The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 59-74

A Study on the Least Squared Estimator of Autoregressive Models when Consecutive Missing Observations Exist

Gui Yeol Ryu[1)]

Abstract

The properties of the residuals are investigated when k-consecutive observations are interpolated. The central limit theorem is also proved for the LSE for autoregressive parameters when k-consecutive observations are contaminated. The performance of the interpolated LSE in small samples is investigated by simulation. And the interpolated LSE is compared with the Yule-Walker type estimator.

1. Introduction

Autoregressive moving average(ARMA) models are commonly used to model univariate time series data. Most statistical inferences for the ARMA model were developed under the assumption that the observations are available at consecutive equally spaced intervals of time. It is, however, not unusual to encounter data which have missing observations. They cause serious problems in the model identification, the parameter estimation, and the forecasting because the covariance matrix of the observed variables is no longer Toeplitz, a structure that the complete data procedures make use of. There are two problems related to missing observations, which are the estimation of the missing observations and the estimation of the model parameters.

To handle these problems, there would be three approaches. The first is estimating parameters without estimating missing observations. Dunsmuir and Robinson(1981) considered the Yule-Walker(Y-W) type estimator for autoregressive models when missing observations are present and investigated the asymptotic normality. Reinsel and Wincek(1987) considered the maximum likelihood estimator(MLE) and the weighted least squared estimator for first order autoregressive models(AR(1)). They considered the parameter estimation procedure when missing observations occur randomly or sampling is periodic.

The second is that Johnes(1980) and Kohn and Ansley(1986), etc, considered the MLE of missing observations and parameters simultaneously using the state space representation. To

1) Full-time Lecturer, Department of Applied Statistics, Seo Kyeong University, Chongnung-Dong Songbuk-Ku Seoul, 136-704, Korea

reduce the computational burden they adopted the Kalman filter for a recursive evaluation.

The third is to estimate missing observations and parameters iteratively using EM algorithm(Little and Rubin(1987)) such as Pourhmadi(1989), etc. Furthermore in the area of the detection of outliers in time series, many authors such as Bruce and Martin(1989), Ljung(1989), Lee(1990), Ryu(1991) and Cho *et. al.* (1994), etc, considered the detection procedures under the series that k-consecutive suspicious observations were replaced by estimates. They adopt the EM algorithm for estimating the parameters. In this paper we will consider the parameter estimation procedure using the EM algorithm such that it computes k-consecutive observations and parameter estimates iteratively, and investigate the properties of the procedure.

Our paper is organized as follow. In Section 2, we will consider the estimation of missing observations and investigate the properties of the interpolated residuals. Section 3 contains the conditions of the asymptotic normality of the LSE when k-consecutive observations are contaminated for AR(h) and interpolated for AR(1)(interpolated LSE). And in Section 4, we will investigate the small sample properties of the interpolated LSE by simulation and compare with the Y-W type estimator.

From now on, we assume that time series x_i's follow h^{th}-order autoregressive model such that

$$x_t = \pi_0 + \pi_1 x_{t-1} + \ldots + \pi_h x_{t-h} + \varepsilon_t, \quad t = h+1, \ldots, n$$

where $\pi' = (\pi_0, \pi_1, \ldots, \pi_h)$ lies in the stationary region and ε_t's are white noises, which are independently and identically distributed as a normal distribution with a mean 0 and a variance σ^2_ε.

2. Estimation of Missing Observations

In this section, we will consider the estimation procedure of the missing observations and investigate the properties of the interpolated residuals.

In case of estimating a block of missing observation, there are two approaches. We first introduce a few notations. When time series of size n with k-consecutive missing observations beginning with T_0 is observed, then partition $\mathbf{x}' = (x_1, \ldots, x_{T_0-1}, x_{T_0}, \ldots, x_n)$ as follows:

$$\mathbf{x}' = (\mathbf{x}_B', \mathbf{x}_M', \mathbf{x}_A')$$

where $\mathbf{x}_B' = (x_1, \ldots, x_{T_0-1})$, $\mathbf{x}_M' = (x_{T_0}, \ldots, x_{T_0+k-1})$ and $\mathbf{x}_A' = (x_{T_0+k}, \ldots, x_n)$. We will exclude the cases of missing observations near the start and the end of the series for simplicity. We may classify the estimators into two types, one is the predicted estimator denoted by $\hat{\mathbf{x}}_M$ and the other is the interpolated estimator denoted by $\hat{\mathbf{x}}_M^I$.

Jones(1980), Harvey and Pierse(1984), Ansley and Kohn(1985), and Kohn and Ansley(1986) used the predicted estimator $\hat{x}_M$ which is the orthogonal projection of $\hat{x}_M$ onto the space spanned by $\{x_t,\ t < T_0\}$. It is obvious $\hat{x}_M = E(x_M \mid x_t, t < T_0)$. To obtain the MLE of parameters and $\hat{x}_M$ simultaneously, the state space representation was used and the Kalman filter for a recursive evaluation was employed to reduce the computational burden. Nevertheless this method seems to require heavy and complicated computations. Hence, it is difficult to apply this without computer package.

Brubacher and Wilson(1976), Abraham(1981), Miller and Ferreiro(1984) and Pourahmadi(1989) considered the interpolated estimator of x_M, which can be expressed as $\hat{x}_M^I = \hat{x}_M + C(x_A - \hat{x}_A)$ for some matrix C, where $\hat{x}_M^I$ is the orthogonal projection of x_M onto the space spanned by $\{x_t,\ t < T_0$ and $t \geq T_0+k\}$. They considered the above type of estimator since the space spanned by $\{x_t,\ t < T_0\}$ is not orthogonal to the space spanned by $\{x_t,\ t \geq T_0+k\}$, but to the space spanned by $\{x_t - \hat{x}_t,\ t \geq T_0+k\}$. They adopted the EM algorithm to estimate missing observations and parameters iteratively. Peña and Tiao(1991) compared missing values as unknown parameters with as random variables suggested by Newbold(1974) and Box and Jenkins(1976). The latter procedure is more relevant to estimate the missing observations than the former and under the Gaussian assumption, is the same as the $\hat{x}_M^I$. Hence, we will adopt the Pourahmadi(1989)'s estimators as missing observations estimators. He suggested the best linear interpolator in the case of interpolating k-consecutive observations as follows :

$$\hat{x}_M^I = \hat{x}_M + (D(I_k + E'E)^{-1}T_k)'(x_A - \hat{x}_A) \tag{2.1}$$

where E and D are $(n-T_0-k+1)\times k$ matrices satisfying

$$T_r'E = B_{r,k}$$
$$T_rD = E \qquad \text{for } r = n-T_0-k+1$$

$$T_r' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_1 & 1 & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ b_{r-1} & b_{r-2} & \cdots & 1 \end{pmatrix}$$

$$B_{r,k} = (b_{k+i-j}) \quad \text{for } i = 0,\ldots,r-1, \text{ and } j = 0,\ldots,k-1$$

and b_i's are the MA parameters of $\{x_t\}$. $\hat{x}_M$ and $\hat{x}_A$ are the prediction of x_M and x_A based on $\{x_t,\ t < T_0\}$.

In the case of estimating a single missing observation, the best linear estimator of x_T is

$$\hat{x}_T = E(x_T|\ X(T)) \tag{2.2}$$

where $X(T) = (x_1, x_2, \ldots, x_{T-1}, x_{T+1}, \ldots, x_n)'$, which is the same as the optimum interpolation of x_T in the mean squares error sense. Grenander and Rosenblatt(1957) showed that the expected value of a missing value given the rest of the data is

$$E(x_T|X(T)) = -\sum_{j=1}^{m} \rho_j^i s(j) \tag{2.3}$$

where ρ_j^i is the j^{th} inverse autocorrelation coefficient(Wei(1990)), $s(j) = x_{T-1} + x_{T+1}$ and $m = \min(T-n,\ n-T)$. For h^{th}-order autoregressive models ρ_j^i is also the j^{th} coefficient in the generating function $(\sum_{i=1}^{h} \pi_i^2) \pi(B) \pi(F)$, which can be interpreted as the inverse autocorrelation function of the process. (2.3) is the same as (2.1) in the case of estimating a single missing observation.

Following Proposition investigates the properties of the T^{th} interpolated residual which is the difference of the T^{th}observation and its interpolator.

Proposition 2.1 When T^{th} observation is interpolated, for AR(h), the interpolated residual is

$$\begin{aligned} \tilde{\omega}_T &= x_T - \hat{x}_T^I \\ &= (\varepsilon_T - \sum_{i=1}^{h} \pi_i \varepsilon_{T+i})/K \end{aligned} \tag{2.4}$$

where $K = \sum_{i=1}^{h} \pi_i^2$ and $\hat{x}_T^I$ is obtained from (2.3).

Proof. For AR(1),

$$\begin{aligned} \tilde{\omega}_T &= x_T - \hat{x}_T^I \\ &= \pi_1 x_{T-1} + \varepsilon_T - \pi_1 (x_{T-1} + x_{T+1})/(1 + \pi_1^2) \\ &= (\varepsilon_T - \pi_1 \varepsilon_{T+1}) / \sum_{i=0}^{1} \pi_i^2. \end{aligned}$$

For AR(2),

$$\begin{aligned} \tilde{\omega}_T &= x_T - \hat{x}_T^I \\ &= \pi_1 x_{T-1} + \pi_2 x_{T-2} + \varepsilon_T - ((\pi_1 - \pi_1 \pi_2 (x_{T-1} + x_{T+1}) + \pi_2 (x_{T-2} + x_{T+2})) / \sum_{i=0}^{2} \pi_i^2 \\ &= (\varepsilon_T - \pi_1 \varepsilon_{T+1} - \pi_2 \varepsilon_{T+2}) / \sum_{i=0}^{2} \pi_i^2. \end{aligned}$$

For AR(3),

$$\tilde{\omega}_T = x_T - \hat{x}_T^I$$
$$= (\varepsilon_T - \pi_1 \varepsilon_{T+1} - \pi_2 \varepsilon_{T+2} - \pi_3 \varepsilon_{T+3}) / \sum_{i=0}^{3} \pi_i^2.$$

Continuing calculations in this way, we obtain, for AR(h), h=1,2,3, ...

$$\tilde{\omega}_T = (\varepsilon_T - \sum_{i=0}^{h} \pi_i \varepsilon_{T+i}) / K \qquad \text{Q.E.D.}$$

From the above proposition, we know that the variance of the interpolated residual is smaller than that of the one step ahead forecast, because $Var(\tilde{\omega}_T) = \sigma_\varepsilon^2 / K$ and the variance of the one step ahead forecast is σ_ε^2. In the case of estimating k-consecutive observations, we investigate the properties of the interpolated residuals for AR(1).

Proposition 2.2 When k-consecutive observations are interpolated with the starting time T, for AR(1), the interpolated residual is

$$\tilde{\omega}_{T+j} = x_{T+j} - \hat{x}_{T+j}^I$$
$$= \begin{cases} K^{-1} \sum_{i=0}^{k-j-1} (\pi^{2i} \sum_{l=0}^{j} \pi^l \varepsilon_{T+j-l} - (\sum_{l=0}^{[j/2]} (\pi^{k+i-2l} + \pi^{k+i+2l}) - \pi^{k+i}) \varepsilon_{T+k-i}) & \text{if } j \text{ is even} \\ K^{-1} \sum_{i=0}^{k-j-1} (\pi^{2i} \sum_{l=0}^{j} \pi^l \varepsilon_{T+j-l} - \sum_{l=0}^{[j/2]+1} (\pi^{k+i-2l} + \pi^{k+i+2l}) \varepsilon_{T+k-i}) & \text{if } j \text{ is odd,} \end{cases}$$

where $K = \sum_{i=0}^{k} \pi^{2i}$, [k] is the integer value of k, and $\hat{x}_T^I$ is obtained from (2.3).

Proof. For t = T,

$$\tilde{\omega}_T = x_T - \hat{x}_T^I$$
$$= \varepsilon_T - (\pi^k \sum_{i=0}^{k} \pi^i \varepsilon_{T+k-i}) / K$$
$$= (\sum_{i=0}^{k-1} \pi^{2i} \varepsilon_T - \pi^{k+i} \varepsilon_{T=k-i}) / K$$
$$\text{let} = a'_T \varepsilon$$

where $a_T' = K^{-1}(\sum_{i=0}^{k-1} \pi^{2i}, -\pi^{2k-1}, -\pi^{2k-2}, \ldots, -\pi^k)$, and $\varepsilon' = (\varepsilon_T, \varepsilon_{T+1}, \ldots, \varepsilon_{T+k})$.

For t = T+1,

$$\tilde{\omega}_{T+1} = x_{T+1} - \hat{x}_{T+1}^I$$
$$= K^{-1} \sum_{i=0}^{k-2} (\pi^{2i+1} \varepsilon_T + \pi^{2i} \varepsilon_{T+1} - (\pi^{k+i-1} + \pi^{i+k+1}) \varepsilon_{T+k-i})$$
$$= K^{-1} \sum_{i=0}^{k-2} (\sum_{l=0}^{1} \pi^{2i+l} \varepsilon_{T+1-l} - (\pi^{k+i-1} + \pi^{k+i+1}) \varepsilon_{T+k-i})$$
$$\text{let} = a'_{T+1} \varepsilon$$

where

$$a_{T+1}' = K^{-1}\left(\sum_{i=0}^{k-2}\pi^{2i+1}, \sum_{i=0}^{k-2}\pi^{2k}, -(\pi^{2k-3}+\pi^{2k-1}), -(\pi^{2k-5}+\pi^{2k-3}), \ldots, -(\pi^{k-1}+\pi^{k+1})\right).$$

For t = T+2,

$$\begin{aligned}\tilde{\omega}_{T+2} &= x_{T+2} - \hat{x}_{T+2}^{I} \\ &= K^{-1}\sum_{i=0}^{k-3}(\pi^{2i+2}\varepsilon_{T} + \pi^{2i+1}\varepsilon_{T+1} + \pi^{2i}\varepsilon_{T+2} - (\pi^{k+i-2} + \pi_{k+i} + \pi^{k+i+2})\varepsilon_{T+k-i}) \\ &= K^{-1}\sum_{i=0}^{k-3}\left(\sum_{l=0}^{2}\pi^{2i+l}\varepsilon_{T+2-l} - (\pi^{k+i-2} + \pi_{k+i} + \pi^{k+i+2})\varepsilon_{T+k-i}\right) \\ \text{let} &= a'_{T+2}\varepsilon,\end{aligned}$$

where

$$a_{T+2}' = K^{-1}\left(\sum_{i=0}^{k-3}\pi^{2i+2}, \sum_{i=0}^{k-3}\pi^{2k+1}, \sum_{i=0}^{k-3}\pi^{2k}, -(\pi^{2k-5}+\pi^{2k-3}+\pi^{2k-1}), \ldots, -(\pi^{k-2}+\pi^{k}+\pi^{k+2})\right).$$

Continuing calculations in this way, we obtain the result. Q.E.D.

3. Asymptotic Properties of Least Squared Estimator

From here, we explain the conditions of the asymptotic normality of the LSE of AR(h) when k-consecutive observations are contaminated and the asymptotic normality of the LSE of AR(1) when k-consecutive observations are replaced by the interpolators.

Dunsmuir and Robinson(1981a) considered that the Y-W type parameter estimators for autoregressive models when missing observations are present adopting the amplitude modulated process introduced by Parzen(1961), investigated the asymptotic normality. Because they adopted a different autocovariance function from the usual one, to compute the Y-W type estimator is much complicated. When the fraction of missing observations is not big, it would be reasonable to estimate missing observations and parameters iteratively, which is especially true when missing value techniques are applied to detect outliers. Thus we will consider the parameter estimation procedure when the k-consecutive observations are replaced by the interpolators.

The following theorem says that the LSE in the presence of k-consecutive contaminations has asymptotic normality under some conditions.

Theorem 3.1 Let x_t be an autoregressive time series with the representation

$$x_t = \pi_0 + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \ldots + \pi_h x_{t-h} + \varepsilon_t,$$

where $\boldsymbol{\pi}' = (\pi_0, \pi_1, \ldots, \pi_h)$ lies under the stationary region, and ε_t's are independent

$(0, \sigma_\varepsilon^2)$ random variables with $E(\varepsilon_t^4) = \eta\sigma_\varepsilon^4$. Let y_t be contaminated by ω_t, $t = T, T+1, \ldots, T+k-1$ such that

$$y_t = \begin{cases} x_t, & \text{if } t \neq T, T+1, \ldots T+k-1 \\ x_t + \omega_t, & \text{if } t = T, T+1, \ldots, T+k-1 \end{cases}$$

where parameters k and ω_t are satisfying following conditions:

for every j, $n^{-1/2} \sum_{t=T}^{T+k-1} \omega_t \omega_{t+j}$ converges in probability to 0

for every j, $n^{-1/2} \sum_{t=T}^{T+k-1} \omega_t x_{t+j}$ converges in probability to 0.

Let $\widehat{\boldsymbol{\pi}}_T$ be the LSE for $\boldsymbol{\pi}$ of $\{y_t\}$, then

$$\sqrt{n}\,(\widehat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{L} N(0, A^{-1}\sigma_\varepsilon^2)$$

where $A_n = (a_{ij})_{h\times h}$ with A_n converging in probability to A, $a_{ij} = (n-h)^{-1} \sum_{t=h+1}^{n} x_{t-i}\, x_{t-j}$ and $\xrightarrow{L}$ means convergence in distribution.

Proof. We first introduce the notations as follows:

$$X = \begin{pmatrix} x_h & x_{h-1} & \cdots & x_1 \\ \vdots & \vdots & & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-h} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_{h+1} \\ \vdots \\ x_n \end{pmatrix},$$

and

$$Y = \begin{pmatrix} y_h & y_{h-1} & \cdots & y_1 \\ \vdots & \vdots & & \vdots \\ y_{n-1} & y_{n-2} & \cdots & y_{n-h} \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_{h+1} \\ \vdots \\ y_n \end{pmatrix}.$$

Then we have $Y = X + M$ and $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{m}$, where

$M' = (M(1)'M(2)'M(3)')$, $\boldsymbol{m}' = (\boldsymbol{m}(1)'\, \boldsymbol{m}(2)'\, \boldsymbol{m}(3)')$,

$$M(1) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{(T-h)\times h}, \quad M(2) = \begin{pmatrix} \omega_T & 0 & \cdots & 0 \\ \omega_{T+1} & \omega_T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_{T+k-1} \end{pmatrix}_{(k+h-1)\times h},$$

$$M(3) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{(n-T-k-h+1)\times h},$$

and

$$\boldsymbol{m}(1) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{(T-h-1)\times 1}, \quad \boldsymbol{m}(2) = \begin{pmatrix} \omega_T \\ \vdots \\ \omega_{T+k-1} \end{pmatrix}_{k\times 1}, \quad \boldsymbol{m}(3) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{(n-T-k+1)\times 1}.$$

And we partition X such that

$$X' = (X(1)' \; X(2)' \; X(3)')$$

where

$$X(1) = \begin{pmatrix} x_h & \cdots & x_1 \\ \vdots & & \vdots \\ x_{T-1} & \cdots & x_{T-h} \end{pmatrix}_{(T-h)\times h},$$

$$X(2) = \begin{pmatrix} x_T & \cdots & x_{T-h+1} \\ \vdots & & \vdots \\ x_{T+h+k-2} & \cdots & x_{T+k-1} \end{pmatrix}_{(h+k-1)\times h},$$

and

$$X(3) = \begin{pmatrix} x_{T+h+k-1} & \cdots & x_{T+k} \\ \vdots & & \vdots \\ x_{n-1} & \cdots & x_{n-h} \end{pmatrix}_{(n-T-h-k+1)\times h}.$$

Let $\widehat{\boldsymbol{\pi}}$ and $\widehat{\boldsymbol{\pi}}_T$ be the LSE of $\boldsymbol{\pi}$ of $\{x_t\}$ and $\{y_t\}$, respectively, then $\widehat{\boldsymbol{\pi}} = (X'X)^{-1}X'\boldsymbol{x}$ and

$$\begin{aligned} \widehat{\boldsymbol{\pi}}_T &= (Y'Y)^{-1}Y'\boldsymbol{y} \\ &= (Y'Y)^{-1}X'\boldsymbol{x} + (Y'Y)^{-1}(X'\boldsymbol{m} + M'\boldsymbol{x} + M'\boldsymbol{m}). \end{aligned}$$

We have

$$\widehat{\boldsymbol{\pi}} - \widehat{\boldsymbol{\pi}}_T = ((X'X)^{-1} - (Y'Y)^{-1})X'\boldsymbol{x} - (Y'Y)^{-1}(X'\boldsymbol{m} + M'\boldsymbol{x} + M'\boldsymbol{m})$$

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\pi}} - \widehat{\boldsymbol{\pi}}_T) &= ((n^{-1}X'X)^{-1} - (n^{-1}Y'Y)^{-1})n^{-1/2}X'\boldsymbol{x} \\ &\quad - (n^{-1}Y'Y)^{-1}(n^{-1/2}X'\boldsymbol{m} + n^{-1/2}M'\boldsymbol{x} + n^{-1/2}M'\boldsymbol{m}). \end{aligned} \tag{3.1}$$

We investigate the first term of the right hand side of (3.1). From the following equation

$$n^{-1}Y'Y = n^{-1}X'X - 2n^{-1}M(2)'X(2) + n^{-1}M(2)'M(2). \tag{3.2}$$

After rigorous computations, we have known that the elements of $M(2)'X(2)$ are represented by linear combinations of ω_{t_1} and x_{t_2} and the elements of $M(2)'M(2)$ are represented by linear combinations of ω_{T_1} and ω_{T_2} for some T_1 and T_2. If the conditions are satisfied, the second two terms of the right hand side of (3.2) converge in probability to zero. Thus we can see that

$$\frac{1}{n}(Y'Y) \xrightarrow{P} A.$$

Therefore, the first term of the right hand side of (3.1) converges in probability to 0.

The elements of $n^{-1/2}X'\boldsymbol{m}$ can be represented as $n^{-1/2}\sum_{i=T}^{T-k+1}\omega_i x_{i+j}$ for some j and the elements of $n^{-1/2}M'\boldsymbol{m}$ are $n^{-1/2}\sum_{i=T}^{T-k+1}\omega_i\omega_{i+j}$ for some j. If the conditions of the theorem is satisfied, we see that the second term also converges in probability to 0. Thus $\hat{\pi} - \hat{\pi}_T = o_p(n^{-1/2})$. From Fuller (1976), $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{L} N(0, A^{-1}\sigma_\varepsilon^2)$ where $\hat{\pi}$ is the LSE for $\{x_t\}$.

Thus

$$\sqrt{n}(\hat{\pi}_T - \pi) \xrightarrow{L} N(0, A^{-1}\sigma_\varepsilon^2). \qquad \text{Q.E.D.}$$

If k is fixed, the conditions of theorem are satisfied and $\hat{\pi}_T$ has the asymptotic normality. This theorem can be applied to the estimation problem of parameters when k-consecutive missing observations are present because missing observations can be substituted with estimators and regarded as contaminations.

In the following Theorem, when k-consecutive observations are replaced by the interpolator, the conditions of Theorem 3.1 are satisfied if $k = o(n^{1/2})$.

Theorem 3.2 Let x_t^I be the interpolated series such that

$$x_t^I = \begin{cases} x_t & \text{if } t \neq T, T+1, \cdots, T+k-1 \\ \hat{x}_t^I & \text{if } t = T, \cdots, T+k-1 \end{cases}$$

where $\hat{x}_t^I$ is obtained from (2.1) and T the starting point of interpolation. Let $\hat{\pi}$ be the LSE for π of x_t^I for AR(1), then

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{L} N(0, A^{-1}\sigma_\varepsilon^2)$$

where A is the same as that of Theorem 3.1, if $k = o(n^{1/2})$.

Proof. We have only to show that if $k = o(n^{1/2})$,

$$n^{-1/2} \sum_{t=T}^{T+k-1} \tilde{\omega}_t \, \tilde{\omega}_{t+j} = o_p(1) \quad \text{for any } j, \tag{3.3}$$

$$n^{-1/2} \sum_{t=T}^{T+k-1} \tilde{\omega}_t \, \tilde{x}_{t+j} = o_p(1) \quad \text{for any } j, \tag{3.4}$$

First we will prove (3.3)

$$\sum_{t=T}^{T+k-1} \tilde{\omega}_{T+i} \, \tilde{\omega}_{T+i+j} = \varepsilon' \sum_{i=0}^{k-1} a_{T+i} \, a_{T+i+j}' \, \varepsilon$$

$$\text{let } = \varepsilon' C \varepsilon$$

Where $\tilde{\omega}_t$ and a_t are the same as those of Proposition 2.2.

$$\tilde{\omega}_{T+i} \, \tilde{\omega}_{T+i+j} = \frac{1}{(1-\pi^2)^2} \varepsilon' C_{i,i+j} \varepsilon,$$

where

$$C_{i,i+j}= \boldsymbol{v}_i \ \boldsymbol{v}_{i+j}'$$

$$\boldsymbol{v}_i=\begin{pmatrix} \pi^{i}(1-\pi^{2(k-i)}) \\ \pi^{i-1}(1-\pi^{2(k-i)}) \\ \vdots \\ \pi^{0}(1-\pi^{2(k-i)}) \\ -\pi^{2k-2i-1}(1-\pi^{2(i+1)}) \\ -\pi^{2k-2i-2}(1-\pi^{2(i+1)}) \\ \vdots \\ -\pi^{k-i}(1-\pi^{2(i+1)}) \end{pmatrix}, \ \boldsymbol{v}_{i+j}=\begin{pmatrix} \pi^{i+j}(1-\pi^{2(k-(i+j))}) \\ \pi^{i+j-1}(1-\pi^{2(k-(i+j))}) \\ \vdots \\ \pi^{0}(1-\pi^{2(k-(i+j))}) \\ -\pi^{2k-1-2(i+j)}(1-\pi^{2(i+j+1)}) \\ \vdots \\ -\pi^{k-(i+j)}(1-\pi^{2(i+j+1)}) \end{pmatrix}.$$

Then $C=\sum_{i=0}^{k-1} C_{i,i+j}=(c_{l,m})$. If $\sum_{i=1}^{k}\sum_{j=1}^{k} c_{ij}^2$ is bounded, then $\boldsymbol{\varepsilon}' C\boldsymbol{\varepsilon}$ has the asymptotic distribution from Johnson and Kotz(1970). We also have $\sum_i\sum_j c_{ij}^2 \le \sum_i\sum_j b_{ij}^2$ where

$$B=\frac{1}{(1-\pi^2)^2}\sum_{i=0}^{k-1}\begin{pmatrix} \pi^{i} \\ \pi^{i-1} \\ \vdots \\ \pi^{0} \\ \pi^{2k+1} \\ \pi^{2k} \\ \vdots \\ \pi^{k+i+2} \end{pmatrix}(\pi^{i+j},\dots,\pi^0,-\pi^{2k+1},\dots,-\pi^{k+i+2+j})$$

fter heavy calculations, we obtain as follows:

$$b_{1,1}^2 \le L^{-1}\pi^{2,j}(1-\pi^{2k})^2,\ b_{1,2}^2 \le L^{-1}\pi^{2(j-1)}(1-\pi^{2k})^2,\ \dots,\ b_{1,j+1}^2 \le L^{-1}\pi^{0}(1-\pi^{2k})^2,$$

$$b_{1,j+2}^2 \le L^{-1}\pi^{2}(1-\pi^{2k})^2,\ b_{1,j+3}^2 \le L^{-1}\pi^{4}(1-\pi^{2k})^2,\ \dots,\ b_{1,k}^2 \le L^{-1}\pi^{2(k-j)}(1-\pi^{2k})^2,$$

where $L=(1-\pi^2)^2$. Thus we have

$$\sum_j b_{1,j}^2 \le L^{-2/3}(1-\pi^{2k})^2(1-\pi^{2(j+1)}+\pi^2-\pi^{2(k-j)}),$$

and

$$\lim_{k\to\infty}\sum_{j=1}^{k} b_{1,j}^2 \le \frac{1-\pi^{2(j+1)}+\pi^2}{L^{2/3}} \le M_1 \ \text{ for some } M_1.$$

Similarly we can show that

$$\lim_{k\to\infty}\sum_{j=1}^{k} b_{i,j}^2 \le M_1 \ \text{ for all } i.$$

Thus we have, if $k=o_p(n^{1/2})$,

$$\frac{\sum_{i=1}^{k}\sum_{j=1}^{k} b_{i,j}^2}{k} \le M_1,$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{k-1}\tilde{\omega}_{T+i}\,\tilde{\omega}_{T+i+j} \xrightarrow{P} 0.$$

Now we check the second condition

$$\sum_{i=0}^{k-1}\tilde{\omega}_{T+i}x_{T+i+j} = L^{-3/2}\varepsilon'\sum_{i=0}^{k-1}\begin{pmatrix} \pi^{i}(1-\pi^{2(k-i)}) \\ \pi^{i-1}(1-\pi^{2(k-i)}) \\ \vdots \\ \pi^{0}(1-\pi^{2(k-i)}) \\ \pi^{2k+1}(1-\pi^{-2(i+1)}) \\ \pi^{2k}(1-\pi^{-2(i+1)}) \\ \vdots \\ \pi^{k+i+2}(1-\pi^{-2(i+1)}) \end{pmatrix}(\pi^{T+i+j},\pi^{T+i+j-1},\dots,\pi^{0})\,\varepsilon_{T+i+j},$$

where $\varepsilon_{T+i+j}{}' = (\varepsilon_0,\varepsilon_1,\dots,\varepsilon_{T+i+j})$. Let $\sum_{i=0}^{k-1}\tilde{\omega}_{T+i}x_{T+i+j} = \varepsilon' E\,\varepsilon_{T+k-1+j}$, where $E=(e_{i,j})$. Then we have $\sum\sum e_{i,j}^2 \le \sum\sum d_{i,j}^2$, where

$$D = L^{-3/2}\sum_{i=0}^{k-1}\begin{pmatrix} \pi^{i} \\ \pi^{i-1} \\ \vdots \\ \pi^{0} \\ \pi^{2k+1} \\ \pi^{2k} \\ \vdots \\ \pi^{k+i+2} \end{pmatrix}(\pi^{T+i+j},\dots,\pi^{0}).$$

After heavy calculations, we can see

$$d_{1,1}^2 \le L^{-1}\pi^{2(T+j)}(1-\pi^{2k})^2,\ d_{1,2}^2 \le L^{-1}\pi^{2(T+j-1)}(1-\pi^{2k})^2,\dots,d_{1,T+j+1}^2 \le L^{-1}\pi^{0}(1-\pi^{2k})^2,$$

$$d_{1,T+j+2}^2 \le L^{-1}\pi^{0}(1-\pi^{2k})^2,\ d_{1,T+j+3}^2 \le L^{-1}\pi^{2}(1-\pi^{2k})^2,\dots,d_{1,T+k-1+j}^2 \le L^{-1}\pi^{2(k-1)}(1-\pi^{2k})^2.$$

Thus we can show

$$\sum_j d_{1,j}^2 \le L^{-2/3}(1-\pi^{2k})^2(1-\pi^{2(T+j+1)}+\pi^{2(1-\pi^{2k})^3}).$$

We have

$$\lim_{k\to\infty}\sum_{j=1}^{k} d_{1,j}^2 \le L^{-2/3}(1-\pi^{2(T+j+1)}+\pi^{2)} \le M_2 \text{ for some } M_2.$$

Similarly we can show that

$$\lim_{k\to\infty} \sum_{j=1}^{k} d^2_{i,j} \le M_2 \quad \text{for all } i.$$

Thus we have, if $k=o_p(n^{1/2})$,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{d^2_{i,j}}{k} \le M_2,$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{k-1} \tilde{\omega}_{T+i} x_{T+i+j} \xrightarrow{P} 0. \qquad \text{Q.E.D.}$$

We have proved that $k=o(n^{1/2})$ satisfied conditions of the asymptotic normality of the interpolated LSE are satisfied only for AR(1) models. Even though we can not derive the conditions for general AR(h) models because of complexity, we would conjecture that same condition satisfies the asymptotic normality for AR(h) models. Thus we will investigate the performance of the interpolated LSE for AR(2) model by the simulation study in the following section. To investigate the small sample properties of the interpolated LSE, we also perform the simulation study.

4. Estimation Procedure and Comparison with Y-W type Estimator

Now, we construct the estimation procedure of the interpolated LSE, perform simulation study to investigate small sample properties, and compare with the Y-W type estimator by the simulation study. After observing the nature of the interpolated LSE, we build the estimation procedure by adopting the concept of the EM algorithm, which estimates the parameters and the missing observations iteratively as follow:

Step 1. Compute $\hat{x}_M^I = \hat{x}_M + (D(I_k+E'E)^{-1}T_k)'(x_A - \hat{x}_A)$, which is the same as (2.1).

Step 2. Compute the LSE of the interpolated series.

Given a starting value for π, we should iterate two steps until convergence. The nature of the starting value for π depends on the data set as well as on π. Hence for a starting value π_0, we propose two estimators as follow:

(1) Compute π_0 in a long segment of the observed series with no contaminated observations which can be used to identify and estimate a model for the underlying process.

(2) Compute the LSE with replacing missing observations with the mean value(mean LSE)

Dunsmuir and Robinson(1981a) considered the Y-W type estimator for AR models and showed the central limit theorem and Dunsmuir and Robinson(1981b) obtained the estimator that maximized the Gaussian likelihood of the amplitude modulated process using Newton-Raphson method although clearly the assumption of Gaussianity was false. Many authors adopt the former because the former needs less computing time than the latter and is less complicated. We compare the interpolated LSE with the Y-W type estimator, because it is widely used and we can't compare with the procedures of Dunsmuir and Robinson(1981b), Reinsel and Wincek(1987) and Kohn and Ansley(1986) without their programs.

For simulation, we generate random numbers with mean = 0 and $\sigma_{\varepsilon}^2 = 1$ using the IMSL on CYBER and appropriate series. To eliminate effects of initial observations, we discard the first 100 observations. We perform 500 trials to obtain the mean squared error(MSE), $E(\widehat{\pi}_i - \pi_i)^2$ for each cases. For all Tables, MSE_1 and MSE_2 represent 100 times MSEs of the interpolated LSE, and the Y-W type estimator respectively.

Table 4.1 MSEs of the Interpolated LSE and the Yule-Walker type Estimator in AR(1) when the number of observation is 100

Number of Missing obs.	MSEs	Values of π_1					
		-0.9	-0.6	-0.3	0.3	0.6	0.9
1	MSE_1	.239	.675	.916	.951	.623	.243
	MSE_2	.287	.705	.912	.946	.649	.269
2	MSE_1	.301	.758	.920	.973	.706	.295
	MSE_2	.354	.814	.915	.968	.736	.327
4	MSE_1	.211	.640	.954	.999	.665	.267
	MSE_2	.288	.671	.954	.982	.670	.343

Table 4.2 MSEs of the Interpolated LSE and the Yule-Walker type Estimator in AR(1) when the number of missing observation is 1

Number of Observations	MSEs	Values of π_1					
		-0.9	-0.6	-0.3	0.3	0.6	0.9
50	MSE_1	.595	1.419	1.892	1.947	1.262	.583
	MSE_2	.703	1.564	1.876	1.925	1.360	.674
75	MSE_1	.390	.975	1.194	1.269	.740	.469
	MSE_2	.470	1.035	1.196	1.266	.770	.524
100	MSE_1	.239	.675	.916	.951	.623	.243
	MSE_2	.287	.705	.912	.946	.649	.269
150	MSE_1	.169	.422	.583	.588	.459	.163
	MSE_2	.186	.428	.580	.588	.469	.176

Table 4.1 contains the MSEs when the series follow AR(1), the number of observations(N) = 100 and the number of missing observations(M) = 1, 2 and 4. From Table 4.1, MSE_1 is smaller than MSE_2 for almost all cases except near $\pi_1 = 0$, in which two MSEs are almost same. Because the values of MSE_1 are very small, we can see that the interpolated LSE works well for small k. In Table 4.2, MSE_1 and MSE_2 are computed when the series follow AR(1), N = 50, 75, 100 and 150 and M = 1. We have the same phenomena as Table 4.1. For AR(2) models, we perform simulation when $\pi_1 = 1.0$ and $\pi_2 = -0.5$. Table 4.3 and Table 4.4 contain MSEs of π_1 and π_2 respectively when N = 50, 75, 100 and 150 and M = 1, 2 and 4. From Table 4.3 and Table 4.4, we can see that MSE_1 is smaller than MSE_2 for all cases. Even though we don't compare for all parameter space, we could conjecture that MSE_1 would be smaller than MSE_2 from Table 4.3 and Table 4.4.

From these simulations, we have the conclusion that the interpolated LSE has better performance in the sense of MSE than the Y-W type estimator.

Table 4.3 MSEs of the Interpolated LSE and the Yule-Walker type Estimator for π_1 of AR(2) with $\pi_1 = 1.0$ and $\pi_2 = -0.5$

Number of Missing obs.	MSEs	Number of Observations			
		50	75	100	150
1	MSE_1	1.622	1.071	.740	.488
	MSE_2	2.572	1.428	.904	.560
2	MSE_1	1.680	1.087	.736	.492
	MSE_2	2.364	1.397	.895	.570
4	MSE_1	1.835	1.167	.784	.532
	MSE_2	2.473	1.397	.895	.570

Table 4.4 MSEs of the Interpolated LSE and the Yule-Walker type Estimator for π_2 of AR(2) with $\pi_1 = 1.0$ and $\pi_2 = -0.5$

Number of Missing obs.	MSEs	Number of Observations			
		50	75	100	150
1	MSE_1	1.546	.937	.756	.542
	MSE_2	2.282	1.114	.878	.587
2	MSE_1	1.597	.939	.759	.534
	MSE_2	1.987	1.078	.852	.568
4	MSE_1	1.814	1.043	.823	.586
	MSE_2	2.137	1.085	.859	.609

References

[1] Abraham, B. (1981). Missing observations in time series, *Communications in Statistics, Theory and Methods*, A, 10, 1643-1654.

[2] Ansley, C.F., and Kohn, R. (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, *The Annals of Statistics*, 13, 1286-1316.

[3] Box, G.E.P., and Jenkins, G. (1976). *Time Series Analysis : forecasting and control*, 2nd ed. Holden-Day, New York.

[4] Brubacher, S.R. and Wilson, G.T. (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand, *Applied Statistics*, 25, 107-116.

[5] Bruce, A.G., and Martin, R. D. (1989). Leave- k-out diagnostics for time series (with discussion), *Journal of the Royal Statistical Society, Ser.* B, 51, 375-424.

[6] Cho, S., Ryu, G.Y., Park, B.U. and Lee, J.J. (1994). Outlier Detextion Diagnostic based on Interpolation Method in Autoregressive Models, *Journal of the Korean Statistical Society*, 22, 283-306

[7] Dunsmuir, W., and Robinson, P. M. (1981a). Asymptotic theory for time series containing missing and amplitude modulated observations, *Sankhyǎ*, A, 43, 260-281.

[8] Dunsmuir, W., and Robinson, P.M. (1981b). Estimation of time series models in the presence of missing data, *Journal of the American Statistical Association*, 76, 560-568.

[9] Fuller, W.A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

[10] Grenander, U., and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Wiley, New York.

[11] Harvey, A.C., and Pierse, R. G. (1984). Estimating missing observations in economic time series, *Journal of the American Statistical Association*, 79, 125-131.

[12] Johnson N. L. and Kotz S. (1970). *Continuous univariate distributions-2,* Houghton Mifflin, Boston.

[13] Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, 22, 389-395.

[14] Kohn, R. and Ansley, C.F. (1986). Estimation, Prediction, and Interpolation for ARIMA Models With Missing Data, *Journal of the American Statistical Association*, 81, 751-761.

[15] Lee, J.J. (1990). A Study on Influential Observations in Linear Regression and Time Series, *Unpublished Ph.D. dissertation, University of Wisconsin, Department of Statistics*.

[16] Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data* Wiley, New York.

[17] Ljung, G.M. (1989). Outliers and Missing Observations in Time Series, *ASA Proceedings of the Business and Economic Statistics Section*, 391-401.

[18] Miller, R.B. and Ferreiro (1984). A strategy to complete a time series with missing observations, *In: Proceedings of time series analysis of irreguraly observed data* (Parzen, E., ed.), Springer-Verlag, New York.

[19] Newbold, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process, *Biometrica*, 61, 423-426.

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

[21] Peña, D. and Tiao, G.C. (1991). A Note on Likelihood Estimation of Missing Values in Time Series, *The American Statistician,* 45, 212-213.

[22] Pourahmadi, M.(1989). Estimation and interpolation of missing values of stationary time series, *Journal of Time Series Analysis*, 10, 149-169.

[23] Reinsel, G.C., and Wincek, M. A. (1987). Asymtotic distribution of parameter estimators for nonconsecutively observed time series, *Biometrika*, 74, 115-124.

[24] Ryu, G.Y. (1991). Outlier Detection Diagnostic in Time Series, *Unpublished Ph.D. dissertation, Seoul National University, Department of Computer Science and Statistics.*

[25] Wei, W.S. (1990). *Time Series Analysis*, Addison-Wesley, Redwood City, CA.