SECOND ORDER PROPERTIES OF LOCALLY STATIONARY PROCESSES

BY KENICHIRO TAMAKI

Waseda University

Abstract. In this paper we investigate an optimal property of the maximum likelihood estimator of Gaussian locally stationary processes by the second order approximation. In the case where the model is correctly specified, it is shown that appropriate modifications of the maximum likelihood estimator for Gaussian locally stationary processes is second order asymptotically efficient. We discuss second order robustness properties.

Keywords. Gaussian locally stationary process; maximum likelihood estimator; second order asymptotic efficiency

1. INTRODUCTION

There has been much discussion of the efficiency in estimation of stationary time series. Hosoya (1979), Akahira and Takeuchi (1981) and Taniguchi (1983) deal with higher order efficiencies for time series analysis. Taniguchi (1983) and Taniguchi and Kakizawa (2001) showed that appropriately modified maximum likelihood and quasi maximum likelihood estimators of Gaussian autoregressive moving average processes is second order asymptotically efficient in the sense of degree of concentration of the sampling distribution up to second order. This concept of efficiency was introduced by Akahira and Takeuchi (1981), and these results are reviewed in Section 2.

Although the analysis for stationary time series is well established, there are many cases where the stationary assumption seems to be restrictive. Because all the results above deal with stationary processes we are led to the problem of efficiently estimating parameters of non-stationary processes.
Dahlhaus (1996a, 1996b, 1996c, 2000) has introduced a class of locally stationary processes (non-stationary processes), and formulated in a rigorous asymptotic framework.

In this paper, we investigate the problems of efficiently estimating parameters of multivariate Gaussian locally stationary processes in the sense of Akahira and Takeuchi (1981). In Section 3, we discuss second order robustness properties.

2. SECOND ORDER EFFICIENCY

We consider the approach of Akahira and Takeuchi (1981) whose argument proceeds as follows. Let \( X_1, T, \ldots, X_{T,T} \) be a sequence of random variables forming a stochastic process, and possessing the probability measure \( P^T_\theta[\cdot] \), where \( \theta = (\theta_1, \ldots, \theta^p) \in \Theta \), a subset of \( \mathbb{R}^p \). We assume that \( \theta_2 = (\theta^2, \ldots, \theta^p) \) is a nuisance parameter (see, Section 1.2 and 4.4 in Akahira and Takeuchi, 1981). If an estimator \( \hat{\theta}_1 \) of \( \theta_1 \) satisfies the equation

\[
\lim_{T \to \infty} \sqrt{T} |P^T_\theta[\sqrt{T}(\hat{\theta}_1 - \theta_1)] \leq 0| - 1/2 = 0,
\]

then \( \hat{\theta}_1 \) is called a second order asymptotically median unbiased (second order AMU) estimator. For this \( \hat{\theta}_1 \), the asymptotic distribution functions

\[
F^+_{\theta_0}(x) + G^+_{\theta_0}(x)/\sqrt{T} \quad \text{and} \quad F^-_{\theta_0}(x) + G^-_{\theta_0}(x)/\sqrt{T}
\]

are defined to be the second order asymptotically distribution of \( \sqrt{T}(\hat{\theta}_1 - \theta_1) \) if

\[
\lim_{T \to \infty} \sqrt{T} |P^T_\theta[\sqrt{T}(\hat{\theta}_1 - \theta_1)] \leq x_1| - F^+_{\theta_0}(x_1) - G^+_{\theta_0}(x_1)/\sqrt{T} = 0
\]

for all \( x_1 \geq 0 \),

\[
\lim_{T \to \infty} \sqrt{T} |P^T_\theta[\sqrt{T}(\hat{\theta}_1 - \theta_1)] \leq x_1| - F^-_{\theta_0}(x_1) - G^-_{\theta_0}(x_1)/\sqrt{T} = 0
\]

for all \( x_1 < 0 \).

For \( \theta_0 = (\theta^1_0, \ldots, \theta^p_0) \in \Theta \), consider the problem of testing hypothesis \( H : \theta_1 = \theta^1_0 + x^1/\sqrt{T} \) (\( x^1 > 0 \)) against alternative \( A : \theta = \theta_0 \). We define \( \beta^+_{\theta_0}(x^1) \) and \( \gamma^+_{\theta_0}(x^1) \) as follows:

\[
\sup_{\{\phi_T \in \Phi_{1/2}\}} \lim_{T \to \infty} \sup \sqrt{T} \{E^T_{\theta_0}[\phi_T] - \beta^+_{\theta_0}(x^1) - \gamma^+_{\theta_0}(x^1)/\sqrt{T}\} = 0,
\]
where

$$\Phi_{1/2} = \{ \phi_T : E_{\theta_0}^{T} [\phi_T] = 1/2 + o(1/\sqrt{T}), 0 \leq \phi_T \leq 1 \}. \quad (5)$$

Then we have for $x^1 \geq 0$

$$F_{\theta_0}^+(x^1) \leq \beta_{\theta_0}^+(x^1) \quad \text{and} \quad G_{\theta_0}^+(x^1) \leq \gamma_{\theta_0}^+(x^1). \quad (6)$$

Also consider the problem of testing hypothesis $H : \theta^1 = \theta_0^1 + x^1/\sqrt{T} \ (x^1 < 0)$ against alternative $A : \theta = \theta_0$. We define $\beta_{\theta_0}^-(x^1)$ and $\gamma_{\theta_0}^-(x^1)$ as follows:

$$\inf_{\{\phi_T \in \Phi_{1/2}\}} \lim_{T \to \infty} \sqrt{T} \{ E_{\theta_0}^{T} [\phi_T] - \beta_{\theta_0}^-(x^1) - \gamma_{\theta_0}^-(x^1) / \sqrt{T} \} = 0. \quad (7)$$

In the same way as for the case $x^1 > 0$, we have for each $x^1 < 0$

$$F_{\theta_0}^-(x^1) \geq \beta_{\theta_0}^-(x^1) \quad \text{and} \quad G_{\theta_0}^-(x^1) \geq \gamma_{\theta_0}^-(x^1). \quad (8)$$

Thus we make the following definition.

**DEFINITION 1** (Akahira and Takeuchi, 1981). A second order AMU estimator $\hat{\theta}^1$ is called second order asymptotically efficient if for each $\theta \in \Theta$

$$P_{\theta}^{T} [\sqrt{T} (\hat{\theta}^1 - \theta^1) \leq x^1] = \begin{cases} \beta_{\theta_0}^+(x^1) + \gamma_{\theta_0}^+(x^1)/\sqrt{T} + o(1/\sqrt{T}) & \text{for all } x^1 \geq 0 \\ \beta_{\theta_0}^-(x^1) + \gamma_{\theta_0}^-(x^1)/\sqrt{T} + o(1/\sqrt{T}) & \text{for all } x^1 < 0. \end{cases} \quad (9)$$

The above definition means that second order asymptotic efficiency implies highest probability concentration around the true value with respect to the second order asymptotic distribution.

**3. SECOND ORDER EFFICIENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN LOCALLY STATIONARY PROCESSES**

In this section we shall show that if we appropriately modify the maximum likelihood estimator in Gaussian locally stationary processes, then it is second order asymptotically efficient in the sense of
Definition 2.1. First we give the precise definition of multivariate locally stationary processes which is due to Dahlhaus (2000).

**Definition 2.** A sequence of Gaussian multivariate stochastic processes $X_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(d)})'$ $(t = 1, \ldots, T)$ is called locally stationary with transfer function matrix $A^\circ$ and mean function vector $\mu$ if there exists a representation

$$X_{t,T} = \mu \left( \frac{t}{T} \right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^\circ_{\lambda,T}(\lambda) d\xi(\lambda)$$

(10)

with the following properties:

(i) $\xi(\lambda)$ is a complex valued Gaussian vector process on $[-\pi, \pi]$ with $\xi_a(\lambda) = \xi_a(-\lambda)$, $E\xi_a(\lambda) = 0$ and

$$E\{d\xi_a(\lambda)d\xi_b(\mu)\} = \delta_{ab}\eta(\lambda + \mu)d\lambda d\mu,$$

(11)

where $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period $2\pi$ extension of the Dirac delta function.

(ii) There exist $2\pi$-periodic matrix valued functions $A : [0, 1] \times \mathbb{R} \to \mathbb{C}^{d \times d}$ with $A(u, -\lambda) = \overline{A(u, \lambda)}$ and

$$\sup_{t,\lambda} \left| A^\circ_{\lambda,T}(\lambda)_{ab} - A(t/T, \lambda)_{ab} \right| = O(T^{-1})$$

(12)

for all $a, b = 1, \ldots, d$ and $T \in \mathbb{N}$. $A(u, \lambda)$ and $\mu(u)$ are assumed to be continuous in $u$.

$f(u, \lambda) := A(u, \lambda)\overline{A(u, \lambda)}$ is called the time varying spectral density of the process.

Throughout this section we assume $A^\circ_{\lambda,T}(u, \lambda) = A_{\theta,t,T}(u, \lambda)$ and $\mu(u) = \mu_{\theta}(u)$, so that efficiency is discussed when the model is correctly specified.

We now set down the following assumptions.

**Assumption 1.** (i) There exist $2\pi$-periodic matrix valued functions $A_{\theta} : [0, 1] \times \mathbb{R} \to \mathbb{C}^{d \times d}$ with $A_{\theta}(u, -\lambda) = \overline{A_{\theta}(u, \lambda)}$ whose components are four times differentiable in $\theta$ and

$$\sup_{t,\lambda} \left| \frac{\partial^k}{\partial \theta_1 \cdots \partial \theta_k} \left\{ A^\circ_{\theta,t,T}(\lambda)_{ab} - A_{\theta}(t/T, \lambda)_{ab} \right\} \right| = O(T^{-1})$$

for $k = 0, 1, 2, 3$, (13)
where \( \partial^k_{j_1\ldots j_k} = \partial^k / \partial \theta^{j_1} \ldots \partial \theta^{j_k} \). The components of \( \partial^k_{j_1\ldots j_k} A_\theta(u, \lambda) \) are differentiable in \( u \) and \( \lambda \) with uniformly bounded derivatives.

(ii) All eigenvalues of \( f_\theta(u, \lambda) = A_\theta(u, \lambda) \bar{A}_\theta(u, \lambda)' \) are bounded from below by some \( C > 0 \) uniformly in \( u \) and \( \lambda \).

(iii) The components of \( \mu_\theta(u) \) are four times differentiable in \( \theta \). The components of \( \partial^k_{j_1\ldots j_k} \mu_\theta(u) \) are differentiable in \( u \) with uniformly bounded derivatives.

Second we give the bound distributions of \( \beta^+_{\theta_0}(x^1) + \gamma^+_{\theta_0}(x^2)/\sqrt{T} \) and \( \beta^+_{\theta} \gamma^+_{\theta_0}(x^1) + \gamma^+_{\theta_0}(x^2)/\sqrt{T} \) defined in the previous section. Using the fundamental lemma of Neyman and Pearson these are given by the likelihood ratio test. Thus we consider the problem of testing hypothesis \( H : \theta = \theta_0 + x/\sqrt{T} \)

against the alternative \( A : \theta = \theta_0 \), where \( x = (x^1, \ldots, x^p) \) and \( x_2 = (x^2, \ldots, x^p) \) is an arbitrary but fixed constant. Let \( \bar{X} = (X_{1,T}', \ldots, X_{T,T}')' \), \( \mu = (\mu(1/T)', \ldots, \mu(T/T)')' \) and \( \Sigma_T(A, B) \) be \( T \times T \) block matrix whose \((r, s)\) block is

\[
\left[ \Sigma_T(A, B) \right]_{r,s} = \int_{-\infty}^{\infty} \exp\{i\lambda(r-s)\} A_r(T) B_s(T)(-\lambda)' d\lambda
\]

(14)

\( r, s = 1, \ldots, T \). The log likelihood function based on \( \bar{X} \) is given by

\[
L_T(\theta) = -\frac{d}{2} \log(2\pi) - \frac{1}{2T} \log \det \Sigma_\theta - \frac{1}{2T} (\bar{X} - \mu_\theta)' \Sigma^{-1}_\theta (\bar{X} - \mu_\theta),
\]

(15)

where \( \Sigma_\theta = \Sigma_T(A_\theta^0, A_\theta^0) \). Let \( LR = L_T(\theta_0) - L_T(\theta_0 + x/\sqrt{T}) \). Using Lemma A.8 in Dahlhaus (2000), we can show that

\[
E_{\theta_0}[LR] = \frac{1}{2} I_{ij} x^i x^j + \frac{1}{6\sqrt{T}} (3J_{ij,k} + K_{ijk}) x^i x^j x^k + o(T^{-1}),
\]

\[
cum_{\theta_0}[LR, LR] = I_{ij} x^i x^j + \frac{1}{\sqrt{T}} J_{ij,k} x^i x^j x^k + o(T^{-1}),
\]

\[
cum_{\theta_0}[LR, LR, LR] = -\frac{1}{\sqrt{T}} K_{ijk} x^i x^j x^k + o(T^{-1}),
\]

(16)

\[
E_{\theta_0+x/\sqrt{T}}[LR] = -\frac{1}{2} I_{ij} x^i x^j - \frac{1}{6\sqrt{T}} (3J_{ij,k} + 2K_{ijk}) x^i x^j x^k + o(T^{-1}),
\]

\[
cum_{\theta_0+x/\sqrt{T}}[LR, LR] = I_{ij} x^i x^j + \frac{1}{\sqrt{T}} (J_{ij,k} + K_{ijk}) x^i x^j x^k + o(T^{-1}),
\]

\[
cum_{\theta_0+x/\sqrt{T}}[LR, LR, LR] = -\frac{1}{\sqrt{T}} K_{ijk} x^i x^j x^k + o(T^{-1}),
\]

5
where

\[
I_{ij}(\theta) = -\frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[(\partial^1_if_\theta)(\partial^1_jf_{\theta})^{-1}]d\lambda du \\
+ \frac{1}{2\pi} \int_0^1 \{\partial^1_i\mu_\theta(u)\}'f_\theta(u,0)^{-1}\{\partial^1_j\mu_\theta(u)\}du,
\]

\[
J_{ij,k}(\theta) = -\frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[(\partial^2_{ij}f^{-1}_\theta)(\partial^1_kf_\theta)]d\lambda du \\
+ \frac{1}{2\pi} \int_0^1 \{\partial^2_{ij}\mu_\theta(u)\}'f_\theta(u,0)^{-1}\{\partial^1_k\mu_\theta(u)\}du \\
+ \frac{1}{2\pi} \int_0^1 \{\partial^1_i\mu_\theta(u)\}'\{\partial^1_j\mu_\theta(u)\}'\{\partial^1_k\mu_\theta(u)\}du[2,ij],
\]

\[
K_{ij,k}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[(f^{-1}_j\theta f_\theta)f^{-1}_\theta(\partial^1_i\mu_\theta(u))^\top(\partial^1_jf_\theta)f^{-1}_\theta(\partial^1_kf_\theta)]d\lambda du[2] \\
- \frac{1}{2\pi} \int_0^1 \{\partial^1_i\mu_\theta(u)\}'\{\partial^1_j\mu_\theta(u)\}'\{\partial^1_k\mu_\theta(u)\}du[3].
\]

Here we use the Einstein summation convention and the simpler notations \(I_{ij}, J_{ij,k}, K_{ij,k}\) etc. are evaluated at \(\theta = \theta_0\). By (4) and the fundamental lemma of Neyman and Perason, the asymptotic power of the most powerful test \(LR\) is given by

\[
\Phi(\sigma) + \frac{1}{6\sqrt{T\sigma}}\phi(\sigma)(3J_{ij,k} + K_{ij,k})x^i\lambda x^jx^k + o(T^{-1/2}),
\]

where \(\Phi(z) = \int_0^z \phi(\sigma)d\sigma, \phi(\sigma) = (2\pi)^{-1/2}\exp(-u^2/2), \sigma = (I_{ij}x^ix^j)^{1/2}x^k\).

Denote by \(I^{ij}\) the \((i,j)\)-th element of the inverse matrix of \(I = \{I_{ij}\}\). The partition \(x = (x^1, x^2)\) induces the following corresponding partition

\[
I(\theta) = \begin{pmatrix} I_{(11)}(\theta) & I_{(12)}(\theta) \\ I_{(21)}(\theta) & I_{(22)}(\theta) \end{pmatrix}.
\]

Since \(x_2\) can take arbitrary values, then the power function of the tests is not larger than the infimum of (16) with respect to \(x_2\). A \(x_2\) minimizing \(\sigma\) is given by \(x_2 = (I_{(22)})^{-1}I_{(21)}x^1\), then \(\sigma^2 = (I^{11})^{-1}(x^1)^2\).

Thus we have the following:
Theorem 1. If \( \hat{\theta}^1 \) is second order AMU and

\[
P_{\theta_0}^T \left[ \sqrt{T} \left( \hat{\theta}^1 - \theta_0^1 \right) \right] \leq x^1 = \Phi \left( x^1 \left( I^{11} \right)^{-1/2} \right) + \frac{(x^1)^2}{6 I^{11} 5/2 \sqrt{T}} \Phi \left( x^1 \left( I^{11} \right)^{-1/2} I^{11} I^{1j} I^{1k} (3 J_{ij,k} + K_{ijk}) + o(T^{-1/2}) \right)
\]

is satisfied, then \( \hat{\theta}^1 \) is second order asymptotically efficient estimator.

Let \( \hat{\theta}_{ML} = (\hat{\theta}_{ML}^1, \ldots, \hat{\theta}_{ML}^p) \) be maximum likelihood estimator which is defined by a value of \( \theta \) that satisfies the equation

\[
0 = \partial^1 L_{T}(\theta).
\]

Write

\[
U^i = \sqrt{T}(\hat{\theta}_{ML}^i - \theta_0^i), \quad Z_i(\theta) = \sqrt{T} \left[ \partial^i L_{T}(\theta) - E \{ \partial^i L_{T}(\theta) \} \right],
\]

\[
Z_{ij}(\theta) = \sqrt{T} \left[ \partial^2_{ij} L_{T}(\theta) - E \{ \partial^2_{ij} L_{T}(\theta) \} \right]
\]

Then we can show the following.

Lemma 1.

\[
U^i = I^{ij} Z_j + \frac{1}{\sqrt{T}} I^{ij} I^{kl} Z_{jk} Z_l - \frac{1}{2 \sqrt{T}} I^{i'i'} I^{j'j} I^{k'k} (J_{i'j'k'3} + K_{i'j'k'}) Z_{j'k'} + o_p(T^{-1/2}).
\]

It is seen that

\[
E_{\theta_0}[U^i] = - \frac{1}{2 \sqrt{T}} I^{ij} I^{kl} (J_{kl,j} + K_{jkl}) + o(T^{-1/2}),
\]

\[
\text{cum}_{\theta_0}[U^i, U^j] = I^{ij} + o(T^{-1/2}),
\]

\[
\text{cum}_{\theta_0}[U^i, U^j, U^k] = -T^{-1/2} I^{i'i'} I^{j'j} I^{k'k} (J_{i'j'k'3} + 2 K_{i'j'k'}) + o(T^{-1/2}),
\]

\[
\text{cum}_{\theta_0}^{J}[U^{i_1}, \ldots, U^{i_J}] = O(T^{-J/2+1}) \quad \text{for} \ J \geq 3.
\]

Applying a general Edgeworth expansion formula (e.g., Taniguchi and Kakizawa, 2000, p.168-170), we have the following theorem
Theorem 2.

\[
P_{\theta_0}[\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \leq x] = \int_{-\infty}^{x} \phi(z, I^{-1}) \left[ 1 - \frac{1}{2\sqrt{T}} I^{ij} I^{kk}(J_{kl,j} + K_{jkl}) H_i(z, I^{-1}) \right. \\
\left. - \frac{1}{6\sqrt{T}} I^{ij} I^{kk} J_{ij,j} + 2K_{ijk} H_i(z, I^{-1}) \right] dz + o(T^{-1/2}),
\]

(24)

where \( z = (z^1, \ldots, z^p)' \),

\[
\phi(z, \Omega) = (2\pi)^{-p/2} |\Omega|^{-1/2} \exp \left( -\frac{1}{2} z' \Omega^{-1} z \right),
\]

(25)

the multivariate normal distribution, and multivariate Hermite polynomials:

\[
H_{j_1 \ldots j_s}(z, \Omega) = (-1)^s \frac{\partial^s}{\partial x_{j_1} \ldots x_{j_s}} \phi(z, \Omega).
\]

(26)

From Theorem 2, it can be seen that \( \hat{\theta}_{ML}^1 \) is not second order AMU. Thus we modify \( \hat{\theta}_{ML}^1 \) as follows:

\[
\hat{\theta}_{ML}^1 = \hat{\theta}_{ML} + \frac{1}{2T} I^{ij}(\hat{\theta}_{ML}) I^{kk}(\hat{\theta}_{ML}) \{ J_{ij,k} + K_{ijk} \} \\
- \frac{1}{6T^{1/2}} I^{ij}(\hat{\theta}_{ML}) I^{kk}(\hat{\theta}_{ML}) \{ 3J_{ij,k} + 2K_{ijk} \}.
\]

(27)

Then we obtain

\[
P_{\theta_0}[\sqrt{T}(\hat{\theta}_{ML}^1 - \theta_0) \leq x] = \Phi(x^1 (I^{11})^{-1/2}) \\
+ \frac{(x^1)^2}{6(I^{11})^{3/2}} \phi(x^1 (I^{11})^{-1/2}) I^{ij} I^{kk} (3J_{ij,k} + K_{ijk}) + o(T^{-1/2}).
\]

(28)

Remembering Theorem 1, we can see that (28) coincides with the bound distribution. Thus we have

**Theorem 3.** The modified MLE \( \hat{\theta}_{ML}^1 \) is second order asymptotically efficient.

4. HIGHER ORDER ROBUSTNESS

In this section, we discuss second order misspecified and time varying robustness of the maximum likelihood estimator. To discuss the problem of higher order asymptotic estimation for parameters of locally stationary processes, the following assumptions imposed.
ASSUMPTION 2. (i)

\[ A_{t,T}(\lambda) = A_{1,\theta,t,T}^{o}(\lambda) + \frac{1}{\sqrt{T}} A_{2,\theta,t,T}^{o}(\lambda) + \frac{1}{T} A_{3,\theta,t,T}^{o}(\lambda), \]
\[ \mu(u) = \mu_{\theta}(u). \]

(ii) There exist $2\pi$-periodic matrix valued functions $A_{i,\theta} : [0,1] \times \mathbb{R} \to \mathbb{C}^{d \times d}$ with $A_{i,\theta}(u,-\lambda) = A_{i,\theta}(u,\lambda)$ whose components are four times differentiable in $\theta$ and

\[ \sup_{t,\lambda} \left| \partial_{j_{1}\ldots j_{k}} \left\{ A_{i,\theta,t,T}^{o}(\lambda)_{ab} - A_{i,\theta}(u/T,\lambda)_{ab} \right\} \right| = o(T^{-1}) \] (29)

for $k = 0,1,2,3$ and $i = 1,2,3$. The components of $\partial_{j_{1}\ldots j_{k}} A_{i,\theta}(u,\lambda)$ ($i = 1,2,3$) are differentiable in $u$ and $\lambda$ with uniformly bounded derivatives.

(iii) Let

\[ f_{\theta}(u,\lambda) = f_{1,\theta}(u,\lambda) + \frac{1}{\sqrt{T}} f_{2,\theta}(u,\lambda) + \frac{1}{T} f_{3,\theta}(u,\lambda) + o(T^{-1}). \]

Then, $f_{i,\theta}(u,\lambda)$ ($i = 1,2,3$) fulfill Assumption 3.1. (ii).

(iv) $\mu_{\theta}(u)$ fulfills Assumption 3.1. (iii).

We define the MLE $\tilde{\theta}_{\text{MLE}}$ in the misspecified case by a solution of equation

\[ 0 = \partial_{i}^{1} \tilde{L}_{T}(\theta), \quad i = 1, \ldots, p, \] (30)

where

\[ \tilde{L}_{T}(\theta) = -\frac{d}{2} \log(2\pi) - \frac{1}{2T} \log \det \Sigma_{1,\theta} - \frac{1}{2T} (X - \mu_{\theta})^{T} \Sigma_{1,\theta}^{-1} (X - \mu_{\theta}), \] (31)

and $\Sigma_{1,\theta} = \Sigma_{T}(A_{1,\theta}^{o}, A_{1,\theta}^{o})$. 

Write
\[ \tilde{U}^i = \sqrt{T} (\tilde{\theta}_{ML}^i - \theta^i), \quad \tilde{Z}_i(\theta) = -\sqrt{T} [\partial_i \tilde{L}_T(\theta) - E_\theta \{ \partial_i \tilde{L}_T(\theta) \}], \quad (32) \]
\[ \tilde{Z}_{ij}(\theta) = -\sqrt{T} \{ \partial_{ij} \tilde{L}_T(\theta) - E_\theta \{ \partial_{ij} \tilde{L}_T(\theta) \} \}, \]
\[ \tilde{Z}_{ij}(\theta) = -\sqrt{T} \{ \partial_{ij} \tilde{L}_T(\theta) - E_\theta \{ \partial_{ij} \tilde{L}_T(\theta) \} \}, \]
where \( E_\theta \) denotes the expectation under the true model.

In the same way as the previous calculations, it follows that
\[ \tilde{U}^i = I_{ij}^i (\tilde{Z}_j - \Gamma_{ij}^{(1)}) - \frac{1}{\sqrt{T}} I_{ij}^{(2)} \frac{1}{T} \tilde{Z}_j \tilde{Z}_k (\tilde{Z}_l - \Gamma_{ij}^{(1)}) + \frac{1}{\sqrt{T}} I_{ij}^{(2)} \tilde{Z}_j (\tilde{Z}_l - \Gamma_{ij}^{(1)}) + o_p(T^{-1/2}), \quad (33) \]
where
\[ \Gamma_{ij}^{(1)}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[\{ \partial_i f_{1,\theta}(u,0) \} f_{2,\theta}(u) d\lambda du, \]
\[ \Gamma_{ij}^{(2)}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[\{ \partial_i f_{1,\theta}(u,0) \} f_{3,\theta}(u) d\lambda du, \]
\[ \Delta_{ij}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[\{ \partial_i^2 f_{1,\theta}(u,0) \} f_{2,\theta}(u) d\lambda du. \]

From direct verification, we can show that
\[ E_{\theta_0}[\tilde{Z}_i \tilde{Z}_j] = I_{ij} + \frac{1}{\sqrt{T}} \frac{1}{4\pi} \int_0^1 \int_{-\pi}^\pi \text{tr}[\{ \partial_i f_{1,\theta_0}(u,0) \} f_{2,\theta_0}(u,0) f_{3,\theta_0}(u,0) \} du \]
\[ + \frac{1}{\sqrt{T}} \frac{1}{2\pi} \int_0^1 \{ \partial_i \mu_{\theta_0}(u) \} f_{1,\theta_0}(u,0) f_{2,\theta_0}(u,0) f_{3,\theta_0}(u,0) \} du \]
\[ = I_{ij} + \frac{1}{\sqrt{T}} \Delta_{ij} + o(T^{-1/2}) \quad \text{(say),} \]
\[ E_{\theta_0}[\tilde{Z}_{ij} \tilde{Z}_k] = J_{ij,k} + O(T^{-1/2}), \]
\[ E_{\theta_0}[\tilde{Z}_i \tilde{Z}_j \tilde{Z}_k] = T^{-1/2} K_{ijk} + O(T^{-1}), \quad (34) \]

10
and the Jth \((J \geq 3)\) order cumulant of \(\tilde{Z}_i(\theta), \ldots, \tilde{Z}_{i,J}(\theta), \tilde{Z}_{j,k}(\theta), \ldots, \tilde{Z}_{J,J,k}(\theta)\) \((J_1 + J_2 = J)\) satisfies

\[
\text{cum}^{(J)}[\tilde{Z}_i(\theta), \ldots, \tilde{Z}_{i,J}(\theta), \tilde{Z}_{j,k}(\theta), \ldots, \tilde{Z}_{J,J,k}(\theta)] = O(T^{-J/2+1}).
\]  

From (33)-(35), it is seen that

\[
E_{\theta_0}[\tilde{U}^j] = -I^{ij}I^{(1)}_j - \frac{1}{\sqrt{T}}I^{ij}I^{(2)}_j + \frac{1}{\sqrt{T}}I^{ij}I^{kl}\Delta_{jk}I^{(1)}_l
\]

\[
- \frac{1}{2\sqrt{T}}I^{ij}I^{kl}I^{kk'}(J_{i'j',k'}[3] + K_{i'j',k'})I^{(1)}_j I^{(1)}_k
\]

\[
- \frac{1}{2\sqrt{T}}I^{ij}I^{kl}(J_{kl,j} + K_{jkl}) + o(T^{-1/2}),
\]

\[
\text{cum}_{\theta_0}[\tilde{U}^i, \tilde{U}^j] = I^{ij} + \frac{1}{\sqrt{T}}I^{ik}I^{j}I^{kl}(\Delta_{1,kl} - 2\Delta_{kl})
\]

\[
+ \frac{1}{\sqrt{T}}I^{ij}I^{kl}I^{kk'}(J_{i'j',k'}[3] + J_{i'j',k'} + 2K_{i'j',k'}) + o(T^{-1/2}),
\]

\[
\text{cum}^2_{\theta_0}[\tilde{U}^i, \tilde{U}^j, \tilde{U}^k] = -T^{-1/2}I^{i'd}I^{j'd}I^{kk'}(J_{d'j',k'}[3] + 2K_{d'j',k'}) + O(T^{-1/2}),
\]

\[
\text{cum}^J_{\theta_0}[\tilde{U}^{i_1}, \ldots, \tilde{U}^{i_J}] = O(T^{-J/2+1}) \quad \text{for} \ J \geq 4.
\]

Applying a general formula (e.g., Taniguchi and Kakizawa, 2000, p.168-170), we have

**Theorem 4.** If \(\Gamma_i^{(1)} = 0 \ (i = 1, \ldots, p)\), then the Edgeworth expansion of the distribution function of \(\sqrt{T}(\hat{\theta}_T - \theta_0)\) is given by

\[
P_{\theta_0}^{\hat{\theta}_T}[(\sqrt{T}(\hat{\theta}_T - \theta_0)) \leq z] = \int_{-\infty}^{z} \phi(x, I^{-1}) \left[ 1 - \frac{1}{2\sqrt{T}}I^{ij}\{I^{(2)}_j + I^{kl}(J_{kl,j} + K_{jkl})\}H_i(x, I^{-1}) + \frac{1}{2\sqrt{T}}I^{ik}I^{j}I^{kl}(\Delta_{1,kl} - 2\Delta_{kl})H_{ij}(x, I^{-1}) \right. \\
- \frac{1}{6\sqrt{T}}I^{ij}I^{kl}I^{kk'}(J_{i'j',k'}[3] + 2K_{i'j',k'})H_{ij}(x, I^{-1}) \left. \right] dx + o(T^{-1/2}).
\]  

**Remark 1.** The condition \(\Gamma_i^{(1)} = 0 \ (i = 1, \ldots, p)\) ensures that the distribution of \(\sqrt{T}(\hat{\theta}_{ML} - \theta)\) converges to the multivariate normal distribution with zero mean vector. If \(\Gamma_i^{(2)} = 0 \ (i = 1, \ldots, p)\) is satisfied, then the bias of \(\hat{\theta}_{ML}\) is equal to that of \(\hat{\theta}_{ML}\) up to second order.
From

\[ I_{ij}(\theta) = \frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr}[\partial_{i} f_{\theta}(u) f_{\theta}^{-1} \partial_{j} f_{\theta}(u)] d\lambda du \]

\[ + \frac{1}{2\pi} \int_{0}^{1} \{ \partial_{i} \mu_{\theta}(u) \} f_{\theta}(u, 0)^{-1} \{ \partial_{j} \mu_{\theta}(u) \} du \]

\[ = I_{ij}(\theta) + \frac{1}{\sqrt{T}} \frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr}[-(\partial_{i} f_{1,\theta}) f_{1,\theta}^{-1} (\partial_{j} f_{1,\theta}) f_{1,\theta}^{-1} f_{2,\theta} f_{1,\theta}^{-1}[2]] d\lambda du \]

\[ + \frac{1}{\sqrt{T}} \frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr}[(\partial_{i} f_{2,\theta}) f_{1,\theta}^{-1} (\partial_{j} f_{1,\theta}) f_{1,\theta}^{-1}[2]] d\lambda du \]

\[ - \frac{1}{\sqrt{T}} \frac{1}{2\pi} \int_{0}^{1} \{ \partial_{i} \mu_{\theta}(u) \} f_{1,\theta}(u, 0)^{-1} f_{2,\theta}(u, 0) f_{1,\theta}(u, 0) f_{1,\theta}(u, 0)^{-1} \{ \partial_{j} \mu_{\theta}(u) \} du + o(T^{-1/2}) \]

\[ = I_{1,ij}(\theta) + \frac{1}{\sqrt{T}} \Delta_{2,ij}(\theta) + o(T^{-1/2}) \quad \text{(say),} \]

we have

\[ P^{ij} = I_{i}^{j} - \frac{1}{\sqrt{T}} I^{ik} I^{jl} \Delta_{2,kl} + o(T^{-1/2}). \]

It is easy to see that \( \Gamma_{1}^{(i)} = 0 \) implies \( 2\Delta_{ij} - \Delta_{1,ij} - \Delta_{2,ij} = 0 \). Therefore, if we put

\[ \tilde{\theta}_{ML}^{ij} = \tilde{\theta}_{ML} + \frac{1}{T} P^{ij} (\tilde{\theta}_{ML}) V_{j}^{(2)} (\tilde{\theta}_{ML}), \]

then we obtain

\[ P^{T}_{\tilde{\theta}_{0}} [\sqrt{T}(\tilde{\theta}_{ML} - \theta_{0}) \leq z] = P^{T}_{\theta_{0}} [\sqrt{T}(\tilde{\theta}_{ML} - \theta_{0}) \leq z] + o(T^{-1/2}). \]

Thus we have

**Corollary 1.** If \( \Gamma_{1}^{(i)} = 0 \ (i = 1, \ldots, p) \) is satisfied, then the distribution function of the modified maximum likelihood estimator \( \tilde{\theta}_{ML} \) is equal to that of the \( \hat{\theta}_{ML} \) with an error \( o(T^{-1/2}) \).
is satisfied, then we say that the estimator \( \tilde{\theta}_{ML} \) of \( \theta \) is asymptotically misspecified robustness with an error \( o(T^{-1/2}) \).

**Corollary 2.** If \( \Gamma_{(1)}^j = \Gamma_{(2)}^j = 0 \) is satisfied, then \( \tilde{\theta}_T \) is asymptotically misspecified robustness with an error \( o(T^{-1/2}) \).

If
\[
\int_0^1 \int_{-\pi}^{\pi} \text{tr}[(\partial_{f_\theta} f_\theta^{-1})f_\theta]d\lambda du = 0
\]
is satisfied, then we say that the parameter \( \theta \) is innovation-free w.r.t. \( f_\theta \).

**Remark 2.** From (31), if the parameter \( \theta \) is innovation-free w.r.t. \( f_{1,\theta} \), \( f_{2,\theta} = af_{1,\theta} \) and \( f_{3,\theta} = bf_{1,\theta} \), \( a, b \in \mathbb{R} \), then \( \Gamma_{(1)}^j = \Gamma_{(2)}^j = 0 \) holds.

We consider the situation where all of the quantities appearing in second order Edgeworth expansion for an estimator have the form
\[
\int_0^1 \int_{-\pi}^{\pi} g_1(\lambda, u)d\lambda du + \int_0^1 g_2(u)du.
\]
If \( g_1(\lambda, u) \) and \( g_2(u) \) are independent of \( u \), then we say that the estimator is time varying robustness up to second order.

**Corollary 3.** If
\[
A_\theta(u, \lambda) = B(u)C_\theta(\lambda), \quad \mu_\theta(u) = B(u)\nu_\theta, \tag{38}
\]
are satisfied, then \( \tilde{\theta}_{ML}, \hat{\theta}_{ML} \) are time varying robustness.

**Remark 3.** If the condition (37) holds, then locally stationary processes \( X_{t,T} \) can be written as
\[
X_{t,T} = B\left(\frac{t}{T}\right) \left\{ \nu_\theta + \int_{-\pi}^{\pi} \exp(i\lambda t)C_\theta^0(\lambda)d\xi(\lambda) \right\},
\]
\[
= B\left(\frac{t}{T}\right) \times \{ \text{stationary process} \}.
\]
EXAMPLE 1. To observe the non-stationary effect, we consider the following model:

\[ X_{t,T} + b_\theta \left( \frac{t}{T} \right) X_{t-1,T} = a_\theta \left( \frac{t}{T} \right) \varepsilon_t, \quad t = 1, \ldots, T, \]

where \( a_\theta(u) = a \exp\left(-\frac{(u - \theta)^2}{2}\right), \) \( b_\theta(u) = u \theta, \) \( \theta^1 < 0, \) \( 1 < \theta^1, \) \( |\theta^2| < 1 \) and \( \varepsilon_t \)'s are i.i.d. \((0, 1)\) random variables. Then the time varying spectral density is given by

\[ f_\theta(u, \lambda) = \frac{1}{2\pi} \left| \frac{a_\theta(u)}{1 + b_\theta(u) e^{-i\lambda u}} \right|^2, \quad \theta = (\theta^1, \theta^2). \]

By the residue theorem, it is shown that

\[ I_{11} = 2 \int_0^1 \left\{ \frac{\partial_1 a_\theta(u)}{a_\theta(u)} \right\}^2 du, \quad I_{12} = 0, \]

\[ I_{21} = 0, \quad I_{22} = \int_0^1 \left\{ \frac{\partial_2 b_\theta(u)}{1 - \{b_\theta(u)\}^2} \right\}^2 du, \]

\[ J_{11,1} = 2 \int_0^1 \frac{\partial_1 a_\theta(u)}{a_\theta(u)} \frac{\partial_2 a_\theta(u)}{a_\theta(u)} du - \frac{3}{4} K_{111}, \quad J_{11,2} = J_{12,1} = 0, \]

\[ J_{12,2} = J_{22,1} = \frac{1}{2} K_{122}, \quad J_{22,2} = -\frac{1}{3} K_{222} + \int_0^1 \frac{\partial_2 b_\theta(u) \partial_2 b_\theta(u)}{1 - \{b_\theta(u)\}^2} du, \]

and

\[ K_{111} = 8 \int_0^1 \left\{ \frac{\partial_1 a_\theta(u)}{a_\theta(u)} \right\}^3 du, \quad K_{112} = 0, \]

\[ K_{122} = 4 \int_0^1 \frac{\partial_1 a_\theta(u)}{a_\theta(u)} \left\{ \frac{\partial_2 b_\theta(u)}{1 - \{b_\theta(u)\}^2} \right\}^2 du, \quad K_{222} = 6 \int_0^1 \frac{\partial_2 b_\theta(u) \partial_2 b_\theta(u)}{[1 - \{b_\theta(u)\}^2]^2} du, \]

Let \( \Delta^S(u) \) be \( \Delta^{LS} \) in stationary case (i.e., \( u \) is treated as a known parameter). We introduce the criterion

\[ D(\theta) = \int_0^1 \left\{ \Delta^{LS} - \Delta^S(u) \right\}^2 du, \]

which measures the time varying effect in efficient estimation.
(i) Suppose that $\theta^1$ is unknown, and that $\theta^2$ is known. Then it is easy to show

$$\Delta_{LS} = \frac{3}{4} \frac{(1 - \theta^1)^4 - (\theta^1)^4}{(1 - \theta^1)^3 + (\theta^1)^3}, \quad \Delta_S(u) = \frac{1}{3(u - \theta^1)}.$$ 

In Figure 1, we plotted $D(\theta^1)$ with $-2 < \theta^1 < 0$ and $1 < \theta^1 < 3$. From the figure we observe that the time varying effect becomes large as $\theta^1 \nearrow 0$ or $\theta^1 \searrow 1$.

(ii) Suppose that $\theta^2$ is unknown, and that $\theta^1$ is known. Then it is easy to show

$$\Delta_{LS} = \frac{1}{6} \left\{ \frac{1}{(\theta^2)^2} + \frac{1}{2(\theta^2)^3} \log \frac{1 + \theta^2}{1 - \theta^2} \right\}^{-2} \left[ 3 \left\{ 3 - 2(\theta^2)^2 \right\} - \frac{9}{2(\theta^2)^4} \log \frac{1 + \theta^2}{1 - \theta^2} \right],$$

$$\Delta_S(u) = \theta^2.$$ 

In Figure 2, we plotted $D(\theta^2)$ with $-1 < \theta^2 < 1$. From the figure we observe that the time varying effect becomes large as $|\theta^2| \nearrow 1$.

(iii) Suppose that $\theta^1$ is a parameter of interest, and that $\theta^2$ is a nuisance parameter. Then it is easy to show

$$\Delta_{LS} = \frac{3}{4} \frac{(1 - \theta^1)^4 - (\theta^1)^4}{(1 - \theta^1)^3 + (\theta^1)^3} + \frac{3}{4} \frac{1}{(1 - \theta^1)^3 + (\theta^1)^3} \left\{ - \frac{1}{(\theta^2)^2} + \frac{1}{2(\theta^2)^3} \log \frac{1 + \theta^2}{1 - \theta^2} \right\},$$

$$\times$$

$$\Delta_S(u) = \frac{5}{6(u - \theta^1)}.$$ 

(iv) Suppose that $\theta^2$ is a parameter of interest, and that $\theta^1$ is a nuisance parameter. From $J_{112} = K_{112} = 0$, it is seen that the modification term is not affected by the nuisance parameter. Hence, $\Delta_{LS}$ and $\Delta_S(u)$ are the same as the case (ii).
ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Professor Masanobu Taniguchi for his encouragement and guidance.

REFERENCES


Figure 1:

Figure 2: