Acta Math. Univ. Comenianae
Vol. LXI, 2(1992), pp. 251–261

MEAN SQUARE ERROR MATRIX OF AN APPROXIMATE LEAST SQUARES ESTIMATOR IN A NONLINEAR REGRESSION MODEL WITH CORRELATED ERRORS

F. ŠTULAJTER

Abstract. A nonlinear regression model with correlated, normally distributed errors is investigated. The bias and the mean square error matrix of the approximate least squares estimator of regression parameters are derived and their limit properties are studied.

1. Introduction

Let us consider a nonlinear regression model

$$Y_t = f(x_t, \theta) + \varepsilon_t; \quad t = 1, \dots, n$$

where f is a model function, x_t; $t = 1, \dots, n$ are assumed to be known k dimensional vectors, $\theta = (\theta_1, \dots, \theta_p)'$ is an unknown vector of regression parameters which belongs to some open set Θ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ is some random vector of errors with zero mean value. Next we assume that the functions $f(x_t; \cdot)$ have for every fixed t continuous derivatives

$$\frac{\partial^2 f(x_t, \theta)}{\partial\theta_i \partial\theta_j} = \frac{\partial^2 f(x_t, \theta)}{\partial\theta_j \partial\theta_i} \qquad \text{for all } i, j = 1, 2, \dots, p.$$

Let us denote by $\hat\theta$ the least squares estimator of θ: that means:

$$\hat\theta = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} (Y_t - f(x_t, \theta))^2$$

We shall assume that this estimator exists and is unique. In this connection see Pázman (1984a).

It is well known that $\hat\theta$ is a biased estimator. The covariance or the mean square error matrix of $\hat\theta$ was derived by Clarke (1980) using a stochastic expansion for $\hat\theta$

Received March 25, 1992.
1980 *Mathematics Subject Classification* (1991 *Revision*). Primary 62J02.
Key words and phrases. Nonlinear regression, appproximate least squares estimator, bias, mean square error matrix.

and a suitable transformation of the vector of uncorrelated errors with a common variance σ^2. A connection between measures of nonlinearity and members of the covariance matrix can be seen from his results. The approximate distribution of $\hat{\theta}$ was given by Pázman (1984b).

In most papers devoted to the problems of nonlinear regression, it is assumed that the errors are independent identically distributed random variables. The problems with auto correlated errors were studied by Gallant and Goebel (1976) and Gallant (1987) using a strong theory of martingales and mixingales.

The aim of this article is to give a direct expression for the mean square error matrix of the approximate $\tilde{\theta}$ of θ without using any transformation of the vector of errors and assuming that the errors are normally distributed with zero mean value and a covariance matrix Σ. Under these conditions and conditions imposed on the nonlinear model function f, the limit properties of the bias and covariance matrix of $\tilde{\theta}$ are studied.

The approximate least squares estimator $\tilde{\theta}$ is derived on the idea which was used by Box (1971) for derivation of an approximate bias of $\hat{\theta}$. Let us denote by $f(\theta)$ the $n\times 1$ vector $(f(x_1,\theta),\dots,f(x_n,\theta))'$ and let $\boldsymbol{j}_t(\theta)$ be the $p\times 1$ vector with components

$$\frac{\partial f(x_t,\theta)}{\partial\theta_i}; \quad i=1,\dots,p, \quad t=1,2,\dots,n\,.$$

Let

$$J(\theta)=\begin{pmatrix} \boldsymbol{j}_1'(\theta)\\ \vdots \\ \boldsymbol{j}_n'(\theta)\end{pmatrix}$$

be the $n\times p$ matrix of the first derivatives of $f(\theta)$.

Let H_t; $t=1,2,\dots,n$ be the $p\times p$ matrices of second derivatives with

$$(H_t)_{ij}=\frac{\partial^2 f(x_t,\theta)}{\partial\theta_i\partial\theta_j}; \qquad i,j=1,\dots,p.$$

Then, since $\hat{\theta}$ is the least squares estimator of θ, the equality

$$\sum_{t=1}^{n}(Y_t-f(x_t,\theta))\,\frac{\partial f(x_t,\theta)}{\partial\theta_i}\Big|_{\theta=\hat{\theta}}=0; \qquad i=1,2,\dots,p$$

must hold. This equality can be written, denoting $Y=(Y_1,\dots,Y_n)'$, in the following matrix form

$$J(\hat{\theta})'(Y-f(\hat{\theta}))=0\,. \tag{1}$$

2. An Approximate Least Squares Estimator And Its Bias

It was shown in Box (1971), using (1) and Taylor expansions of $J(\theta)$ and $f(\theta)$ that the LSE $\hat{\theta}$ of θ can be approximated by the estimator $\tilde{\theta}$ given by

$$\tilde{\theta} = \theta + (J'J)^{-1}J'\varepsilon + (J'J)^{-1}\left[U'(\varepsilon)M\varepsilon - 1/2J'H(\varepsilon)\right], \tag{2}$$

where $J = J(\theta)$, $M = I - J(J'J)^{-1}J$, $U(\varepsilon)$ denotes the $n \times p$ random matrix,

$$U(\varepsilon) = \begin{pmatrix} \varepsilon' A' H_1 \\ \vdots \\ \varepsilon' A' H_n \end{pmatrix},$$

where $A = (J'J)^{-1}J$ and $H(\varepsilon)$ is the $n \times 1$ random vector with components $\varepsilon' A' H_t A\varepsilon$; $t = 1, \ldots, n$.

Using these results, it was shown by Box (1971) that if ε_t; $t = 1, 2, \ldots, n$ are i.i.d. random variables with zero mean and with a variance σ^2, then

$$E_\theta[\tilde{\theta}] = \theta - \frac{\sigma^2}{2}(J'J)^{-1}J' \begin{pmatrix} \operatorname{tr}((J'J)^{-1}H_1) \\ \vdots \\ \operatorname{tr}((J'J)^{-1}H_n) \end{pmatrix}$$

Now we shall assume that $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is a random vector with zero mean value and with some covariance matrix Σ. Then we have:

$$E_\theta[\tilde{\theta}] = \theta + (J'J)^{-1}E_\theta\left[U'(\varepsilon)M\varepsilon\right] - 1/2(J'J)^{-1}J'E_\theta\left[H(\varepsilon)\right].$$

$U'(\varepsilon)M\varepsilon$ can by written as

$$U'(\varepsilon)M\varepsilon = (H_1 A\varepsilon, \ldots, H_n A\varepsilon)M\varepsilon$$

and for the j-th component of this random vector we get:

$$\begin{aligned}(U'(\varepsilon)M\varepsilon)_j &= \sum_{i=1}^{n}(U'(\varepsilon))_{ji}(M\varepsilon)_i = \sum_{k=1}^{n}\sum_{l=1}^{n}\left(\sum_{i=1}^{n}(H_iA)_{jk}M_{il}\right)\varepsilon_k\varepsilon_l \\ &= \varepsilon' N_j \varepsilon; \quad j = 1, 2, \ldots, p, \qquad \text{where}\end{aligned}$$

$$(N_j)_{kl} = \sum_{i=1}^{n}(H_iA)_{jk}M_{il}; \quad k, l = 1, 2, \ldots, n. \tag{3}$$

We can also write

$$(U'(\varepsilon)M\varepsilon)_j = \varepsilon'\left(\frac{N_j + N_j'}{2}\right)\varepsilon : \qquad j = 1, 2, \ldots, n$$

as quadratic forms with symmetric matrices.

Using the equality $E[\varepsilon' C\varepsilon] = \operatorname{tr}(C\Sigma)$, which holds for any matrix C and for any random vector ε with mean value zero and a covariance matrix Σ, we get

$$E_\theta[\tilde{\theta}] = \theta + (J'J)^{-1}\left[\begin{pmatrix}\operatorname{tr}(N_1\Sigma)\\ \vdots \\ \operatorname{tr}(N_p\Sigma)\end{pmatrix} - \frac{1}{2}J'\begin{pmatrix}\operatorname{tr}(A'H_1A\Sigma)\\ \vdots \\ \operatorname{tr}(A'H_nA\Sigma)\end{pmatrix}\right], \tag{4}$$

where N_j; $j = 1, 2, \ldots, p$ are given by (3).

In the special case $\Sigma = \sigma^2 I$ of uncorrelated errors we get

$$\operatorname{tr}(N_j\Sigma) = \sigma^2 \operatorname{tr}(N_j) = \sigma^2 \sum_{i=1}^{n} (H_i AM)_{ji} = 0,$$

since $AM = 0$ and

$$\operatorname{tr}(A'H_jA\Sigma) = \sigma^2 \operatorname{tr}(AA'H_j) = \sigma^2 \operatorname{tr}((J'J)^{-1}H_j)$$

and we see that (4) agrees with the bias of $\tilde{\theta}$ given by Box (1979) for uncorrelated errors.

Now we shall study the limit properties of the bias given by (4) of the approximate LSE $\tilde{\theta}$.

Since $\operatorname{tr}(AB') = \sum_{i,j=1}^{n} A_{ij}B_{ij}$ is an inner product in the space of square matrices, we can write $|\operatorname{tr}(AB')| \le \|A\|\,\|B\|$, where $\|A\| = \left(\sum_{i,j=1}^{n} A_{ij}^2\right)^{1/2}$ is the Euclidean norm of a matrix A, for which the inequality $\|AB\|^2 \le \|A\|^2\|B\|^2$ holds. Thus we can write:

$$|\operatorname{tr}(N\Sigma)| \le \|N\|\,\|\Sigma\| \qquad \text{and} \tag{5}$$

$$\begin{aligned}\|N_j\|^2 &= \sum_{k,l=1}^{n}\left(\sum_{i=1}^{n}(H_iA)_{jk}M_{il}\right)\left(\sum_{s=1}^{n}(H_sA)_{jk}M_{sl}\right)\\ &= \sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{n}(H_iA)_{jk}(H_sA)_{jk}M_{is} = \sum_{s=1}^{n}\sum_{i=1}^{n}(H_iAA'H_s)_{jj}M_{is}\\ &= \sum_{i=1}^{n}(H_i(J'J)^{-1}H_i)_{jj} - \sum_{s=1}^{n}\sum_{i=1}^{n}(H_i(J'J)^{-1}H_s)_{jj}\boldsymbol{j}_i'(J'J)^{-1}\boldsymbol{j}_s,\end{aligned} \tag{6}$$

since $M = M' = I - P = M^2$.

Now, for the j-th component of the second term of the bias, we have:

$$\left(J'\begin{pmatrix}\operatorname{tr}(A'H_1A\Sigma)\\ \vdots \\ \operatorname{tr}(A'H_nA\Sigma)\end{pmatrix}\right)_j = \sum_{i=1}^{n}(\boldsymbol{j}_i)_j \operatorname{tr}(A'H_iA\Sigma) = \operatorname{tr}(A'\sum_{i=1}^{n}(\boldsymbol{j}_i)_j H_iA\Sigma),$$

since $J' = (\boldsymbol{j}_1, \ldots, \boldsymbol{j}_n)$.

Let us denote by $B_j = \sum_{i=1}^n (\boldsymbol{j}_i)_j H_i$, for simplicity. Then we have:

$$|\operatorname{tr}(A'B_jA\Sigma)| \le \|A'B_jA\|\,\|\Sigma\| = \left[\operatorname{tr}(B_j(J'J)^{-1}B_j(J'J)^{-1})\right]^{1/2} \cdot \|\Sigma\|. \tag{7}$$

The limit properties of the estimator $\tilde{\theta}$ are based on the following assumptions.

Assumption 1. The matrix $(J'J)^{-1}$ is of the order $\frac{1}{n}$ (we write $(J'J)^{-1} = O_G(\frac{1}{n})$) by which we mean that $(J'J)^{-1} = \frac{1}{n}G_n$ and there exists a nonnegative definite matrix G such that $\lim_{n\to\infty} G_n = G$.

Assumption 2. The following limits

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \frac{\partial f(x_t,\theta)}{\partial \theta_i} \cdot \frac{\partial^2 f(x_t,\theta)}{\partial \theta_j \partial \theta_k} \qquad \text{and}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 f(x_t,\theta)}{\partial \theta_i \theta_j} \frac{\partial^2 f(x_t,\theta)}{\partial \theta_k \partial \theta_l}$$

exist and are finite for every fixed i, j, k, l.

Theorem 1. *Let the Assumptions* 1 *and* 2 *hold and let for the covariance matrix* Σ *of the vector* ε *of errors* $\lim_{n\to\infty} \frac{1}{n}\|\Sigma\| = 0$. *Then for the bias of the approximate least squares estimator* $\tilde{\theta}_n$ *we have:*

$$\lim_{n\to\infty} E_\theta[\tilde{\theta}_n] = \theta.$$

Proof. It is a direct consequence of (5), (6), (7), and the assumptions of Theorem 1, that there exist finite limits

$$\lim_{n\to\infty} \|N_j\| \quad \text{and} \quad \lim_{n\to\infty} \|A'B_jA\|.$$

$\square$

Remarks.

1. In the case when ε_i; $i = 1, 2, \ldots, n$ are i.i.d. random variables with $E[\varepsilon_i] = 0$ and $D[\varepsilon_i] = \sigma^2$ we have

$$\lim_{n\to\infty} \frac{1}{n}\|\Sigma\| = \lim_{n\to\infty} \frac{\sigma^2}{n^{1/2}} = 0$$

and the condition for the vector of errors in theorem is fulfilled.

2. If $\{\varepsilon_t; \quad t = 1, 2, \ldots\}$ is a stationary time series with a covariance function $R(\cdot)$ such that $\lim_{t\to\infty} R(t) = 0$, then

$$\frac{1}{n}\|\Sigma\| = \left(\frac{R^2(0)}{n} + \frac{2}{n}\sum^n \left(1 - \frac{t}{n}\right) R^2(t)\right)^{1/2}, \quad \lim_{n\to\infty} \frac{1}{n}\,\|\Sigma\| = 0$$

and the condition of theorem is fulfilled.

3. The Mean Square Error Matrix Of The Approximate Least Squares Estimator

We have shown in the preceding part that

$$\tilde{\theta} = \theta + A\varepsilon + (J'J)^{-1}\left[\begin{pmatrix}(\varepsilon' N_1 \varepsilon)\\ \vdots \\ (\varepsilon' N_p \varepsilon)\end{pmatrix} - \frac{1}{2}J'\begin{pmatrix}(\varepsilon' A' H_1 A\varepsilon)\\ \vdots \\ (\varepsilon' A' H_n A\varepsilon)\end{pmatrix}\right] \tag{8}$$

where $A = (J'J)^{-1}J'$ and the matrices N_j are given by (3).

Let $N(\varepsilon)$ be the $p \times 1$ random vector

$$N(\varepsilon) = \left(\varepsilon' \frac{N_1 + N_1'}{2}\varepsilon, \dots, \varepsilon' \frac{N_p + N_p'}{2}\varepsilon\right)'$$

and $H(\varepsilon)$ be the $n \times 1$ random vector

$$H(\varepsilon) = (\varepsilon' A' H_1 A\varepsilon, \dots, \varepsilon' A' H_n A\varepsilon)' .$$

Then we can write:

$$\begin{aligned} E_\theta[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)'] &= AE_\theta[\varepsilon\varepsilon']A' + (J'J)^{-1}E_\theta[(N(\varepsilon) \\ &\qquad - \frac{1}{2}J'H(\varepsilon))(N(\varepsilon) - \frac{1}{2}J'H(\varepsilon))'](J'J)^{-1} \\ &= A\Sigma A' + (J'J)^{-1}\Big\{E_\theta[N(\varepsilon)N(\varepsilon)'] - \frac{1}{2}E_\theta[N(\varepsilon)H(\varepsilon)]J \\ &\qquad - \frac{1}{2}J'E_\theta[H(\varepsilon)N(\varepsilon)'] + \frac{1}{4}J'E_\theta[H(\varepsilon)H(\varepsilon)']J\Big\}(J'J)^{-1}, \end{aligned} \tag{9}$$

assuming that the vector ε of errors is such that all its third moments are equal to zero. This condition is fulfilled also for the case when ε has the $N_n(0, \Sigma)$ distribution, what we shall assume in the sequel. In this case we can use the following known formula

$$E[\varepsilon' B\varepsilon \cdot \varepsilon' C\varepsilon] = 2\,\mathrm{tr}\,(B\Sigma C\Sigma) + \mathrm{tr}\,(B\Sigma)\,\mathrm{tr}\,(C\Sigma)$$

which holds for any symmetric matrices B and C and any normally distributed random vector ε. According to this we can write

$$\begin{aligned} S(1,1)_{ij} &= (E[N(\varepsilon)N(\varepsilon)'])_{ij} = 2\,\mathrm{tr}\left(\frac{N_i + N_i'}{2}\Sigma\frac{N_j + N_j'}{2}\Sigma\right) + \mathrm{tr}(N_i\Sigma)\,\mathrm{tr}(N_j\Sigma) \\ S(1,2)_{ij} &= (E[N(\varepsilon)H\varepsilon)'])_{ij} = 2\,\mathrm{tr}\left(\frac{N_i + N_i'}{2}\Sigma A' H_j A\Sigma\right) + \mathrm{tr}(N_i\Sigma)\,\mathrm{tr}(A' H_j A\Sigma) \\ S(2,1) &= E[H(\varepsilon)N(\varepsilon)'] = S(1,2)' \qquad \text{and} \\ S(2,2)_{ij} &= (E[H(\varepsilon)H\varepsilon)'])_{ij} = 2\,\mathrm{tr}\,(A' H_i A\Sigma A' H_j A\Sigma) + \mathrm{tr}(A' H_j A\Sigma). \end{aligned}$$

Let us consider now the case when $\Sigma = \sigma^2 I$.

Then we have, using the same algebra as in (6):

$$S(1,1)_{ij} = \sigma^4 \operatorname{tr}(N_i N_j') = \sigma \sum_{k,l=1}^{n} \left(H_k (J'J)^{-1} H_l\right)_{ij} M_{kl},$$

since we can easily get from (3) and from the equality $AM = 0$ that

$$\operatorname{tr}(N_i) = \operatorname{tr}(N_i N_j) = 0.$$

Next, $S(1,2) = 0$, since

$$S(1,2)_{ij} = \sigma^4 \operatorname{tr}\left(\frac{N_i + N_i'}{2} A' H_j A\right) = \sigma^4 \operatorname{tr}(A' H_j A N_i) = 0 \qquad \text{for all } i, j.$$

For $S(2,2)$ we get, using $AA' = (J'J)^{-1}$

$$\begin{aligned} S(2,2)_{ij} &= 2\sigma^4 \operatorname{tr}(A' H_i A A' H_j A) + \sigma^4 \operatorname{tr}(A' H_i A) \operatorname{tr}(A' H_j A) \\ &= 2\sigma^4 \operatorname{tr}(H_i (J'J)^{-1} H_j (J'J)^{-1}) + \sigma^4 \operatorname{tr}(H_i (J'J)^{-1}) \operatorname{tr}(H_j (J'J)^{-1}). \end{aligned}$$

Using these results we can get the mean square error matrix for the approximate least squares estimator $\tilde{\theta}$ as follows:

$$\begin{aligned} E_\theta[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)'] = \sigma^2 (J'J)^{-1} + \sigma^4 (J'J)^{-1} \Bigg[& \sum_{i,j=1}^{n} M_{ij} H_i (J'J)^{-1} H_j \\ & + \frac{1}{4} \sum_{i,j=1}^{n} \Big(2 \operatorname{tr}(H_i (J'J)^{-1} H_j (J'J)^{-1}) \\ & + \operatorname{tr}(H_i (J' J^{-1}) \operatorname{tr}(H_j (J'J)^{-1}) \Big) \boldsymbol{j}_i \boldsymbol{j}_j' \Bigg] (J'J)^{-1} \end{aligned} \tag{10}$$

We can prove the following theorem.

Theorem 2. *Let the Assumptions* 1 *and* 2 *hold and let* ε *has the* $N_n(0, \sigma^2 I)$ *distribution. Then the mean square error matrix* $E_\theta[(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)']$ *is given by* (10) *and we have*

$$\lim_{n \to \infty} n E_\theta[(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)'] = \sigma^2 G.$$

Proof. The theorem will be proved if we show that

$$\lim_{n \to \infty} \sum_{i,j=1}^{n} M_{ij} H_i (J'J)^{-1} H_j (J'J)^{-1} = 0.$$

But, since $M_{ij} = \delta_{ij} - \boldsymbol{j}_i'(J'J)^{-1}\boldsymbol{j}_j$, we can write

$$\lim_{n\to\infty} \sum_{i,j=1}^{n} M_{ij} H_i (J'J)^{-1} H_j (J'J)^{-1} = \lim_{n\to\infty} \left(\sum_{i=1}^{n} H_i (J'J)^{-1} H_i (J'J)^{-1} - \sum_{i,j=1}^{n} \boldsymbol{j}'_i (J'J)^{-1} \boldsymbol{j}_j H_i (J'J)^{-1} H_j (J'J)^{-1} \right) = 0$$

and the proof is complete. □

Let us consider now the case when ε is $N_n(0,\Sigma)$ distributed random vector. Then we have the inequalities

$$\begin{aligned} &\|A\Sigma A'\|^2 \le \|\Sigma\|^2 \|AA'\|^2 = \|\Sigma\|^2 \|(J'J)^{-1}\|^2, \\ &\left| \operatorname{tr}\left(\frac{N_i + N_i'}{2} \Sigma \frac{N_j + N_j'}{2} \Sigma \right) \right| \le \left\| \frac{N_i + N_i'}{2} \right\| \cdot \|\Sigma\|^2 \cdot \left\| \frac{N_j + N_j'}{2} \right\| \\ &\qquad \le \|N_i\| \cdot \|N_j\| \, \|\Sigma\|^2. \end{aligned} \tag{11}$$

Thus

$$|S(1,1)_{ij}| \le 3\|N_i\| \, \|N_j\| \, \|\sigma\|^2, \tag{12}$$

where the expression for $\|N_j\|^2$ is given by (6). By analogy

$$\begin{aligned} |(S(1,2))_{il}| &= \left| \sum_{j=1}^{n} 2 \operatorname{tr}(N_i \Sigma A' H_j A \Sigma)(\boldsymbol{j}_j)_l + \operatorname{tr}(N_i\Sigma) \operatorname{tr}(A' B_l A \Sigma) \right| \\ &\le 3\|N_i\| \, \|A' B_l A\| \, \|\Sigma\|^2. \end{aligned} \tag{13}$$

It is easy to show that

$$\|A' B_l A\|^2 = \operatorname{tr}((J'J)^{-1} B_l (J'J)^{-1} B_l). \tag{14}$$

By analogy

$$\begin{aligned} (J'S(2,2)J)_{kl} &= \left((\boldsymbol{j}_1, \dots, \boldsymbol{j}_n) S(2,2) \begin{pmatrix} \boldsymbol{j}_1' \\ \vdots \\ \boldsymbol{j}_n' \end{pmatrix} \right)_{kl} \\ &= \sum_{i,j=1}^{n} (\boldsymbol{j}_i)_k \left(2 \operatorname{tr}(A' H_i A \Sigma A' H_j A \Sigma) + \operatorname{tr}(A' H_i A \Sigma) \operatorname{tr}(A' H_j A \Sigma) \right) (\boldsymbol{j}_j)_l) \\ &= 2 \operatorname{tr}(A' B_k A \Sigma A' B_l A \Sigma) + \operatorname{tr}(A' B_k A \Sigma) \operatorname{tr}(A' B_l A \Sigma) \quad \text{for } k,l = 1,\dots,p. \end{aligned}$$

Thus we have:

$$|(J'S(2,2)J)_{kl}| \le 3\|A' B_k A\| \, \|A' B_l A\| \, \|\Sigma\|^2. \tag{15}$$

From the inequalities (12), (13), (15) and from the equalities (6), (9) and (14) the following theorem follows easily.

Theorem 3. *Let the Assumptions* 1 *and* 2 *hold and let* ε *has the* $N_n(0,\Sigma)$ *distribution, where* $\lim_{n\to\infty} 1/n\|\Sigma\| = 0$. *Then for the approximate least squares estimator* $\tilde{\theta}_n$ *given by* (8) *we have:*

$$\lim_{n\to\infty} E[(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)'] = 0.$$

Proof. It follows from (6) and (14) and from the assumptions of the theorem that finite limits $\lim_{n\to\infty} \|N_k\|$ and $\lim_{n\to\infty} \|A'B_kA\|$ exist for every $k = 1,\dots,p$. Thus we see, using (9), (11), (12), (13), (15) and the assumption $(J'J)^{-1} = O_G(1/n)$ that every member of the mean square error matrix of the approximate estimator $\tilde{\theta}_n$ converges to zero if n tends to infinity. □

Remark. According to Remark 2 the condition imposed on Σ is fulfilled if ε is a stationary time series with a covariance function $R(\cdot)$ such that $\lim_{t\to\infty} R(t) = 0$.

4. Simulation Results

Let us consider the nonlinear regression model

$$X(t) = \beta_1 + \beta_2 t + \gamma_1 \cos\lambda_1 t + \gamma_2 \sin\lambda_1 t + \gamma_3 \cos\lambda_2 t + \gamma_4 \sin\lambda_2 t + \varepsilon(t);$$

$t = 1,2,\dots,n$, where $\theta = (\beta_1,\beta_2,\lambda_1,\lambda_2,\gamma_1,\gamma_2,\gamma_3,\gamma_4)'$ is an unknown vector of regression parameters and ε is an $AR(1)$ process given by

$$\varepsilon(t) = \rho\varepsilon(t-1) + e(t)$$

with a white noise e having variance $\sigma^2 = 1$.

We have simulated data following this nonlinear regression model with different values of an autoregression parameter ρ and a given value of θ. For every fixed value of the parameters ρ and θ one observation of X of the length $n = 51$, one of the length $n = 101$ and one of the length $n = 149$ were simulated. The modified Marquard's method was used to compute the LSE $\hat{\theta}$ for θ.

A comparison of the LSE $\hat{\theta}$ and the approximate LSE estimator $\tilde{\theta}$ was done in Štulajter and Hudáková (1991). It was shown that $\hat{\theta}$ and $\tilde{\theta}$ are nearly the same in many cases.

The aim of this simulation study is to investigate an influence of different values of the parameter ρ on the LSE $\hat{\theta}$ and dependence of this influence on n, the length of an observation.

The initial values for iterations were found as follows. First, from $X(\cdot)$ the LSE β^0 for β was found. Then we have computed the periodogram for the partial residuals $X(t) - \beta_1^0 - \beta_2^0 t$; $t = 1,2,\dots,n$ and the frequences λ_1^0, λ_2^0 in which there are the two greatest values of the periodogram were found. In the model

$$X(t) - \beta_1^0 - \beta_2^0 t = \gamma_1 \cos\lambda_1^0 t + \gamma_2 \sin\lambda_1^0 t + \gamma_3 \cos\lambda_2^0 t + \gamma_4 \sin\lambda_2^0 t + \varepsilon(t)$$

we have used again the least squares method for finding γ^0 – the LSE estimator for γ.

The value $\theta^0 = (\beta^{0'}, \lambda^{0'}, \gamma^{0'})'$ of an unknown parameter θ was used as an initial value for computing the LSE $\hat{\theta}$ of θ using the Marquard's method. The least squares estimators, each computed from one simulation of the corresponding length, are given in the following tables.

Least squares estimates

$n = 51$	$\rho =$ -0.99	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	0.99
$\beta_1 = 3$	3.02	3.02	3.09	2.85	2.83	2.98	3.09	3.15	3.23	2.88	0.71
$\beta_2 = 2$	2.01	1.99	1.99	2.01	2.00	2.00	2.00	1.99	1.99	2.00	2.08
$\lambda_1 = 0.75$	0.99	0.75	0.75	0.74	0.75	0.75	0.75	0.75	0.74	0.74	0.74
$\lambda_2 = 0.25$	0.75	0.25	0.24	0.25	0.25	0.25	0.24	0.24	0.24	0.24	0.25
$\gamma_1 = 4$	-1.67	3.73	3.21	4.14	4.19	4.32	3.91	3.94	3.92	3.84	3.78
$\gamma_2 = 3$	-10.59	2.64	3.15	2.38	2.40	2.78	3.02	3.03	2.50	2.88	3.20
$\gamma_3 = 2$	3.6	1.78	2.11	1.51	1.52	2.31	1.88	2.02	2.43	2.47	0.56
$\gamma_4 = 4$	2.77	3.91	4.01	3.83	3.74	3.95	3.66	3.55	3.93	3.54	4.29

Least squares estimates

$n = 101$	$\rho =$ -0.99	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	0.99
$\beta_1 = 3$	3.20	2.98	2.92	3.09	2.94	2.89	3.18	2.95	3.47	4.07	3.78
$\beta_2 = 2$	1.99	2.00	2.00	1.99	2.00	2.00	1.99	2.00	1.99	1.97	1.98
$\lambda_1 = 0.75$	0.99	0.74	0.75	0.75	0.75	0.75	0.74	0.75	0.75	0.74	0.75
$\lambda_2 = 0.25$	0.75	0.25	0.24	0.25	0.25	0.25	0.24	0.25	0.25	0.24	0.25
$\gamma_1 = 4$	-0.70	4.12	3.60	4.00	3.88	3.73	4.13	3.88	3.77	4.04	3.75
$\gamma_2 = 3$	-10.51	3.50	3.52	2.93	3.72	3.29	2.88	3.09	3.24	2.73	3.30
$\gamma_3 = 2$	3.66	2.06	2.15	2.19	2.05	2.35	2.43	2.29	2.19	2.20	1.51
$\gamma_4 = 4$	2.90	3.99	3.86	3.78	3.98	3.74	3.61	3.63	4.01	3.62	4.40

Least squares estimates

$n = 149$	$\rho =$ -0.99	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	0.99
$\beta_1 = 3$	3.04	2.94	2.97	2.98	2.86	3.09	2.95	3.34	3.10	3.20	3.38
$\beta_2 = 2$	1.99	2.00	2.00	2.00	2.00	1.99	2.00	1.99	1.99	1.99	1.99
$\lambda_1 = 0.75$	0.74	0.75	0.75	0.74	0.75	0.75	0.75	0.75	0.75	0.75	0.75
$\lambda_2 = 0.25$	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.24	0.25
$\gamma_1 = 4$	4.22	3.77	4.14	4.31	3.92	3.85	4.18	4.03	4.07	40.3	3.97
$\gamma_2 = 3$	2.68	3.24	2.93	2.80	3.45	2.80	3.01	2.93	3.01	3.12	2.97
$\gamma_3 = 2$	2.05	1.91	1.77	1.91	1.82	2.15	1.95	2.39	2.23	2.67	1.95
$\gamma_4 = 4$	4.01	4.10	3.98	4.00	4.09	3.85	3.83	3.64	3.87	4.06	4.27

We can see from the tables that the only difficulty with estimation is for $\rho = -0.99$, where the influence of the spectral density of $AR(1)$ process on the periodogram occurs. Here $\lambda_1 = 0.75$ is discovered as a second peak of the periodogram and the estimates of corresponding γ's are 3.60 and 2.77 for $n = 51$ and 3.66 and 2.90 for $n = 101$ instead of 4 and 3 respectively. The value $\hat{\lambda}_1 = 0.99$ is due to the spectral density of the $AR(1)$ process and to this frequency correspond also the estimates of γ's. This effect does not occur for $n = 149$. For other values of ρ the LSE $\hat{\theta}$ of θ are satisfactory, as we can see from the tables even for $n = 51$, a relatively small length of observations.

References

Bates D. M. and Watts D. G., *Relative curvature measures of nonlinearity*, J. Roy. Stat. Soc. **B 42** (1980), 1–25.

Box M. J., *Bias in nonlinear estimation*, J. Roy. Stat. Soc. **B 33** (1971), 171–201.

Clarke G. P. Y., *Moments of the least squares estimators in a nonlinear regression model*, J. Roy. Stat. Soc. **B 42** (1980), 227–237.

Cook R. D. and Tsai C. L., *Residuals in nonlinear regressions*, Biometrika **72** (1985), 23–29.

Cook R. D., Tsai C. L. and Wei B. C., *Bias in nonlinear regression*, Biometrika **73** (1986), 615–623.

Gallant A. R. and Goebel J. J., *Nonlinear regression with auto correlated errors*, Jour. of the Amer. Stat. Assoc. **71** (1976), 961–967.

Gallant A. R., *Nonlinear Statistical Modeling*, Wiley, New York, 1987.

Hougaard R., *Parameterization of non-linear models*, J. Roy. Stat. Soc. **B 44** (1982), 244–252.

———, *The appropriateness of the asymptotic distribution in nonlinear regression model in relation to curvature*, J. Roy. Stat. Soc. **B 47** (1985), 103–114.

Jennrich R. I., *Asymptotic properties of nonlinear least squares estimators*, Ann. Math. Stat. **40** (1969), 633–643.

Pázman A., *Nonlinear least squares – Uniqueness versus Ambiguity*, Math. Oper. f. Stat. **15** (1984a), 323–336.

———, *Probability distribution of the multivariate nonlinear least squares estimates*, Kybernetika **20** (1984b), 209–230.

Rattkowsky D. A., *Nonlinear regression modeling*, Marcel Deer, Inc., New York, 1983.

Schmidt W. H. and Zwanzig S., *Second order asymptotics in nonlinear regression*, J. Multivar. Anal. **18** (1986), 187–215.

Štulajter F. and Hadáková J., *An approximate least squares estimator in nonlinear regression and some of its properties*, Submitted for publication..

Wu C. F., *Asymptotic theory of nonlinear least squares estimators*, Ann. Stat. **9** (1981), 501–513.

F. Štulajter, Department of Probability and Statistics, Faculty of Mathematics and Physics, Comenius University, 842 15 Bratislava, Czechoslovakia