Ann. Inst. Statist. Math.
32 (1980), Part A, 393-400

FITTING AUTOREGRESSION WITH REGULARLY MISSED OBSERVATIONS

HIDEAKI SAKAI

(Received Jan. 23, 1978; revised June 27, 1980)

Abstract

The effect of regularly missed observations on the estimation of parameters of an autoregressive (AR) process is investigated by using the frequency domain method. For first order AR processes, numerical results are shown to see a behavior of variances of the estimate due to the missed observations. In some cases, we can positively utilize the concept of missed observations to decrease the variances if the number of observations is fixed but time instants at with the observations are made can be changed.

1. Introduction

The problem of time series with missed observations was first treated by Jones [5] where the instants of missed observations were assumed to be periodic and their effect on spectral analysis was investigated.

After this work, several papers concerning this aspect, for example, [2], [3], [6], [7], [8], [10], have been published, but all these papers were concerned with the analysis in the frequency domain, or in other words, spectral analysis based on variously modified Blackman-Tukey procedures.

Recently, several works have been done concerning the effects of missed observations on the estimates of parametric models. For example, the author reported one result in [9] where the relation between fitting autoregression and periodogram, an important quantity in the frequency domain analysis, is presented and utilized to derive the asymptotic error covariance matrix of the estimate of the parameters of an autoregressive (AR) process with randomly missed observations, the situation treated by Scheinok [10].

In this paper, using the same technique, we calculate the asymptotic error covariance matrix for the situation with regularly missed observations, treated by Jones [5].

2. Regularly missed observations

Let a zero-mean stationary time series $\{x_t\}$ be sampled in groups of α consecutive time instants separated by β missed observations ($\alpha > \beta$). This situation may occur in the radar studies of the moon surface since during the reception of the radar echo, one must systematically cease the signal transmission so that there are time intervals without the reflections of the signals [8].

Let

$$a_t = \begin{cases} 1 & \text{if } x_t \text{ is read}, \\ 0 & \text{if } x_t \text{ is not read}. \end{cases}$$

Hence, $\{a_t\}$ is a sequence with period $\alpha+\beta$. According to [5], we define the limit of ratio of N, the total sample size, to the number of pairs available for estimating $r_k \overset{\Delta}{=} \mathrm{E}[x_t x_{t+k}]$ by

$$(1) \qquad c_k = \lim_{N\to\infty} \frac{N}{\sum_{t=1}^{N-|k|} a_t a_{t+|k|}} .$$

Then $c_k = c_{-k}$ and $\{c_k\}_{k=0}^{\infty}$ is also a sequence with period $\alpha+\beta$. The values of c_k during one period are as follows; $c_k = (\alpha+\beta)/(\alpha-k)$ $(0 \leqq k \leqq \beta)$, $c_k = (\alpha+\beta)/(\alpha-\beta)$ $(\beta \leqq k \leqq \alpha)$ and $c_k = (\alpha+\beta)/(k-\beta)$ $(\alpha \leqq k \leqq \alpha+\beta)$. It is obvious that the consistent estimator for r_k is given by

$$(2) \qquad \hat{r}_k = \frac{1}{N} \sum_{t=1}^{N-|k|} c_k a_t a_{t+|k|} x_t x_{t+|k|} .$$

By defining the modified periodogram as

$$(3) \qquad I'_N(s) = \frac{1}{2\pi N} \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} a_\nu a_\mu c_{\nu-\mu} x_\nu x_\mu \exp[-i(\nu-\mu)s] ,$$

$\hat{r}_k$ is expressed in terms of this periodogram as

$$(4) \qquad \hat{r}_k = \int_{-\pi}^{\pi} I'_N(s) \exp(iks) ds .$$

From (3) it is easily shown that

$$(5) \qquad \mathrm{E}[I'_N(s)] = \mathrm{E}[I_N(s)] + O(N^{-1})$$

where $I_N(s)$ is the usual periodogram without missed observations. The term of order N^{-1} in (5) arises from the fact that $\left(c_k - N \Big/ \sum_{t=1}^{N-|k|} a_t a_{t+|k|}\right)$ is of order N^{-1} but as is readily seen later, this term is of no importance for the ensuing analysis.

3. Parameter estimation of an AR process

Let $\{x_t\}$ be generated by a *Gaussian* stationary pth order AR process

$$x_t - b_1 x_{t-1} - \cdots - b_p x_{t-p} = u_t \tag{6}$$

with $\mathrm{E}[u_t]=0$ and $\mathrm{E}[u_t u_s]=\sigma^2\delta_{t,s}$. When the parameter estimation procedure is solving the well-known Yule-Walker equations (c.f. Akaike [1]) with the estimated $\hat{r}_k$'s in the place of true r_k's, it is the question how the asymptotic variances and covariances of the estimates for the unknown parameters $b_1, b_2, \cdots, b_p$ and σ^2 are affected by the above mentioned missed observations.

These estimators $\hat{b}_1, \cdots, \hat{b}_p$ and $\hat{\sigma}^2$ are obviously consistent and noting the relations (4), (5), the basic formula in [9] remains valid with replacing $I_N(s)$ by $I'_N(s)$ (cf. [9], (12)). Thus the asymptotic covariance matrix of the estimation error $\Delta\boldsymbol{b}=(\hat{b}_1-b_1, \cdots, \hat{b}_p-b_p)^T$ is given by

$$(\boldsymbol{R}\,\mathrm{E}[\Delta\boldsymbol{b}\Delta\boldsymbol{b}^T]\boldsymbol{R}^T)_{m,n}=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)\,\mathrm{Cov}\,[I'_N(s), I'_N(t)] \times \exp[i(ms+nt)]ds dt + O(N^{-2}) \tag{7}$$

where $(\boldsymbol{R})_{i,k}\overset{\Delta}{=}(i,k)$th element of $\boldsymbol{R}=r_{i-k}$ and $B(s)\overset{\Delta}{=}\sum_{k=0}^{p}(-b_k)\exp(-iks)$ $(-b_0=1)$. The effect of the second term in (5) is absorbed in $O(N^{-2})$ term by the similar argument to derive [9], (12). Hence, it is sufficient to know $\mathrm{Cov}\,[I'_N(s), I'_N(t)]$ for calculating (7).

According to Jones [5], $a_\nu a_\mu c_{\nu-\mu}$ in (3) is periodic so that it has the following two-dimensional representation;

$$a_\nu a_\mu c_{\nu-\mu}=\sum_{k,j} H_{k,j}\exp(-i\nu\lambda_k+i\mu\lambda_j) \tag{8}$$

where $\lambda_k\overset{\Delta}{=}2\pi k/(\alpha+\beta)$ and $k, j=-(\alpha+\beta-1)/2, -(\alpha+\beta-3)/2, \cdots, (\alpha+\beta-1)/2$ if $\alpha+\beta$ is odd, or $-(\alpha+\beta-2)/2, -(\alpha+\beta-4)/2, \cdots, (\alpha+\beta)/2$ if $\alpha+\beta$ is even. In general, $H_{k,j}$ is very complicated as indicated in [4], p. 458 but for $\beta=1$ it reduces to $H_{k,k}=\delta_{k,0}$ and $H_{k,j}=(\alpha^{-1}-\delta_{k,0}-\delta_{j,0})/(\alpha-1)$ $(k\neq j)$. Substituting (8) into (3) and introducing the discrete Fourier transform

$$J_N(s)\overset{\Delta}{=}\sum_{t=1}^{N} x_t\exp(-its)\ , \tag{9}$$

$I'_N(s)$ is reexpressed as

$$I'_N(s)=\frac{1}{2\pi N}\sum_{k,j} H_{k,j}J_N(s+\lambda_k)J_N(-s-\lambda_j)\ . \tag{10}$$

Since by the Gaussian assumption $J_N(s)$ is also Gaussian, it easily follows that

$$\text{(11)}\quad \begin{aligned}&\operatorname{Cov}[I_N'(s), I_N'(t)]\\ &=\frac{1}{(2\pi N)^2}\sum_{k,j}\sum_{k',j'} H_{k,j}H_{k',j'}\\ &\quad\times\{\mathrm{E}[J_N(s+\lambda_k)J_N(t+\lambda_{k'})]\,\mathrm{E}[J_N(-s-\lambda_j)J_N(-t-\lambda_{j'})]\\ &\quad+\mathrm{E}[J_N(s+\lambda_k)J_N(-t-\lambda_{j'})]\,\mathrm{E}[J_N(-s-\lambda_j)J_N(t+\lambda_{k'})]\}\ .\end{aligned}$$

On the other hand, from Brillinger [4], p. 93, it follows that

$$\text{(12)}\qquad \mathrm{E}[J_N(s)J_N(t)]=2\pi f(s)D_N(s+t)+O(1)$$

where we define

$$\text{(13)}\qquad f(s)\stackrel{\Delta}{=}\frac{1}{2\pi}\sum_{k=-\infty}^{\infty} r_k\exp(-iks)$$

$$\text{(14)}\qquad D_N(s)\stackrel{\Delta}{=}\sum_{k=1}^{N}\exp(-iks)\ ,$$

respectively. The second term of (12) is uniform in s, t so that its contribution to the integrals below can be neglected. From (11), (12) we have

$$\text{(15)}\quad \begin{aligned}&\operatorname{Cov}[I_N'(s), I_N'(t)]\\ &=\frac{1}{N^2}\sum_{k,j,k',j'} H_{k,j}H_{k',j'}[f(s+\lambda_k)f(t+\lambda_{j'})D_N(s+t+\lambda_k+\lambda_{k'})\\ &\quad\times D_N(-s-t-\lambda_j-\lambda_{j'})+f(s+\lambda_k)f(t+\lambda_{k'})\\ &\quad\times D_N(s-t+\lambda_k-\lambda_{j'})D_N(-s+t-\lambda_j+\lambda_{k'})]\ .\end{aligned}$$

Substituting this into (7) and using (14), the integral due to the first term in the square bracket of (15) is written as

$$\text{(16)}\quad \begin{aligned}&\sum_{q=1}^{N}\sum_{q'=1}^{N}\int_{-\pi}^{\pi} B(s)f(s+\lambda_k)\exp[-is(q-q'-m)]ds\\ &\quad\times\int_{-\pi}^{\pi} B(t)f(t+\lambda_{j'})\exp[-it(q-q'-n)]dt\\ &\quad\times\exp[-i(\lambda_k+\lambda_{k'})q+i(\lambda_j+\lambda_{j'})q']\ .\end{aligned}$$

From the definition of $B(s)$ and $f(s)$, we readily obtain

$$\text{(17)}\quad \begin{aligned}&\int_{-\pi}^{\pi} B(s)f(s+\lambda_k)\exp[-is(\nu-m)]ds\\ &=\sum_{j=0}^{p}(-b_j)r_{\nu-m+j}\exp[i(\nu-m+j)\lambda_k]\stackrel{\Delta}{=}\Theta_{\nu-m}(\lambda_k)\ .\end{aligned}$$

Hence, (16) reduces to

$$(18)\quad \sum_{\nu=0}^{N-1} \Theta_{\nu-m}(\lambda_k)\Theta_{\nu-n}(\lambda_{j'}) \exp[-i(\lambda_k+\lambda_{k'})\nu] \times \sum_{q=1}^{N-\nu} \exp[-i(\lambda_k+\lambda_{k'}-\lambda_j-\lambda_{j'})q] + \sum_{\nu=-N+1}^{-1} \Theta_{\nu-m}(\lambda_k)\Theta_{\nu-n}(\lambda_{j'}) \exp[-i(\lambda_k+\lambda_{k'})\nu] \times \sum_{q=1-\nu}^{N} \exp[-i(\lambda_k+\lambda_{k'}-\lambda_j-\lambda_{j'})q] = \sum_{\nu=-\infty}^{\infty} \Theta_{\nu-m}(\lambda_k)\Theta_{\nu-n}(\lambda_{j'}) \exp[-i(\lambda_k+\lambda_{k'})\nu] \times D_N(\lambda_k+\lambda_{k'}-\lambda_j-\lambda_{j'})+O(1)$$

where we use the fact that $|\Theta_\nu(\lambda_k)|$ is exponentially decreasing. It is well known that

$$(19)\quad D_N(s)=\begin{cases} N & \text{for } s\equiv 0 \pmod{2\pi} \\ O(1) & \text{otherwise}. \end{cases}$$

Thus, the value of (18) is of order N if and only if $\lambda_k+\lambda_{k'}-\lambda_j-\lambda_{j'}\equiv 0 \pmod{2\pi}$. Since $-\pi<\lambda_k<\pi$ if $\alpha+\beta$ is odd and $-\pi<\lambda_k\leqq\pi$ if $\alpha+\beta$ is even, it follows that $|\lambda_k+\lambda_{k'}-\lambda_j-\lambda_{j'}|<4\pi$ regardless of the parity of $\alpha+\beta$. Hence, from the definition of λ_k, the above possibilities are $k+k'-j-j'=0, \pm(\alpha+\beta)$.

In a similar way, integral due to the second term in the square bracket in (15) is calculated as

$$(20)\quad \sum_{\nu=-\infty}^{\infty} \Theta_{\nu-m}(\lambda_k)\Theta_{-\nu-n}(\lambda_{k'}) \exp[-i(\lambda_k-\lambda_{j'})\nu] \times D_N(-\lambda_j+\lambda_{k'}+\lambda_k-\lambda_{j'})+O(1)\,.$$

This is also of order N if and only if $k+k'-j-j'=0, \pm(\alpha+\beta)$. Thus, the asymptotic value of (7) is given by

$$(21)\quad N^{-1} \sum_{\substack{k,j,k',j' \\ k+k'-j-j'=0,\pm(\alpha+\beta)}} H_{k,j}H_{k',j'} \sum_{\nu=-\infty}^{\infty} \{\Theta_{\nu-m}(\lambda_k)\Theta_{\nu-n}(\lambda_{j'}) \times \exp[-i(\lambda_k+\lambda_{k'})\nu] + \Theta_{\nu-m}(\lambda_k)\Theta_{-\nu-n}(\lambda_{k'}) \exp[-i(\lambda_k-\lambda_{j'})\nu]\}\,.$$

4. A simple example

To obtain the explicit values of (21), let $\{x_t\}$ be a first order AR process with $r_k=b^{|k|}$ ($|b|<1$). For $m=n=1$, (17) is

$$\Theta_{\nu-1}(\lambda)=\{r_{\nu-1}\exp(-i\lambda)-br_\nu\}\exp(i\nu\lambda)\,.$$

From this, the infinite summation in (21) is given by

$$(z_k-b^2)(z_{j'}-b^2)\frac{z_{k'-j'}}{1-b^2z_{k'-j'}}+(z_k-1)(z_{j'}-1)\frac{b^2}{1-b^2z_{j'-k'}}$$
$$+(z_k-b^2)(z_{k'}-1)\frac{b^2z_{k'-j'}}{1-b^2z_{k'-j'}}+b^2(z_k-1)(z_{k'}-1)$$
$$+(z_k-1)(z_{k'}-b^2)\frac{b^2z_{j'-k'}}{1-b^2z_{j'-k'}}$$

with $z_k \stackrel{\Delta}{=} \exp(-i\lambda_k)$. Numerical calculations were performed for various values of α, b with $\beta=1$. To see the effect of missed observations, we compare $N\,\mathrm{E}\,[(\Delta b)^2]_{\text{miss.}}$ with the error variance from data of length $2N/3$ without missed observations, since for $\alpha=2$, $\beta=1$, the number of the net observations is $2N/3$. The latter is [1]

$$\frac{2}{3}N\,\mathrm{E}\,[(\Delta b)^2]_{\text{cont.}} \simeq 1-b^2\,. \tag{22}$$

Table 1 shows these values for $b=0.1, 0.2, \cdots, 0.9$. It is interesting to note that as the correlation of the data becomes strong, that is $|b|>0.8$, the degrading effect of regularly missed observations disappears.

Table 1. Comparison of the variances with and without missed observations

$\lvert b\rvert$	$N\,\mathrm{E}\,[(\Delta b)^2]_{\text{miss.}}$	$N\,\mathrm{E}\,[(\Delta b)^2]_{\text{cont.}}$	$\lvert b\rvert$	$N\,\mathrm{E}\,[(\Delta b)^2]_{\text{miss.}}$	$N\,\mathrm{E}\,[(\Delta b)^2]_{\text{cont.}}$
0.1	2.940	1.485	0.6	1.289	0.960
0.2	2.765	1.440	0.7	0.884	0.765
0.3	2.486	1.365	0.8	0.527	0.540
0.4	2.126	1.250	0.9	0.231	0.285
0.5	1.714	1.125			

Table 2 shows the behavior of $N\,\mathrm{E}\,[(\Delta b)^2]_{\text{miss.}}$ for increasing values of α with $\beta=1$ fixed. The convergence to $1-b^2$ is apparent but converging rates are fairly different. That is, for small $|b|$, the rate is high whereas for larger $|b|$ near 1, the convergence is considerably slow. This phenomenon also occurs in the case of randomly missed observations [9] and can be explained as follows. Since the covariance estimator (2) is based on filling the missed observation with zero, a priori mean, this estimate does not make any use of the information about the data correlation. Thus, the degrading effect vanishes promptly as $\alpha\to\infty$ in the white noise case whereas it is still non-neglegible for the data with strong correlation. Also we can note a quite curious phenomenon in Table 2. That is, at $b=0.9$, the variance for $\alpha=2$ is smaller than those for $\alpha=10$, 20, 30! At present there is no explanation for this counterintuitive result. Perhaps, this is due to the suboptimality of the present estimation procedure and the maximum likelihood estimate may not possess such a property.

Table 2. Convergent behavior of $N\,\mathrm{E}\,[(\Delta b)^2]_{\mathrm{miss.}}$ for increasing α

$\|b\|$	$\alpha=10$	$\alpha=20$	$\alpha=30$	$\alpha=\infty$
0.1	1.222	1.093	1.053	0.990
0.2	1.181	1.063	1.022	0.960
0.3	1.129	1.014	0.978	0.910
0.4	1.055	0.944	0.908	0.840
0.5	0.959	0.853	0.818	0.750
0.6	0.839	0.740	0.707	0.640
0.7	0.689	0.607	0.575	0.510
0.8	0.501	0.447	0.421	0.360
0.9	0.266	0.252	0.239	0.190

To see the validity of the theoretical results, in Table 3 we present simulation results where empirical variances are obtained by averaging squares of estimation errors over M sets of data each of N length. We can see a fairly good agreement between the theoretical and experimental results.

Table 3. The simulation results to show the validity of the theoretical analysis

a	Number of data N	Number of data sets M	$N\,\mathrm{E}\,[(\Delta b)^2]_{\mathrm{miss.}}$ by theory	By simulations
0.8	1000	500	$\alpha=2$ 0.527 $\alpha=10$ 0.501	$\alpha=2$ 0.576 $\alpha=10$ 0.493
0.9	1000	500	$\alpha=2$ 0.231 $\alpha=10$ 0.266 $\alpha=\infty$ 0.190	$\alpha=2$ 0.262 $\alpha=10$ 0.277 $\alpha=\infty$ 0.202
0.9	1000	900	$\alpha=2$ 0.231 $\alpha=10$ 0.266	$\alpha=2$ 0.262 $\alpha=10$ 0.282
0.95	1000	500	$\alpha=2$ 0.108 $\alpha=10$ 0.134	$\alpha=2$ 0.134 $\alpha=10$ 0.147

5. Conclusion

At first sight we are apt to think negative effects of missed observations. But from the above results, in some cases, we can positively utilize the concept of missed observations to improve the performance of the estimate if the number of observations is fixed but time instants at which the observations are made can be changed. For example, for a first order AR process with $b=0.9$, about 20% reduction of the variance is gained if we allocate the total observations of length, say, $N=500$ over 750 instants to form regularly missed observations with $\alpha=2$, $\beta=1$.

Acknowledgement

The author is grateful to Prof. H. Tokumaru of Kyoto University and Prof. T. Soeda of Tokushima University for their useful advice and constant encouragement. Thanks are also due to the reviewer for his interest in this work and pointing out several errors in early versions of the manuscript.

KYOTO UNIVERSITY

REFERENCES

[1] Akaike, H. (1970). Statistical predictor identification, *Ann. Inst. Statist. Math.*, **22**, 203-217.

[2] Alekseev, V. G. and Savitsky, Yu. A. (1973). On spectral analysis of Gaussian random processes with missing observations, *Prob. Inform. Transmiss.*, **9**, 66-72.

[3] Bloomfield, P. (1970). Spectral analysis with randomly missing observations, *J. R. Statist. Soc.*, B, **32**, 369-380.

[4] Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*, Holt, Rinehart, and Winston, Inc., New York.

[5] Jones, R. H. (1962). Spectral analysis with regularly missed observations, *Ann. Math. Statist.*, **33**, 455-461.

[6] Jones, R. H. (1971). Spectrum estimation with missing observations, *Ann. Inst. Statist. Math.*, **23**, 387-398.

[7] Neave, H. R. (1970). Extending the frequency range of spectrum estimate by the use of two data recorders, *Technometrics*, **12**, 877-890.

[8] Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation, *Sankhyā*, A, **25**, 383-392.

[9] Sakai, H., Soeda, T. and Tokumaru, H. (1979). On the relation between fitting autoregression and periodogram with applications, *Ann. Statist.*, **7**, 96-107.

[10] Scheinok, P. A. (1965). Spectral analysis with randomly missed observations: the binomial case, *Ann. Math. Statist.*, **36**, 971-977.