The Annals of Statistics
1986, Vol. 14, No. 3, 1214–1225

SECOND-ORDER RISK STRUCTURE OF GLSE AND MLE IN A REGRESSION WITH A LINEAR PROCESS[1]

BY YASUYUKI TOYOOKA

Osaka University

In a regression model with an error that is a general linear process, the second-order expansion of the risk matrix of GLSE or MLE is obtained. A set of sufficient conditions for the effect of estimating the structural parameter of the linear process to vanish in the above expansion is obtained. The relation of the covariance matrix of SLSE with those of GLSE and MLE up to $O(T^{-2})$ is elucidated.

1. Introduction. In this paper we consider the estimation of β when θ is unknown in the regression model

$$y_t = x_t'\beta + u_t, \tag{1.1}$$

where $\{x_t\}$ is a sequence of p-dimensional fixed designed vectors, $\beta \in R^p$, and u_t is a general linear process,

$$u_t = \sum_{j=0}^{\infty} g_j(\theta)\epsilon_{t-j}, \tag{1.2}$$

with $g_0(\theta) = 1$, $\sum_{j=0}^{\infty} g_j(\theta)^2 < \infty$, and $\{\epsilon_t\}$ is a sequence of i.i.d. $N(0, \sigma^2)$ random variables, where

$$\sigma^2 = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_\theta(\lambda)\, d\lambda\right\}$$

with

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} g_j(\theta)e^{i\lambda j}\right|^2 = \frac{\sigma^2}{2\pi}g(\lambda,\theta) \quad \text{(say)}. \tag{1.3}$$

The parameter space of θ is an open subset θ of R^1. As is well known, the finite parameter stationary models, such as the autoregressive model, moving average model, and autoregressive-moving average model, can be expressed as the general linear process $\{u_t\}$ in (1.2).

When $\{y_t\colon t = 1, \dots, T\}$ and $\{x_t\colon t = 1, \dots, T\}$ are observed, the statistical linear model is

$$y = X\beta + u, \qquad E(u) = 0 \quad \text{and} \quad \text{Cov}(u) = V(\theta), \tag{1.4}$$

Received March 1984; revised August 1985.
[1]Research partially supported by U.S. Bureau of the Census through Joint Statistical Agreement J.S.A. 84-2.
AMS 1980 *subject classifications*. Primary 62F10; secondary 62J10.
Key words and phrases. GLSE, Grenander's condition, linear process, MLE, regression with a linear process, second-order risk, SLSE.

where $y = [y_1, \dots, y_T]'$, $X = [x_1, \dots, x_T]'$, and $u = [u_1, \dots, u_T]'$. Assume that $\operatorname{rank}(X) = p$ if $T \geq p$ for simplicity of discussion. The covariance structure $V(\theta)$ does not generally satisfy the condition (Mitra and Rao (1969)) for which the simple least squares estimator (SLSE) $\hat{\beta} = (X'X)^{-1}X'y$ is equivalent to the best linear unbiased estimator (BLUE)

$$\hat{\beta}_W = \left\{X'V^{-1}(\theta)X\right\}^{-1}X'V^{-1}(\theta)y. \tag{1.5}$$

Since θ is unknown, alternative estimators other than SLSE are the generalized least squares estimator (GLSE) and the maximum likelihood estimator (MLE). When $\{u_t\}$ is a stationary autoregressive process with parameter θ, Toyooka (1985) proved that the maximum likelihood estimator (MLE) $\hat{\beta}_{\mathrm{MLE}}$ is a GLSE of the form

$$\hat{\beta}_{\hat{W}_1} = \left\{X'V^{-1}(\hat{\theta}_1(\tilde{u}))X\right\}^{-1}X'V^{-1}(\hat{\theta}_1(\tilde{u}))y (= \hat{\beta}_{\mathrm{MLE}}),$$

where $\hat{\theta}_1(\tilde{u})$ is some function of $\tilde{u} = [I - X(X'X)^{-1}X']u$. The risk matrix of a usual GLSE

$$\hat{\beta}_{\hat{W}_2} = \left\{X'V^{-1}(\hat{\theta}_2(\tilde{u}))X\right\}^{-1}X'V^{-1}(\hat{\theta}_1(\tilde{u}))y,$$

where

$$\hat{\theta}_2(\tilde{u}) = \sum_{t=2}^{T} \tilde{u}_t\tilde{u}_{t-1} \Big/ \sum_{t=2}^{T} \tilde{u}_{t-1}^2,$$

is equivalent to that of $\hat{\beta}_{\mathrm{MLE}}$ up to $O(T^{-2})$ in the previous paper. Moreover, Toyooka gave sufficient conditions for the estimation effect of θ to vanish from the expansion of the risk matrix of $\hat{\beta}_{\hat{W}_2}$ up to $O(T^{-2})$.

In the present paper we examine the above sufficient condition under the more general error process (1.2). In Section 2, we formulate the problem. We give the second-order expansion of the risk matrix of GLSE or MLE and the sufficient condition for the estimation effect of θ contained in this expansion to vanish in Section 3. In Section 4 we give the statistical implication of the sufficient condition. An extension to the case where θ is multidimensional is straightforward, and therefore is omitted.

2. The estimator of the structural parameter θ. Let the estimated residual be

$$\begin{aligned} \tilde{u} &= y - X\hat{\beta} \\ &= \left\{I - X(X'X)^{-1}X'\right\}u. \end{aligned} \tag{2.1}$$

We use the Whittle functional for $\tilde{u}$ (see Walker (1964)), that is,

$$\begin{aligned} U_T(\tilde{u}, \theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T}\tilde{u}_t e^{i\lambda t}\right|^2 \{g(\lambda, \theta)\}^{-1}\,d\lambda \\ &= T\sum_{s=-(T-1)}^{T-1} \alpha_s(\theta)C_s, \end{aligned} \tag{2.2}$$

where

$$\alpha_s(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\iota\lambda s}\{g(\lambda,\theta)\}^{-1}\,d\lambda$$

and

$$C_s = \sum_{t=1}^{T-|s|} \tilde{u}_t\tilde{u}_{t+|s|}/T. \tag{2.3}$$

As the estimator $\hat{\theta}$ of θ, we use the value of θ that minimizes $U_T(\tilde{u},\theta)$ with respect to θ.

For the linear process $\{u_t\}$ defined in (1.2), the covariance matrix of u is

$$V(\theta) = \left[\int_{-\pi}^{\pi} e^{\iota(t-s)\lambda} f_\theta(\lambda)\,d\lambda\right]_{t,s=1,\dots,T}. \tag{2.4}$$

So our GLSE is

$$\hat{\beta}_{\hat{W}} = \{X'V^{-1}(\hat{\theta})X\}^{-1}X'V^{-1}(\hat{\theta})y. \tag{2.5}$$

First we get

LEMMA 2.1. *The estimator $\hat{\theta}$ for the structural parameter is an even function of $\tilde{u}$.*

PROOF. From (2.2), the normal equation is

$$\frac{\partial}{\partial\theta}U_T(\tilde{u},\hat{\theta}) = T\sum_{s=-(T-1)}^{T-1}\frac{\partial}{\partial\theta}\alpha_s(\hat{\theta})C_s = 0.$$

Then from the implicit function theorem, $\hat{\theta}$ is a function of C_s ($s = -(T-1),\dots,(T-1)$), which implies that $\hat{\theta}$ is an even function of $\tilde{u}$. □

Moreover,

LEMMA 2.2. *The GLSE $\hat{\beta}_{\hat{W}}$ is an unbiased estimator for β.*

PROOF. Since $V(\theta)$ is a continuous function of θ and $\hat{\theta}$ is an even function of u, $\hat{\beta}_{\hat{W}}$ is an odd function of u. Therefore, $E(\hat{\beta}_{\hat{W}}) = \beta$. □

An interesting lemma by Kariya and Toyooka (1985) is

LEMMA 2.3. *For the GLSE $\hat{\beta}_{\hat{W}}$,*

$$E[(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)(\hat{\beta}_W - \beta)'] = 0,$$

where $\hat{\beta}_W = \{X'V^{-1}(\theta)X\}^{-1}X'V^{-1}(\theta)y$ is BLUE.

From Lemmas 2.2 and 2.3,

$$E[(\hat{\beta}_{\hat{W}} - \beta)(\hat{\beta}_{\hat{W}} - \beta)'] = \mathrm{Cov}(\hat{\beta}_{\hat{W}}) \tag{2.6}$$
$$= \mathrm{Cov}(\hat{\beta}_W) + E[(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)'].$$

The first term is the covariance matrix of the BLUE and the second term is the estimation effect of θ.

3. Asymptotic evaluation of $E[(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)']$. One main object is to evaluate the leading term of the second term of (2.6). We consider the situation in which $\{x_t\}$ is a sequence of bounded designed functions of t.

Let, for $i, j = 1, \dots, p$,

$$a_{ij}^T(h) = \sum_{t=1}^{T-h} x_{it}x_{jt+h}, \qquad h = 0, 1, \dots$$
$$= \sum_{t=1-h}^{T} x_{it}x_{jt+h}, \qquad h = 0, -1, \dots.$$

We impose the following regularity conditions on the regression functions $\{x_t\}$ (see Grenander (1954)):

R.1 $a_{ii}^T(0) = \|x_i\|_T^2 \to \infty$ as $T \to \infty$, where $\|x_i\|_T = (\sum_{t=1}^T x_{it}^2)^{1/2}$ for $i = 1, \dots, p$.
R.2 $\lim_{T\to\infty} x_{iT+1}^2 / a_{ii}^T(0) = 0$ for $i = 1, \dots, p$.
R.3 The limit of

$$a_{ij}^T(h)/T = \gamma_{ij}^T(h) \quad \text{as } T \to \infty$$

exists for every $i, j = 1, \dots, p$ and $h = 0, \pm 1, \dots$. Let

$$\lim_{T\to\infty} \gamma_{ij}^T(h) = \rho_{ij}(h)$$

for $i, j = 1, \dots, p$ and $h = 0, \pm 1, \dots$ and let $R(h) = [\rho_{ij}(h)]$.
R.4 $R(0)$ is nonsingular.

Then under these conditions, there exists a matrix-valued regression spectral measure $M(\lambda)$ such that

$$R(h) = \int_{-\pi}^{\pi} e^{ih\lambda}\, dM(\lambda). \tag{3.1}$$

From Walker's result (1964), we have the following:

LEMMA 3.1. *If $\sum_{s=-(T-1)}^{T-1} (\partial/\partial\theta)\alpha_s(\theta) = o(T\sqrt{T})$, then as $T \to \infty$, $\sqrt{T}(\hat{\theta} - \theta) \to_D N(0, w^{-1})$, where D denotes convergence in distribution and*

$$w = \frac{1}{4\pi}\int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta}\log g(\lambda, \theta)\right\}^2 d\lambda.$$

Proof. Let

$$U_T(u, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^{T} u_t e^{i\lambda t} \right|^2 \{g(\lambda, \theta)\}^{-1} d\lambda$$

$$= T \sum_{s=-(T-1)}^{T-1} \alpha_s(\theta) C_s,$$

where

$$C_s = \sum_{t=1}^{T-|s|} u_t u_{t+|s|} \Big/ T$$

and let

$$U_T(\tilde{u}, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^{T} \tilde{u}_t e^{i\lambda t} \right|^2 \{g(\lambda, \theta)\}^{-1} d\lambda$$

$$= T \sum_{s=-(T-1)}^{T-1} \alpha_s(\theta) \tilde{C}_s,$$

where

$$\tilde{C}_s = \sum_{t=1}^{T-|s|} \tilde{u}_t \tilde{u}_{t+|s|} \Big/ T.$$

Let $\bar{\theta}$ be the minimizing value of $U_T(u, \theta)$ and $\hat{\theta}$ be the minimizing value of $U_T(\tilde{u}, \theta)$. Then

$$T \sum_{s=-(T-1)}^{T-1} \frac{\partial}{\partial \theta} \alpha_s(\bar{\theta}) C_s = 0$$

and

$$T \sum_{s=-(T-1)}^{T-1} \frac{\partial}{\partial \theta} \alpha_s(\hat{\theta}) \tilde{C}_s = 0.$$

Therefore

$$-\sum_{s=-(T-1)}^{T-1} \frac{1}{T} \frac{\partial^2}{\partial \theta^2} \alpha_s(\theta^*) \sqrt{T}(\bar{\theta} - \theta) = \frac{1}{\sqrt{T}} \sum_{s=-(T-1)}^{T-1} \frac{\partial}{\partial \theta} \alpha_s(\theta) C_s,$$

where $\theta^* = \lambda_1 \bar{\theta} + (1 - \lambda_1)\theta_0$ and

$$-\sum_{s=-(T-1)}^{T-1} \frac{1}{T} \frac{\partial^2}{\partial \theta^2} \alpha_s(\theta^{**}) \sqrt{T}(\hat{\theta} - \theta) = \frac{1}{\sqrt{T}} \sum_{s=-(T-1)}^{T-1} \frac{\partial}{\partial \theta} \alpha_s(\theta) \tilde{C}_s,$$

where $\theta^{**} = \lambda_2\hat{\theta} + (1 - \lambda_2)\theta_0$. From Walker's result,

$$\begin{aligned}\operatorname{plim}\frac{1}{T}\sum_{s=-(T-1)}^{T-1}\frac{\partial^2}{\partial\theta^2}\alpha_s(\theta^*) &= \operatorname{plim}\frac{1}{T}\sum_{s=-(T-1)}^{T-1}\frac{\partial^2}{\partial\theta^2}\alpha_s(\theta)\\ &= \operatorname{plim}\frac{1}{T}\sum_{s=(T-1)}^{T-1}\frac{\partial^2}{\partial\theta^2}\alpha_s(\theta^{**})\\ &= 2\sigma^2 w.\end{aligned}$$

Remark that

$$C_s - \tilde{C}_s = O_p(1/T)$$

and

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{s=-(T-1)}^{T-1}\frac{\partial}{\partial\theta}\alpha_s(\theta)\tilde{C}_s &= \frac{1}{\sqrt{T}}\sum_{s=-(T-1)}^{T-1}\frac{\partial}{\partial\theta}\alpha_s(\theta)(\tilde{C}_s - C_s)\\ &\quad + \frac{1}{\sqrt{T}}\sum\frac{\partial}{\partial\theta}\alpha_s(\theta)C_s.\end{aligned}$$

From this, if $\sum_{s=-(T-1)}^{T-1}(\partial/\partial\theta)\alpha_s(\theta) = o(T\sqrt{T})$, then

$$\frac{1}{\sqrt{T}}\sum_{s=-(T-1)}^{T-1}\frac{\partial}{\partial\theta}\alpha_s(\theta)\tilde{C}_s \sim \frac{1}{\sqrt{T}}\sum_{s=-(T-1)}^{T-1}\frac{\partial}{\partial\theta}\alpha_s(\theta)C_s.$$

Therefore, from Walker's result for the asymptotic normality of the right-hand side,

$$\sqrt{T}(\hat{\theta} - \theta) \sim N(0, 1/w). \qquad \square$$

From the fact that the parameters β and θ are orthogonal in the Fisher information matrix sense, we can get, by using a discussion similar to Hildreth (1969),

LEMMA 3.2. *As* $T \to \infty$,

$$\begin{bmatrix} T^{1/2}(\hat{\theta} - \theta) \\ T^{-1/2}X'V^{-1}(\theta)u \\ T^{-1/2}X'\dfrac{\partial}{\partial\theta}V^{-1}(\theta)u \end{bmatrix} \to_D N\left[0, \begin{pmatrix} w^{-1} & 0 & 0 \\ 0 & \lim\limits_{T\to\infty} T^{-1}X'V^{-1}(\theta)X & \lim\limits_{T\to\infty} T^{-1}X'\dfrac{\partial}{\partial\theta}V^{-1}(\theta)X \\ 0 & \lim\limits_{T\to\infty} T^{-1}X'\dfrac{\partial}{\partial\theta}V^{-1}(\theta)X & \lim\limits_{T\to\infty} T^{-1}X'\dfrac{\partial}{\partial\theta}V^{-1}(\theta)V(\theta)\dfrac{\partial}{\partial\theta}V^{-1}(\theta)X \end{pmatrix}\right].$$

By using these results and applying an argument similar to that of Toyooka (1985), we obtain

LEMMA 3.3. *The second term of* (2.6) *is*

$$E[(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)'] = T^{-2}w^{-1} \lim_{T\to\infty} T\{X'V^{-1}(\theta)X\}^{-1}$$

$$\times\Big[\lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)V(\theta)\frac{\partial}{\partial\theta}V^{-1}(\theta)X$$

$$(3.2)\qquad - \lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X \lim_{T\to\infty} T\{X'V^{-1}(\theta)X\}^{-1}$$

$$\times \lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X\Big]$$

$$\times \lim_{T\to\infty} T\{X'V^{-1}(\theta)X\}^{-1} + o(T^{-2}).$$

We remark that

$$\frac{\partial}{\partial\theta}V^{-1}(\theta) = -V^{-1}(\theta)\frac{\partial}{\partial\theta}V(\theta)V^{-1}(\theta)$$

and

$$\frac{\partial^2}{\partial\theta^2}V^{-1}(\theta) = 2V^{-1}(\theta)\frac{\partial}{\partial\theta}V(\theta)V^{-1}(\theta)\frac{\partial}{\partial\theta}V(\theta)V^{-1}(\theta)$$

$$-V^{-1}(\theta)\frac{\partial^2}{\partial\theta^2}V(\theta)V^{-1}(\theta).$$

Then

$$X'V^{-1}(\theta)\frac{\partial}{\partial\theta}V(\theta)V^{-1}(\theta)\frac{\partial}{\partial\theta}V(\theta)V^{-1}(\theta)X$$

$$(3.3)\qquad = \frac{1}{2}\left[X'\frac{\partial^2}{\partial\theta^2}V^{-1}(\theta)X + X'V^{-1}(\theta)\frac{\partial^2}{\partial\theta^2}V(\theta)V^{-1}(\theta)X\right].$$

The limiting behaviour of the first term of (3.3) is evaluated as

$$(3.4)\qquad \lim_{T\to\infty} T^{-1}X'\frac{\partial^2}{\partial\theta^2}V^{-1}(\theta)X$$

$$= \int_{-\pi}^{\pi}\{2f_\theta(\lambda)f_\theta'(\lambda)^2 - f_\theta''(\lambda)f_\theta(\lambda)^2\}f_\theta(\lambda)^{-4}\,dM(\lambda),$$

where $f_\theta''(\lambda) = (\partial^2/\partial\theta^2)f_\theta(\lambda)$ and $f_\theta'(\lambda) = (\partial/\partial\theta)f_\theta(\lambda)$. This expression can be simplified when we assume that $M(\lambda)$ has only one jump point at $\lambda = 0$ for which the jump is $\Delta M(0) = R$. Then (3.4) is

$$-f_\theta''(0)f_\theta(0)^{-2}\Delta M(0) + 2f_\theta'(0)^2 f_\theta(0)^{-3}\Delta M(0).$$

On the other hand,

$$\lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X \lim_{T\to\infty} T\{X'V^{-1}(\theta)X\}^{-1} \lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X$$

$$(3.5)\quad \begin{aligned} &= \left\{\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi} f_\theta(\lambda)^{-1}\,dM(\lambda)\right\}\left\{\int_{-\pi}^{\pi} f_\theta(\lambda)^{-1}\,dM(\lambda)\right\}^{-1} \\ &\quad \times\left\{\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi} f_\theta(\lambda)^{-1}\,dM(\lambda)\right\} \\ &= f_\theta'(0)^2 f_\theta(0)^{-3}\Delta M(0)\end{aligned}$$

under the condition $M(\lambda)$ has only one jump point at $\lambda = 0$. Therefore, the term in square brackets in (3.2) is, from (3.4) and (3.5),

$$\begin{aligned} &\lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)V(\theta)\frac{\partial}{\partial\theta}V^{-1}(\theta)X \\ &\quad - \lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X \lim_{T\to\infty} T\{X'V^{-1}(\theta)X\}^{-1} \\ &\quad \times \lim_{T\to\infty} T^{-1}X'\frac{\partial}{\partial\theta}V^{-1}(\theta)X \end{aligned}$$

$$(3.6)\quad \begin{aligned} &= -\frac{1}{2}f_\theta''(0)f_\theta(0)^{-2}\Delta M(0) + f_\theta'(0)^2 f_\theta(0)^{-3}\Delta M(0) \\ &\quad + \frac{1}{2}\lim_{T\to\infty} T^{-1}X'V^{-1}(\theta)\frac{\partial^2}{\partial\theta^2}V(\theta)V^{-1}(\theta)X \\ &\quad - f_\theta'(0)^2 f_\theta(0)^{-3}\Delta M(0) \\ &= \frac{1}{2}\left[\lim_{T\to\infty} T^{-1}X'V^{-1}(\theta)\frac{\partial^2}{\partial\theta^2}V(\theta)V^{-1}(\theta)X - f_\theta''(0)f_\theta(0)^{-2}\Delta M(0)\right].\end{aligned}$$

In order to evaluate $\lim_{T\to\infty} T^{-1}X'V^{-1}(\theta)(\partial^2/\partial\theta^2)V(\theta)V^{-1}(\theta)X$, let

$$\begin{aligned} b_{ij}^T(h) &= \sum_{t=1}^{T-h} z_{it}z_{jt+h}, \qquad h = 0,1,\dots \\ &= \sum_{t=1+h}^{T} z_{it}z_{jt+h}, \qquad h = 0,-1,\dots,\end{aligned}$$

with $Z = V^{-1}(\theta)X$. We assume the following:

S.1 $b_{ii}^T(0) = \|z_i\|_T^2 \to \infty$ as $T\to\infty$.
S.2 $\lim_{T\to\infty} z_{iT+1}^2/b_{ii}^T(0) = 0$ for $i = 1,\dots,p$.
S.3 The limit of

$$b_{ij}^T(h)/T = q_{\theta ij}^T(h) \quad \text{as } T\to\infty$$

exists for every $i, j = 1,\dots,p$ and $h = 0,\pm1,\dots$. Let

$$\lim_{T\to\infty} q_{\theta ij}^T(h) = q_{\theta ij}(h)$$

for $i, j = 1,\dots,p$ and $h = 0,\pm1,\dots$ and let $Q_\theta(h) = [q_{\theta ij}(h)]$.

S.4 $Q_\theta(0)$ is nonsingular.

Then there exists another regression spectral measure $N_\theta(\lambda)$ such that

$$Q_\theta(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dN_\theta(\lambda).$$

So

$$\lim_{T\to\infty} T^{-1}X'V^{-1}(\theta)\frac{\partial^2}{\partial\theta^2}V(\theta)V^{-1}(\theta)X = \int_{-\pi}^{\pi} f_\theta''(\lambda)\, dN_\theta(\lambda). \tag{3.7}$$

Therefore, we obtain the integral representation such as

THEOREM 3.1. *If the conditions* R.1–R.4 *and* S.1–S.4 *hold, then as* $T \to \infty$

$$\begin{aligned} &E[(\hat\beta_{\hat W} - \hat\beta_W)(\hat\beta_{\hat W} - \hat\beta_W)'] \\ &= T^{-2}w^{-1}\left\{\int_{\pi}^{\pi} f_\theta(\lambda)^{-1}\, dM(\lambda)\right\}^{-1} \\ &\quad\times\Bigg[\int_{-\pi}^{\pi} f_\theta'(\lambda)^2 f_\theta(\lambda)^{-3}\, dM(\lambda) - \tfrac{1}{2}\int_{-\pi}^{\pi} f_\theta''(\lambda) f_\theta(\lambda)^{-2}\, dM(\lambda) \\ &\qquad + \tfrac{1}{2}\int_{-\pi}^{\pi} f_\theta''(\lambda)\, dN_\theta(\lambda) - \left\{\int_{-\pi}^{\pi} f_\theta'(\lambda) f_\theta(\lambda)^{-2}\, dM(\lambda)\right\} \\ &\qquad \times\left\{\int_{-\pi}^{\pi} f_\theta(\lambda)^{-1}\, dM(\lambda)\right\}^{-1}\left\{\int_{-\pi}^{\pi} f_\theta'(\lambda) f_\theta(\lambda)^{-2}\, dM(\lambda)\right\}\Bigg] \\ &\quad\times\left\{\int_{-\pi}^{\pi} f_\theta(\lambda)^{-1}\, dM(\lambda)\right\}^{-1} + o(1/T^2). \end{aligned} \tag{3.8}$$

If the regression function $\{x_t\}$ satisfies the condition that $M(\lambda)$ jumps at $\lambda = 0$ only, we obtain

THEOREM 3.2. *As* $T \to \infty$,

$$\begin{aligned} &E[(\hat\beta_{\hat W} - \hat\beta_W)(\hat\beta_{\hat W} - \hat\beta_W)'] \\ &= T^{-2}w^{-1}f_\theta(0)^{-2}\,\Delta M(0)^{-1}2^{-1} \\ &\quad\times\left[\int_{-\pi}^{\pi} f_\theta''(\lambda)\, dN_\theta(\lambda) - f_\theta''(0) f_\theta(0)^{-2}\,\Delta M(0)\right]\Delta M(0)^{-1} + o(1/T^2) \end{aligned}$$

under the condition that $M(\lambda)$ *has only one jump point at* $\lambda = 0$.

On the other hand,

LEMMA 3.4. *If* $N_\theta(\lambda)$ *has only one jump point at* $\lambda = 0$ *for which the jump is* $f_\theta(0)^{-2}R$, *then* (3.7) *is*

$$f_\theta''(0) f_\theta(0)^{-2} R.$$

Moreover,

$$Q_\theta(0) = f_\theta(0)^{-2} R.$$

By using this lemma, we obtain

THEOREM 3.3. *Under the conditions* R.1–R.4 *and* S.1–S.4, *if* $M(\lambda)$ *and* $N_\theta(\lambda)$ *each have only one jump point at* $\lambda = 0$, *where the jumps are* R *and* $f_\theta(0)^{-2}R$, *respectively, then the coefficient of* T^{-2} *in* (3.2) *vanishes.*

COROLLARY 3.1. *Under the same conditions as in Theorem* 3.2,

$$E\left[(\hat{\beta}_{\hat{W}} - \beta)(\hat{\beta}_{\hat{W}} - \beta)'\right] - \operatorname{Cov}(\hat{\beta}_W) = o(1/T^2) \quad \text{as } T \to \infty.$$

REMARK 1. Since $y - X\hat{\beta}_W = [I - X\{X'V^{-1}(\theta)X\}^{-1}X'V^{-1}(\theta)]\tilde{u}$, which depends on y only through $\tilde{u}$, $\hat{\theta}_{\mathrm{MLE}}$ is a function of $\tilde{u}$ only. So $\hat{\beta}_{\mathrm{MLE}}$ is a GLSE even in the present situation and $\hat{\theta}_{\mathrm{MLE}}$ has the same asymptotic distribution as $\hat{\theta}$. So (3.2) for $\hat{\beta}_{\mathrm{MLE}}$ is identical to that for $\hat{\beta}_{\hat{W}}$.

REMARK 2. From the proof of Theorem 3.1, the leading term of $E[(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)(\hat{\beta}_{\hat{W}} - \hat{\beta}_W)']$ has the same expansion (3.8) whenever $\hat{\theta}$ is an estimator of a class of best asymptotically normal estimators. This point is also discussed in the point of the prediction framework (see Toyooka (1982)).

REMARK 3. Under the conditions of Theorem 3.2, there is no difference between the asymptotic value of the covariance matrix of $\hat{\beta}$ and that of $\hat{\beta}_W$ which is equivalent to that of $\hat{\beta}_{\hat{W}}$. This fact is a special statement of Grenander's (1954) result. Moreover, under the conditions of Theorem 3.2, there is no difference between the asymptotic value of the covariance matrix of $\hat{\beta}_W$ and that of $\hat{\beta}_{\hat{W}}$ up to $O(T^{-2})$. Of course both matrices are smaller than that of $\hat{\beta}$ up to $O(T^{-2})$ (see Toyooka (1985)).

4. Implications of Theorems 3.2 and 3.3. Under the conditions of Theorem 3.3, we can compare the risk matrix of $\hat{\beta}$ with that of $\hat{\beta}_{\hat{W}}$ or $\hat{\beta}_{\mathrm{MLE}}$ as stated in Toyooka (1985).

In the first-order autoregression with the autoregressive parameter θ, after simple calculation,

$$\begin{aligned} Q_\theta(h) &= (1 - 4\theta + 6\theta^2 - 4\theta^3 + \theta^4)R \\ &= \int_{-\pi}^{\pi} e^{i\lambda h}\, dN_\theta(\lambda), \end{aligned}$$

which does not depend on h. Therefore, $N_\theta(\lambda)$ has only one jump point at $\lambda = 0$ and the jump is

$$f_\theta(0)^{-2}R = (1 - 4\theta + 6\theta^2 - 4\theta^3 + \theta^4)R.$$

So the structure of the autoregression automatically satisfies the condition for $N_\theta(\lambda)$.

The case in which the error $\{u_t\}$ is a second-order autoregression is a special case of our model (1.2). Pantula and Fuller (1985) compare the empirical risk matrices of two estimated generalized least squares estimators for a linear trend model with second-order autoregressive error in a Monte Carlo experiment. The estimator $\hat{\beta}_{\hat{W}_3}$ is based on autoregressive parameters estimated by using the ordinary least squares residuals and $\hat{\beta}_{\hat{W}_4}$ is based on a bias adjusted estimator for the autoregressive parameters. Their experimental results agree with our theory in that there is little difference between the risk matrix of $\hat{\beta}_{\hat{W}_3}$ and that of $\hat{\beta}_{\hat{W}_4}$ in the most cases. They find that the bias adjustment procedure for the autoregressive estimator is effective in small samples ($n = 25$) for processes with a large positive root. Our theory says that the effect of the bias adjustment procedure in the autoregressive parameters exists in the $o(T^{-2})$ term of expansion (3.8). Their experiment indicates that the high-order asymptotic expansion provides better approximation to the small sample behavior of the estimator in the interior of the parameter space.

From the results of Toyooka (1983), (1985) and the present results, the following facts were elucidated. In the case where the regression function $\{x_t\}$ does not satisfy the Grenander condition that $M(\lambda)$ increases at not more than p values of λ, $0 < \lambda < \pi$, and the sum of the ranks of the increases in $M(\lambda)$ is p,

$$\operatorname{Cov}(\hat{\beta}) - \operatorname{Cov}(\hat{\beta}_W) = O(T^{-1})$$

and

$$\operatorname{Cov}(\hat{\beta}_{\hat{W}}) - \operatorname{Cov}(\hat{\beta}_W) = O(T^{-2}).$$

On the other hand, under Grenander's condition,

$$\operatorname{Cov}(\hat{\beta}) - \operatorname{Cov}(\hat{\beta}_W) = O(T^{-2}),$$

and, moreover, if $M(\lambda)$ and $N_\theta(\lambda)$ satisfy the condition of Theorem 3.2,

$$\operatorname{Cov}(\hat{\beta}_{\hat{W}}) - \operatorname{Cov}(\hat{\beta}_W) = o(T^{-2}).$$

It is not obvious whether the last equality is $O(T^{-3})$ or not. An extension to the case where θ is multidimensional is straightforward and therefore is omitted.

Acknowledgments. The author would like to express his hearty thanks to Professor W. A. Fuller of Iowa State University for insightful discussions. He would like to express his thanks to the referees and Associate Editor for improving this paper.

REFERENCES

GRENANDER, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

HILDRETH, C. (1969). Asymptotic distribution of maximum likelihood estimation in a linear model with autoregressive disturbances. *Ann. Math. Statist.* **40** 583–594.

KARIYA, T. and TOYOOKA, Y. (1985). Nonlinear versions of the Gauss–Markov theorem and GLSE. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 345–354. Elsevier, New York.

MITRA, S. K. and RAO, C. R. (1969). Conditions for optimality and validity of simple least squares theory. *Ann. Math. Statist.* **40** 1617–1624.

PANTULA, S. G. and FULLER, W. A. (1985). Mean estimation bias in least squares estimation of autoregressive processes. *J. Econometrics* **27** 99–121.

TOYOOKA, Y. (1982). Prediction error in a linear model with estimated parameters. *Biometrika* **69** 453–459.

TOYOOKA, Y. (1983). Second-order structure of mean squared error of generalized least squares estimator adjusted by estimated residuals in a location model with autoregressive error. *Math. Japon.* **28** 163–171.

TOYOOKA, Y. (1985). Second-order risk comparisons of SLSE with GLSE and MLE in regression with serial correlation. *J. Multivariate Anal.* **17** 107–126.

WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary nondeterministic time-series. *J. Aust. Math. Soc.* **4** 363–384.

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
TOYONAKA, OSAKA 560
JAPAN