Phi(a) - \phi(a)\left[\frac{1}{2}\left[\frac{b_1}{\sqrt{T}} + \frac{b_2}{T}\right] a + \left[\frac{c_1}{6\sqrt{T}} + \frac{c_2}{6T}\right](a^2-1) + \left[\frac{d_1}{24T} + \frac{b_2}{8T}\right](a^3 - 3a) + \frac{b_1C_1}{12T}(a^4 - 6a^2 + 3) + \frac{C_1{}^2}{72T}(a^5 - 10a^3 + 15a)\right] + o(T^{-1}) \tag{13}$$

and

$$P_{\theta_1}(Y_T \geq a) = \Phi(x') - \phi(x') - \phi(x')\left\{\frac{\beta_3}{6\sqrt{T}} + \frac{\beta_3}{6\sqrt{T}}(x'^2 - 1) + \left[\frac{\beta_2}{2T} + \frac{\beta_3^2}{72T}\right]x' + \left[\frac{\beta_4}{24T} + \frac{\beta_3^2}{36T}\right](x'^3 - 3x') + \frac{\beta_3^2}{72T}(x'^5 - 10x'^3 + 15x')\right\} + o(T^{-1}) \tag{14}$$

where $x' = x\sqrt{I_T}$ and $\beta_3 = -(3J + 2K)/I^{3/2}$,

$$\beta_4 = 3\beta_3^2 - (4P + 3M + 12N + 3H)/I^2,$$

$$\beta_2 = \frac{17}{36}\beta_3^2 - \frac{K^2}{18I^3} - \frac{12P + 9M + 36N + 8H}{12I^2}.$$

4.2 Relation to Curvature Measures

When p = 1 in our model (1) and log likelihood (2), it is easy to show that general formulas of I, J, K, M, N, H, P are reduced as follows:

$$I = \frac{1}{T\sigma^2} \Sigma f_t^{(1)^2}, \quad J = \frac{1}{T\sigma^2} \Sigma f_t^{(1)} f_t^{(1,1)}, \quad K = 0,$$

$$P = \frac{\sigma^2}{T} \frac{\Sigma}{T} f_t^{(1)} f_t^{(1,1,1)}, \quad M = \frac{1}{T\sigma^2} \Sigma f_t^{(1,1)2}$$

$$N = \frac{\sigma^2}{T\sqrt{T}} \Sigma f_t^{(1)^2} f_t^{(1,1)}, \qquad H = 0, \quad \Delta = 0.$$

On the other hand, when p = 1 our V. and V.. defined in Section 3 are

$$V. = \left[f_1^{(1)}, f_2^{(1)}, \ldots f_T^{(1)}\right]'$$

$$V.. = \left[f_1^{(1,1)}, f_2^{(1,1)}, \ldots, f_T^{(1,1)}\right]'.$$

Therefore we obtain

$$LL' = (V.'V.)^{-1} = 1/\sum_{t=1}^{T} f_t^{(1)^2}, \; L = \left[\sum_{t=1}^{T} f_t^{(1)^2}\right]^{-1/2},$$

(15) $$W. = V.L = \left[f_1^{(1)}, f_2^{(1)}, \ldots, f_T^{(1)}\right]' \left[\sum_{t=1}^{T} f_t^{(1,1)^2}\right]^{-1/2},$$

(16) $$W.. = L' \, V..L = \left[f_1^{(1,1)}, f_2^{(1,1)}, \ldots, f_T^{(1,1)}\right]' \left[\sum_{t=1}^{T} f_t^{(1)^2}\right]^{-1/2},$$

and

$$A... = (a_1^T, a_2^N, \ldots, a_T^N)'. \qquad (T\times 1)$$

We can represent A.. in terms of W. and W.. as follows:

Lemma 3 For p = 1, we have

(a) $$\left[a_1^T\right]^2 = \left[\frac{1}{\sum_{t=1}^{T} f_t^{(1)^2}} \times \sum_{t=1}^{T} f_t^{(1)} f_t^{(1,1)}\right]^2$$

and

(b) $$\sum_{t=2}^{T} \left[a_t^N\right]^2 = \frac{1}{\sum_{t=1}^{T} f_t^{(1)^2}} \times \sum_{t=1}^{T} f_t^{(1,1)^2}$$

Proof We can write as $A'..\ A.. = \left[a_1^T\right]^2 + \sum_{t=2}^{T} \left[a_t^N\right]^2$ i.e., sum of parameter effect and intrinsic curvatures, and for p = 1, BW's definition of A.. is reduced to

$A.. = Q'W.. = [W.|N]'\ W..$, where $Q'Q = E_T$, (T×T) identity matrix)

$$A..'A.. = W..'\ (W.'W. + N'N)\ W..$$

$$= W..'\ (W.'W. + E)W.. \quad (\because N'N = E_{T-1}, \text{(T−1)×(T−1) identity matrix})$$

$$= W..'\ (W.'W.)W.. + W..'\ W..$$

Substituting (15) and (16), we have (a) and (b). □

Using these relationships, we can write

$$I = \frac{1}{T\sigma^2} L^{-2},\ J = \frac{L^{-3}}{T\sigma^2} \left[a_1^T\right],\ K = 0,$$

$$H = 0, \qquad M = \frac{P^{-4}}{T\sigma^2} \left[\left[a_1{}^T\right]^2 + \sum_{t=2}^{T} \left[a_t^N\right]^2\right].$$

Note that M includes the intrinsic curvature $\sum_{t=2}^{T} \left[a_t^N\right]^2$, and N and P can not be represented by a_1^T and a_t^N.

Since asymptotic expansions of distributions of Y_T is expressed by I, J, K, ..., N as shown in eqs. (13) and (14), we observe:

Remark 1 The asymptotic expansions of distribution of LR test statistic (13) and (14) are affected by the intrinsic curvature through M in $O(T^{-1})$ terms such as b_2/T in (13) and β_2/T, β_4/T in (14).

Remark 2 (13) and (14) contain terms N and P which are neither parameter effect nor intrinsic curvature. P involve third order derivative $f_t^{(1,1,1)}$.

4.3 Expansion for MLE

Taniguchi (1986, p. 22, Theorem 3) also derived the third order asymptotic expansion for MLE $\hat{\theta}$ for p=1. It is represented by using I, J, ..., N defined above:

$$P\left\{\sqrt{T}\, I\,(\hat{\theta}-\theta) \le x\right\} = \Phi(x) - \phi(x)\left\{\left[\frac{\alpha_1}{\sqrt{T}} + \frac{\gamma_1}{6\sqrt{T}}(x^2-1)\right.\right.$$
$$+\frac{1}{2}\left[\frac{\rho_2}{T}+\frac{\alpha_1^2}{T}\right]x + \left[\frac{\delta_1}{24T}+\frac{\alpha_1\gamma_1}{6T}\right]\left[x^3-3x\right] \tag{17}$$
$$\left.+\frac{\gamma_1^2}{72T}(x^5-10x^3+15x)\right\} + o(T^{-1}),$$

where

$$\alpha_1 = \frac{J+K}{2I^{3/2}},$$

$$\rho_2 = \frac{7J^2 + 14JK + 5K^2}{2I^3} - \frac{L + 4N + H}{I^2} - \frac{\Delta}{I}$$

$$\gamma_1 = \frac{3J + 2K}{I^{3/2}},\ \delta = \frac{12(2J+K)(J+K)}{I^3} - \frac{4L + 12N + 3H}{I^2}.$$

Remark 3 Eq. (17) does not include M, hence the asymptotic expansion of distribution of MLE, $\sqrt{TI}(\hat{\theta}-\theta)$, up to $O(T^{-1})$ is not affected by intrinsic curvature.

Finally we check the relationship between BW's two curvatures and Effron's (1975) statistical curvature. The latter can be written by our notation as

$$\frac{1}{I(\theta)} E\left[Z_2(\theta) - \frac{J(\theta)}{I(\theta)} Z_1(\theta) \right] = \gamma(\theta)$$

It is shown (Taniguchi and Taniguchi (1987)) that

$$\gamma(\theta) = \left\{ M(\theta)\, I(\theta) - J(\theta)^2 \right\}^{1/2} / I(\theta)^{3/2}$$

where I, J, and M are defined in Section 4.1.

Since Effron's statistical curvature involves M which is essentially the sum of parameter effect and intrinsic curvature, it is not a property of parameterization.

5. Summary and Conclusions

Making use of the curvature measures introduced by Bates and Watts (1980), this paper has analyzed the effect of nonlinearity on MLE and LR test in a nonlinear regression model (1). First we developed the formal asymptotic expansion of distribution function of MLE $\hat{\theta}$ (p×1) up to $O\left[T^{-\frac{1}{2}}\right]$, and tried to rewrite the resulting distribution function in terms of the parameter effect and intrinsic curvature defined by BW. As a result we found that the distribution of MLE, $\sqrt{T}(\hat{\theta}-\theta)$, is associated only with parameter effect curvature. Although BW emphasized that "to the extent of Box's bias approximation, bias is strictly a property of the parameterization", our finding allows us to generalize BW statement as such that the discrepancy from normality of distribution of MLE are strictly properties of the parameterization up to $O(T^{-1/2})$.

Second, we explored the third order asymptotic expansion of distribution of MLE and LR test statistic for a single parameter case, i.e., θ (1×1). Observing the resulting expansion we found that the distribution of LR is affected by both curvatures in a term of $O(T^{-1})$, while the distribution of MLE is not affected by the intrinsic curvature up to $O(T^{-1})$. However, we noted that since the third order asymptotic expansions inevitably involves third order derivatives of nonlinear function f( ) in (1), contains other terms than curvatures which are a concept associated with second order derivative.

Finally we examined a relationship between BW's curvatures and Effron's statistical curvature and noticed that Effron's curvature is a mixture of parameter effect and intrinsic curvatures.

APPENDIX

Proof of Lemma 2. We only prove the third order cumulant. Let ℓ^{ij} be (i,j)th element of L^{-1}. Since $S = L^{-1}\sqrt{T}(\hat\theta - \theta) = L^{-1}U$, an element s_a in S is

$$s_a = \sum_{i=1}^{p} \ell^{ai} u_i. \qquad \text{Therefore we have}$$

$$\text{cum}(s_i, s_j, s_k) = \text{cum}\left(\sum_{h=1}^{p} \ell^{ih} u_h,\ \sum_{m=1}^{p} \ell^{jm} u_m,\ \sum_{n=1}^{p} \ell^{kn} u_n\right)$$

$$= \sum_h \sum_m \sum_n \ell^{ih} \ell^{jm} \ell^{kn}\ \text{cum}(u_h, u_m, u_n)$$

$$\text{(A1)} \quad = \sum_h \sum_m \sum_n \ell^{ih}\ell^{jm}\ell^{kn} \sum_{a,b,c} I^{ha} I^{mb} I^{nc} (J_{abc} + J_{bca} + J_{cab}) + o\left[T^{-\frac{1}{2}}\right]$$

where I^{rs}, J_{rst} are defined in Theorem 3. Since we have (see BW eq. (2.24))

$$V_{.}'V_{.} = R'Q'QR = R'R = \tilde{R}'\tilde{R} = (LL')^{-1},$$

and

$$I = \left[\frac{1}{T\sigma^2} \sum_{t=1}^{T} f_t^{(i)} f_t^{(j)}\right] = \left[\frac{1}{T\sigma^2} V_{.}'V_{.}\right],$$

we obtain $I^{-1} = T\sigma^2 LL'$, and $I^{mn} = T\sigma^2 \sum_\beta \ell^{m\beta}\ell^{\beta n}$. Substituting I^{mn} and J_{abc}, we have

$$\sum_{h,m,n} \ell^{ih}\ell^{jm}\ell^{kn} \sum_{a,b,c} I^{ha} I^{mb} I^{nc} J_{abc}$$

$$= \sum_{h,m,n} \ell^{ih}\ell^{jm}\ell^{kn} \sum_{a,b,c} \left(\sum_\alpha \ell_{h\alpha}\ell_{\alpha a}\right)\left(\sum_\beta \ell_{m\beta}\ell_{\beta b}\right)\left(\sum_\gamma \ell_{n\gamma}\ell_{\gamma c}\right)\left(\sum_t f_t^{(a)} f_t^{(b,c)}\right)$$

$$= \sum_t \left[\sum_{h,m,n} \ell^{ih}\ell^{jm}\ell^{kn} \sum_a \{\sum_\alpha \ell_{h\alpha}\ell_{\alpha a}\} f_t^{(a)} \sum_{b,c} \{\sum_\beta \ell_{m\beta}\ell_{\beta b}\}\{\sum_\gamma \ell_{n\gamma}\ell_{\gamma c}\} f_t^{(b,c)}\right.$$

$$= \sum_t \left[\sum_{h,m,n} \ell^{ih}\ell^{jm}\ell^{kn} \sum_\alpha \ell_{h\alpha} \sum_\alpha \ell_{\alpha a} f_t^{(a)} \sum_{\beta,\gamma} \ell_{m\beta}\ell_{n\gamma} \sum_{b,c} \ell_{\beta b} f_t^{(b,c)} \ell_{\gamma c}\right]$$

$$= \sum_t \left[\sum_{h,m,n} \ell^{ih} \ell^{jm} \ell^{kn} \sum_\alpha \ell_{h\alpha} w_{t\alpha} \sum_{\beta,\gamma} \ell_{m\beta} \ell_{n\gamma} w_{t\beta\gamma} \right]$$

$$= \sum_t \left[\sum_h \ell^{ih} \sum_\alpha \ell^{h\alpha} w_{t\alpha} \sum_{m,n} \sum_{\beta,\gamma} \ell^{jm} \ell^{kn} \ell_{m\beta} \ell_{n\gamma} w_{t\beta\gamma} \right]$$

$$= \sum_t \left[\sum_h \ell^{ih} \ell_{hi} w_{ti} \sum_m \ell^{jm} \ell_{mj} \sum_n \ell^{kn} \ell_{nk} w_{tjk} \right]$$

$$= \sum_t (w_{ti} w_{tjk}),$$

where we use the relation $L^{-1}L = E$ or

$$\sum_{i,j} \ell^{ij} \ell_{jk} = 0 \quad \text{if } i \neq j$$

$$= 1 \quad \text{if } i = j.$$

Similarly we can calculate other two terms in (A.1) associated with J_{bca} and J_{cab}. Finally we obtain $\text{cum}(s_i, s_j, s_k) = \sum_t (w_{ti} w_{tjk} + w_{tj} w_{tik} + w_{tk} w_{tij}) + O\left[T^{-\frac{1}{2}}\right]$. Calculations of $\text{cum}(s_i)$ and $\text{cum}(s_i, s_j)$ are entirely analogous.

REFERENCES

Bates, D.M. and Watts, D.G. (1980). Relative Curvature Measures of Nonlinearity, *J.R. Statis. Soc.* B, 42, 1–25.

Box, M.J. (1971). Bias in Nonlinear Estimation, *J.R. Statist. Soc.* B, 32, 171–201.

Effron, B. (1975). Defining the Curvature of Statistical Problem. *Ann. Statist.*, 3, 1189–1242.

Ivanov, A.V. (1976). An Asymptotic Expansion for the Distribution of the Least Squares Estimator of the Non–Linear Regression Parameter, *Theory of Probability and Its Applications*, XXI, 557–570.

Maekawa, K. and Lu Xianzi (1988). Second Order Properties of Nonlinear Least Squares Estimator (in Japanese), *Hiroshima Economic Review*, 12, 33–50.

Takeuchi, K. (1974) Asymptotic Theory of Statistical Inference (in Japanese) Kyoiku Shuppan Co. Tokyo.

Taniguchi, M. (1986). Third Order Asymptotic Properties of MLE for Gaussian ARMA Processes, *Jour. of Multivariate Analysis* 18, 1–31.

Taniguchi, M. and Taniguichi, R. (1987). Asymptotic Ancillarity in Time Series Analysis, *Technical Report*, No. 197, Statistical Research Group, Horishima Univ.

Taniguchi, M. and Maekawa, K. (1988). Asymptotic Expansions of Distributions of Statistics Related to the Spectral Density Matrix in Multivariate Time Series and Their Applications, to appear in *Econometric Theory*.

SUPPLEMENT

Proofs and derivations in this paper are largely dependant on the previous works, some of which are not easily accessible from outside of Japan because they are not published or written in Japanese. So I decided to write this supplement to give the details to interested readers and referees.

(1) Proof of Theorem 1

MLE $\hat{\theta}$ is a solution of the first order condition $\frac{\partial \ell(\theta)}{\partial \theta} = 0(p\times 1)$. If we develop this equation in Taylor expansion about $\theta = \hat{\theta}$ (p×1), we have

$$0 = \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}}$$
$$= \frac{\partial \ell(\theta)}{\partial \theta} + \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}(\hat{\theta}-\theta) + \frac{1}{2}F_{\ldots}\circ(\hat{\theta}-\theta)\circ(\hat{\theta}-\theta)$$

where $F_{\ldots}$ is a p×p×p array with $(i,j,k)^{th}$ element

$$\frac{\partial^3 \ell(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$$

and $F_{\ldots}\circ(\hat{\theta}-\theta)\circ(\hat{\theta}-\theta)$ is a (p×1) column vector with i^{th} element

$$\sum_{k,\ell} \frac{\partial^3 \ell(\theta)}{\partial \theta_i \partial \theta_k \partial \theta_\ell}\left[\hat{\theta}_k-\theta\right]\left[\hat{\theta}_k-\theta\right].$$

Therefore we have

$$\sqrt{T}(\hat\theta-\theta) \doteq \left[E\left[\frac{1}{T}\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right] + \frac{1}{T}\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'} - E\left[\frac{1}{T}\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right]\right]^{-1}\left[\frac{1}{\sqrt{T}}\frac{\partial \ell(\theta)}{\partial\theta}\right]$$
$$-\frac{1}{2\sqrt{T}}\left[\frac{1}{T}\frac{\partial^2(\theta)}{\partial\theta\partial\theta'}\right]^{-1}\frac{1}{T}F\ldots\circ\sqrt{T}(\hat\theta-\theta)\circ\sqrt{T}(\hat\theta-\theta).$$

Substituting

$$\frac{1}{T}E\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right] = -I(\theta) + O(T^{-1}),$$

we obtain

$$= [I - I(\theta)^{-1}\frac{Z_2(\theta)}{\sqrt{T}} + O(T^{-1})]^{-1} I(\theta)^{-1} Z_1(\theta)$$
$$+\frac{1}{2\sqrt{T}} I(\theta)^{-1}\frac{F\ldots}{T}\circ\sqrt{T}(\hat\theta-\theta)\circ\sqrt{T}(\hat\theta-\theta)$$
$$\doteq [I + \frac{I(\theta)^{-1}Z_2(\theta)}{\sqrt{T}} + O(T^{-1})] I(\theta)^{-1} Z_1(\theta)$$
$$+\frac{1}{2\sqrt{T}} I(\theta)^{-1} E\left[\frac{F\ldots}{T}\right]\circ I(\theta)^{-1}Z_1(\theta)\circ I(\theta)^{-1}Z_1(\theta).$$

Therefore we can write

$$\sqrt{T}(\hat\theta-\theta) = I(\theta)^{-1}Z_1(\theta) + \frac{1}{\sqrt{T}} I(\theta)^{-1} Z_2(\theta) I(\theta)^{-1} Z(\theta)$$
$$+\frac{1}{2\sqrt{T}} I(\theta)^{-1} E\left[\frac{\partial^3 \ell(\theta)}{T}\right]\circ I(\theta)^{-1}Z_1(\theta)\circ I(\theta)^{-1}Z_1(\theta) + o_p\left[T^{-\frac{1}{2}}\right].$$

By taking i^{th} element out, we obtain Eq.(4) in Theorem 1.

(2) <u>A general formula of Edgeworth type asymptotic expansion used in Theorem 2.</u>

Let $Y_T = (Y_1, Y_2, \ldots, Y_p)$ were Y_j is a random variable, and suppose they have cumulants such as

$$E(Y_j) = \frac{c_j}{\sqrt{T}} + o\left[T^{-\frac{1}{2}}\right],$$

$$\text{cum}(Y_j, Y_k) = c_{jk} + o\left[T^{-\frac{1}{2}}\right],$$

$$\text{cum}(Y_j, Y_k, Y_\ell) = \frac{c_{jk\ell}}{\sqrt{T}} + o\left[T^{-\frac{1}{2}}\right].$$

Then the asymptotic expansion of joint distribution of Y_j, up to $O\left[T^{-\frac{1}{2}}\right]$, is given by

$$P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_p \le y_p)$$
$$= \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_p} g(\underset{\sim}{y}, \Omega)[1 + \frac{1}{\sqrt{T}} \sum_{i=1}^{p} c_i H_i(\underset{\sim}{y})$$
$$+ \frac{1}{6\sqrt{T}} \sum_{i,j,k=1}^{p} c_{ijk} H_{ijk}(\underset{\sim}{y})] dy_1 \ldots dy_p$$

where H_i, H_{ijk} are Hermite polynomials:

$$H_{j_1 j_2 \ldots j_s}(\underset{\sim}{y}) = \frac{(-1)^s}{g(\underset{\sim}{y}, \Omega)} \frac{\partial^s}{\partial y_{j_i} \partial y_{j_2} \ldots \partial y_{j_s}} g(\underset{\sim}{y}, \Omega),$$

$$g(\underset{\sim}{y}, \Omega) = (2\pi)^{-P/2} |\Omega|^{\frac{1}{2}} \exp\left[-\frac{1}{2} \underset{\sim}{y}' \Omega \underset{\sim}{y}\right],$$

$$\underset{\sim}{y} = (y_1, \ldots, y_p)'$$

For a more general case, see Taniguchi (1986, Eq. (3.8)).

(3) Proof of Theorem 3

Derivatives of $\ell(\theta)$ wrt θ_i up to third order are given by

(3.1) $$\frac{\partial \ell(\theta)}{\partial \theta_i} = \frac{1}{\sigma^2} \sum_t (y_t - f(x_t,\theta)) f_i^{(i)} = \frac{1}{\sigma^2} \sum_t \varepsilon_t f_t^{(i)}$$

where $y_t - f(x_t,\theta) = \varepsilon_t$. Similarly

(3.2) $$\frac{\partial^2 \ell(\theta)}{\partial \theta_k \partial \theta_\ell} = -\frac{1}{\sigma^2} \sum_t f_t^{(\ell)} f_t^{(k)} + \frac{1}{\sigma^2} \sum_t \varepsilon_t f_t^{(k,\ell)}$$

(3.3) $$\frac{\partial^3 \ell(\theta)}{\partial \theta_i \partial \theta_k \partial \theta_\ell} = \frac{-1}{\sigma^2} \sum_t \left[f_t^{(i,\ell)} f_t^{(k)} + f_t^{(\ell)} f_t^{(k,i)} - f_t^{(i)} f_t^{(k,\ell)} + \varepsilon_t f_t^{(i,k,\ell)} \right]$$

As is defined in Section 2, we have

(3.4) $$z_{1i} = \frac{1}{\sqrt{T}} \frac{\partial \ell(\theta)}{\partial \theta_i},$$

(3.5) $$Z_{2k\ell} = \frac{1}{\sqrt{T}} \left[\frac{\partial^2 \ell(\theta)}{\partial \theta_k \partial \theta_\ell} - E\left[\frac{\partial^2 \ell(\theta)}{\partial \theta_k \partial \theta_\ell} \right] \right].$$

Substituting (3.1) and (3.2) into (3.4) and (3.5), we have

(3.6) $$E(z_{ii}) = \frac{1}{\sigma^2} \sum_t E(\varepsilon_t) f_t^{(i)} = 0,$$

(3.7) $$-E\left[\frac{\partial^2 \ell(\theta)}{\partial \theta_k \partial \theta_\ell} \right] = \frac{1}{T\sigma^2} \sum_t f_t^{(\ell)} f_t^{(k)}.$$

By definition

$$
\begin{aligned}
J_{ik\ell}(\theta) &= \mathrm{cov}(z_{1i}, z_{2k\ell}) \\
(3.8) \qquad &= E\{[Z_{1i}-E(Z_{1i})]Z_{2k\ell}\} \\
&= E(z_{1i}z_{1k}z_{1\ell})
\end{aligned}
$$

Substituting (3.4)~(3.7) into (3.8), and using $E\left[\varepsilon_t^2\right] = \sigma^2$, $E(\varepsilon_t,\varepsilon_s) = 0$ for $s \neq t$, we obtain

$$
\begin{aligned}
J_{ik\ell}(\theta) &= \frac{1}{T\sigma^4} E\left\{\left[\sum_t \varepsilon_t f_t^{(i)}\right] \sum_t \left[f_t^{(k)} f_t^{(\ell)} + u_t f_t^{(k,\ell)} + f_t^{(\ell)} f_t^{(k)}\right]\right\} \\
&= \frac{1}{T\sigma^2} \sum_t f_t^{(i)} f_t^{(k,\ell)}.
\end{aligned}
$$

Next we calculate $K_{ik\ell}(\theta)$ and $R_{ik\ell}(\theta)$:

$$
\begin{aligned}
K_{ik\ell}(\theta) &= \mathrm{cum}(z_{1i}, z_{1k}, z_{1\ell}) \\
&= E(z_{1i}, z_{1k}, z_{1\ell}) \\
&= \frac{1}{T\sqrt{T}\sigma^6} E\left\{\left[\sum_t \varepsilon_t f_t^{(i)}\right]\left[\sum_t \varepsilon_t f_t^{(k)}\right]\left[\sum_t \varepsilon_t f_t^{(\ell)}\right]\right\} \\
&= 0 \ (\because E(\varepsilon_s \varepsilon_t \varepsilon_u) = 0) \\
R_{ik\ell}(\theta) &= \frac{1}{T} E\left[\frac{\partial^3 \ell(\theta)}{\partial\theta_i \partial\theta_k \partial\theta_\ell}\right] \\
&= -\frac{1}{T\sigma^2} \sum_t \left[f_t^{(i,\ell)} f_t^{(k)} + f_t^{(\ell)} f_t^{(k,i)} + f_\tau^{(i)} f_t^{(k,\ell)}\right]
\end{aligned}
$$

(4) <u>Representation of A..</u>

For illustrative purpose, the representation of A.. is shown for p=2, and T=3. A.. is defined by

$$(4.1)\qquad \tfrac{1}{2}Q'(\phi-\phi_0)'\,W..(\phi-\phi_0)=\tfrac{1}{2}\begin{bmatrix}(\phi-\phi_0)'A_1...(\phi-\phi_0)\\(\phi-\phi_0)'A_2...(\phi-\phi_0)\\(\phi-\phi_0)'A_3...(\phi-\phi_0)\end{bmatrix},$$

where $W.. = L'V..L$ is defined in Section 3.

Set $(\phi-\phi_0) = a = (a_1,a_2)'$ and $Q = \{q_{ij}\}$, $i,j = 1,2,3$. Then we can write

$$(4.2)\qquad Q'(\phi-\phi_0)'W..(\phi-\phi_0)=Q'\begin{bmatrix}a'W_1..\,a\\a'W_2..\,a\\a'W_3..\,a\end{bmatrix}$$

$$=\begin{bmatrix}q_{11}(a'W_1..a)+q_{12}(a'W_2..a)+q_{13}(a'W_3..a)\\q_{21}(a'W_1..a)+q_{22}(a'W_2..a)+q_{23}(a'W_3..a)\\q_{31}(a'W_1..a)+q_{32}(a'W_2..a)+q_{33}(a'W_3..a)\end{bmatrix}$$

$$=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix},\text{ say.}$$

For instance we have

$$\begin{aligned}x_1 &= q_{11}\left[a_1^2w_{111}+a_1a_2w_{112}+a_2a_1w_{121}+a_2^2w_{122}\right]\\&\quad+q_{12}\left[a_1^2w_{211}+q_{12}w_{211}+a_2a_1w_{221}+a_2^2w_{222}\right]\\&\quad+q_{13}\left[a_1^2w_{311}+a_1a_2w_{312}+a_2a_1w_{321}+a_2^2w_{322}\right]\\&= a_1^2\left[q_{11}w_{111}+q_{12}w_{211}+q_{13}w_{311}\right]\\&\quad+a_1a_2\left[q_{11}w_{112}+q_{12}w_{212}+q_{13}w_{312}\right]\\&\quad+a_2a_1\left[q_{11}w_{121}+q_{12}w_{221}+q_{13}w_{312}\right]\end{aligned}$$

$$+a_2^2\left[q_{11}w_{122}+q_{12}w_{222}+q_{13}w_{322}\right]$$

Similarly we can calculate x_2 and x_3. Substituting those x_1, x_2, and x_3, we obtain

$$(4.2) \qquad = \begin{bmatrix} a'A_{1..}\, a \\ a'A_{2..}\, a \\ a'A_{3..}\, a \end{bmatrix}$$

where $A_{i..}$ is the i^{th} face of A..., i.e.,

$$A.. = \begin{bmatrix} A_{2..} \\ A_{2..} \\ A_{3..} \end{bmatrix}.$$

Form the definition of Q, the first p columns of Q are $w_1,...,w_p$ (see eq. (2.13) in BW), so we have

$$Q = \begin{bmatrix} w_{11} & w_{21} & q_{31} \\ w_{12} & w_{22} & q_{33} \\ w_{13} & w_{23} & q_{33} \end{bmatrix} \text{ for p=2 and T=3.}$$

Therefore we have

$$A_{i..} = \begin{bmatrix} \sum_t w_{it}w_{t11}, & \sum_t w_{it}w_{t12} \\ \sum_t w_{it}w_{t21}, & \sum_t w_{it}w_{t22} \end{bmatrix} \text{ for i = 1,2}$$

and

$$A_{3\cdot\cdot} = \begin{bmatrix} \sum_t w_{3t} w_{t11}, & \sum_t w_{3t} w_{t12} \\ \sum_t w_{3t} w_{t21}, & \sum_t w_{3t} w_{t22} \end{bmatrix}.$$