Ann. Inst. Statist. Math.
Vol. 55, No. 1, 197–216 (2003)

L_1 LINEAR INTERPOLATOR FOR MISSING VALUES IN TIME SERIES*

ZUDI LU[1] AND Y. V. HUI[2]

[1]*Institute of Systems Science, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100080, China*
[2]*Department of Management Sciences, City University of Hong Kong, Academic Building, 83 Tat Chee Avenue, Kowloon, Hong Kong, China,* e-mail: msyervan@cityu.edu.hk

(Received January 17, 2001; revised January 18, 2002)

Abstract. We propose a minimum mean absolute error linear interpolator (MMAELI), based on the L_1 approach. A linear functional of the observed time series due to non-normal innovations is derived. The solution equation for the coefficients of this linear functional is established in terms of the innovation series. It is found that information implied in the innovation series is useful for the interpolation of missing values. The MMAELIs of the AR(1) model with innovations following mixed normal and t distributions are studied in detail. The MMAELI also approximates the minimum mean squared error linear interpolator (MMSELI) well in mean squared error but outperforms the MMSELI in mean absolute error. An application to a real series is presented. Extensions to the general ARMA model and other time series models are discussed.

Key words and phrases: Autoregressive process, innovation departure, linear interpolation, minimum mean absolute error, missing values.

1. Introduction

Numerous efforts have contributed to the interpolation of missing values as well as the estimation of model parameters based on maximum likelihood (ML) methods and least squares (LS) procedures in time series analysis. Parzen (1984) gives a comprehensive review of the earlier developments. Other research includes Dunsmuir and Robinson (1981), Gómez and Maravall (1994), Harvey and Pierce (1984), Jones (1980), Kohn and Ansley (1986), Ljung (1982), Peña and Tiao (1991), Penzer and Shea (1997) and Wincek and Reinsel (1986), for the ML methods, and Abraham (1981), Beveridge (1992), Damsleth (1980), Ferreiro (1987), Ljung (1989), and Luceño (1997), for the LS procedures. A recent review can be found in Dagum *et al.* (1998).

Besides missing observations, the time series data are possibly contaminated by outliers or are heterogenous. In addition, a heavy-tailed phenomenon relative to normality often emerges in the observed data set. If the interpolation of missing values and the estimation of model parameters are heavily dependent on some atypical observations, then the forecasts based on extrapolation from the observed samples would be poor. It could be expected that for incomplete time series observations with non-normal distribution, the normality-based ML and the LS procedures would retain poor performance when atypical points exist. They would induce an inaccurate interpolation of missing

*This research was supported by a CityU Research Grant and Natural Science Foundation of China.

values, leading to poor parameter estimates and bad forecasts. In this paper, we restrict ourselves to the interpolation of missing values.

Assume that we observe a discrete-time series $\{y_t\}$ at times $1 = t_1 < t_2 < \cdots < t_m = n$, where t_i, $i = 1, \ldots, m(\leq n)$ are positive integers. If $m < n$, then the series are observed irregularly, with missing values, of sample size m. When the series $\{y_t\}$ is stationary and only one missing observation y_τ exists, then the optimal least squares linear estimate of y_τ is given by (see Grenander and Rosenblatt ((1957), p. 83) and Whittle (1963))

$$\hat{y}_\tau = \mu - \sum_{j=1}^{\infty} \rho_j \{(y_{\tau-j} - \mu) + (y_{\tau+j} - \mu)\}, \tag{1.1}$$

where μ is the series mean and ρ's are the inverse autocorrelations. When the series is Gaussian, (1.1) equals the minimum mean square error (MMSE) interpolator defined by

$$\hat{y}_\tau = \operatorname{argmin} E[(y_\tau - c)^2 \mid y_s, s \neq \tau], \tag{1.2}$$

where c takes values in the σ-field generated by $\{y_s, s \neq \tau\}$. For the non-Gaussian case, (1.1) usually does not equal (1.2) and it is often difficult to calculate the MMSE interpolator. However, in any case, (1.1) is the minimum mean square error linear interpolator (MMSELI) which minimizes

$$MSE(c) = E[(y_\tau - c)^2] = E\{E[(y_\tau - c)^2 \mid y_s, s \neq \tau]\} \tag{1.3}$$

among the class of linear functions of the observed series, $\{y_s, s \neq \tau\}$, where c takes the form of $\sum_{s \neq \tau} c_s y_s$ with c_s's being real constants. The procedure is extended to a single gap with consecutive missing values by Brubacker and Wilson (1976), and to more general irregular spaces by Beveridge (1992).

Our aim is to explore the (robust) interpolation of missing values in time series when the innovations are non-normally distributed. It is well accepted that the L_1 rule is a good alternative to, and more robust than, the ML and LS rules. In Section 2, we will propose a minimum mean absolute error linear interpolator (MMAELI), *minimizing the mean absolute error* (MAE). Section 3 focuses on the AR(1) models with one missing observation. Section 4 gives the specific solution equations for the coefficients of the linear functional when the innovations follow mixed normals and t distributions respectively. An illustrative example is considered in Section 5, where the L_2 a pproach is shown to lead to a 'bad' interpolated value in comparison with the L_1 approach. The extension to the general AR(p) models is presented in Section 6. Section 7 discusses some further problems on multiple missing values and general ARMA model case. Complex proofs are relegated to the Appendix.

2. Mean absolute error and L_1 linear interpolator

Let $\{y_t\}$ be a stationary discrete-time series, and denote by τ the time period at which the series is not observed. That is, the observed series is $y_1, \ldots, y_{\tau-1}, y_{\tau+1}, \ldots, y_n$ and is denoted by $\{y_{\tau+s}, s \neq 0\}$. Assume that $1 < \tau < n$ and the series mean $\mu = 0$.

We interpolate y_τ by a functional of the observed series and denote by $\hat{y}_\tau$ the interpolator of y_τ. Although the MSE rule has been widely used to measure the closeness of the interpolation, it is less robust. We suggest using the MAE to measure the closeness of the interpolation.

If we can find a functional of the observed series, $\hat{y}_T = c(y_{T+s}, s \neq 0)$, at which the *mean absolute error* conditional on $\{y_{T+s}, s \neq 0\}$

$$MAE_c(c) = E[|y_T - c| \mid y_{T+s}, s \neq 0] \tag{2.1}$$

is minimized almost surely (a.s. P) among all the measurable functions of $\{y_{T+s}, s \neq 0\}$, then $\hat{y}_T = c(y_{T+s}, s \neq 0)$ will be called the *minimum mean absolute error* (*MMAE*) *interpolator* or the *least absolute deviation interpolator*. This interpolator is the conditional median of y_T given $\{y_{T+s}, s \neq 0\}$. The calculation of conditional median is usually very difficult for series with a general distribution. Here we propose a linear interpolator based on the MMAE rule.

DEFINITION 1. Let $\{y_t\}$ be a time series with a finite first absolute moment. If there is a linear functional of the form

$$\hat{y}_T^{L1} = \sum_{s \neq 0} c_s y_{T+s} \tag{2.2}$$

where c_s's are real constants such that $\hat{y}_T^{L1}$ minimizes the conditional mean absolute error $MAE_c(c)$ almost surely among all the linear functionals in (2.2), we call $\hat{y}_T^{L1}$ the *minimum mean absolute error linear interpolator* (*MMAELI*) or the *L_1 linear interpolator* of y_T.

Let $\mathcal{L}_{-T} = \{\sum_{s \neq 0} c_s y_{T+s} : c_s \in R \text{ for } s \neq 0\}$ be the linear space of $\{y_{T+s}, s \neq 0\}$ and $\sigma_{-T} = \sigma(y_{T+s}, s \neq 0)$ be the σ-field generated by $\{y_{T+s}, s \neq 0\}$. The MMAELI of y_T can be redefined mathematically as the $\hat{y}_T^{L1} \in \mathcal{L}_{-T}$ such that

$$MAE_c(\hat{y}_T^{L1}) \leq MAE_c(c) \quad \text{a.s.} \quad P \text{ for any } c \in \mathcal{L}_{-T}. \tag{2.3}$$

Remark 1. Given that the unconditional mean absolute error $MAE(c) = E[|y_T - c|] = E[MAE_c(c)]$ for any $c \in \mathcal{L}_{-T}$, (2.3) can be rewritten as

$$MAE(\hat{y}_T^{L1}) \leq MAE(c) \quad \text{for any} \quad c \in \mathcal{L}_{-T}. \tag{2.3$'$}$$

This shows that $\hat{y}_T^{L1}$ is the optimal linear interpolator with respect to the unconditional MAE.

Similar to the properties of L_1 estimators vs. L_2 estimators in the literature, the MMAELI has certain advantages compared with the MMSELI.

PROPOSITION 2.1. *Denote by $\hat{y}_T^{L2}$ the MMSELI of y_T in* (1.3). *Then*

$$MAE(\hat{y}_T^{L1}) \leq MAE(\hat{y}_T^{L2}) \leq MSE^{1/2}(\hat{y}_T^{L2}) \leq MSE^{1/2}(\hat{y}_T^{L1}). \tag{2.4}$$

PROOF. It follows from the definitions of $\hat{y}_T^{L1}$ and $\hat{y}_T^{L2}$ and the Schwarz's inequality.

Remark 2. (2.4) shows that the MAE of the MMAELI is the smallest among the MAE's and the SMSE's of both MMAELI and MMSELI where SMSE is the square root of the MSE.

3. Characterization of MMAELI in AR(1) model

For simplicity, we first consider the AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{3.1}$$

with $|\phi| < 1$, $\{\varepsilon_t\}$ being an i.i.d. innovation process with a finite first absolute moment and ε_t is independent of $\{y_s, s < t\}$. Assume that there is only one missing value at time $t = \tau$.

If $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = \sigma_\varepsilon^2 < \infty$, then it is well known that the MMSELI of y_τ is

$$\hat{y}_\tau^{L2} = \frac{\phi}{1+\phi^2}[y_{\tau+1} + y_{\tau-1}]. \tag{3.2}$$

The MMSELI and the MMAELI are the same when the series is normally distributed. We investigate the computation and properties of the MMAELI under non-normal innovations.

Let $y_{ob} = (y_1, \ldots, y_{\tau-1}, y_{\tau+1}, \ldots, y_n)$ be the observed sample. The conditional density function of y_τ $(\tau > 1)$ given y_{ob} is

$$p(y_\tau \mid y_{ob}) = \frac{p(y_\tau, y_{ob})}{p(y_{ob})} = \frac{p_\varepsilon(y_\tau - \phi y_{\tau-1})p_\varepsilon(y_{\tau+1} - \phi y_\tau)}{\int p_\varepsilon(u - \phi y_{\tau-1})p_\varepsilon(y_{\tau+1} - \phi u)du} \tag{3.3}$$

where $p_\varepsilon(\cdot)$ is the density functions of ε_t. Hence, $y_\tau \mid y_{ob}$ depends only on $(y_{\tau+1}, y_{\tau-1})$ in the observed series. Thus, $p(y_\tau \mid y_{ob}) = p(y_\tau \mid y_{\tau+1}, y_{\tau-1})$, and the linear interpolator of y_τ defined in (2.2) is reduced to the form

$$\hat{y}_\tau^{L1} = c_1 y_{\tau+1} + c_2 y_{\tau-1} \tag{3.4}$$

with c_1 and c_2 being two real constants.

Now our task is to determine c_1 and c_2 in (3.4) according to the MMAE rule of (2.3). Let

$$u_\tau = y_{\tau+1} - \phi^2 y_{\tau-1}, \qquad v_\tau = y_\tau - \frac{\phi}{1+\phi^2}[y_{\tau+1} + y_{\tau-1}]. \tag{3.5}$$

Together with (3.1) and (3.5), it gives

$$u_\tau = \phi\varepsilon_\tau + \varepsilon_{\tau+1}, \qquad v_\tau = \frac{1}{1+\phi^2}\varepsilon_\tau - \frac{\phi}{1+\phi^2}\varepsilon_{\tau+1}. \tag{3.6}$$

Hence (3.3) can be expressed as

$$\begin{aligned} p(v_\tau \mid y_{\tau+1}, y_{\tau-1}) &= p(v_\tau \mid u_\tau, y_{\tau-1}) = p(v_\tau \mid u_\tau) \\ &= \frac{p_\varepsilon\left(\frac{1}{1+\phi^2}u_\tau - \phi v_\tau\right) p_\varepsilon\left(v_\tau + \frac{\phi}{1+\phi^2}u_\tau\right)}{\int p_\varepsilon\left(\frac{1}{1+\phi^2}u_\tau - \phi v\right) p_\varepsilon\left(v + \frac{\phi}{1+\phi^2}u_\tau\right) dv}. \end{aligned} \tag{3.7}$$

If the MMAELI, $\hat{v}_\tau^{L1}$, of v_τ based on $(y_{\tau+1}, y_{\tau-1})$ is derived, then (3.4) can be obtained from (3.5) and $\hat{v}_\tau^{L1}$. That is,

$$\hat{y}_\tau^{L1} = \hat{v}_\tau^{L1} + \frac{\phi}{1+\phi^2}[y_{\tau+1} + y_{\tau-1}]. \tag{3.8}$$

From (3.7), $\hat{v}_T^{L1}$ equals the MMAELI of v_T based on u_T. Hence, $\hat{v}_T^{L1} = c_0 + d_0 u_T$, where c_0 and d_0 are real constants which minimize $MAE_c(c,d) = E[|v_T - c - du_T| \mid u_T]$. It is clear that $c_0 = 0$ since $Ev_T = Eu_T = 0$. Our next step is to determine d_0 which minimizes

$$\widetilde{MAE}(d) = E[MAE_c(0,d)] = E[|v_T - du_T|]. \tag{3.9}$$

Applying Theorem 2.1 in Pinkus ((1989), p. 14), we deduce that d_0 minimizes (3.9) if and only if it satisfies

$$|E[\operatorname{sgn}(v_T - d_0 u_T)u_T]| \le E[I_{\{v_T - d_0 u_T = 0\}}|u_T|]. \tag{3.10}$$

From (3.6),

$$v_T - du_T = \frac{a\varepsilon_T - b\varepsilon_{T+1}}{1+\phi^2}, \tag{3.11}$$

where $a = a(d) = 1 - \phi(1+\phi^2)d$ and $b = b(d) = \phi + (1+\phi^2)d$.

If $\phi = 0$, then $a = 1$ and $b = d$. Hence the independence between ε_T and ε_{T+1} together with (3.9) and (3.11) gives $d_0 = 0$. Assume that $\phi \ne 0$. If the distribution of ε_T is non-degenerate, then $P(v_T - d_0 u_T = 0) = 0$, from (3.11). From (3.10),

$$E[\operatorname{sgn}(v_T - d_0 u_T)u_T] = 0. \tag{3.12}$$

Set $a_0 = a(d_0)$ and $b_0 = b(d_0)$. (3.6) and (3.11) deduce that the LHS of (3.12) equals

$$\begin{aligned}\phi E\varepsilon_T I_{\{a_0\varepsilon_T > b_0\varepsilon_{T+1}\}} &+ E\varepsilon_{T+1} I_{\{a_0\varepsilon_T > b_0\varepsilon_{T+1}\}} - \phi E\varepsilon_T I_{\{a_0\varepsilon_T < b_0\varepsilon_{T+1}\}} \\ &- E\varepsilon_{T+1} I_{\{a_0\varepsilon_T < b_0\varepsilon_{T+1}\}}.\end{aligned} \tag{3.13}$$

Now if $b_0 = 0$, then $\phi E\varepsilon_T \operatorname{sgn}(a_0\varepsilon_T) = 0$ from (3.12) and (3.13). Hence $a_0 = 0$ for $\phi \ne 0$. It is impossible that $a_0 = b_0 = 0$, for $1+\phi^2$ would be 0 otherwise. Hence $b_0 \ne 0$. Similarly, $a_0 \ne 0$. Using the i.i.d. property of ε_t with mean 0, it follows from (3.12) and (3.13) that

$$\phi E\left[\varepsilon_t F_\varepsilon\left(\frac{a_0}{b_0}\varepsilon_t\right)\right]\operatorname{sgn}(b_0) - E\left[\varepsilon_t F_\varepsilon\left(\frac{b_0}{a_0}\varepsilon_t\right)\right]\operatorname{sgn}(a_0) = 0, \tag{3.14}$$

where $F_\varepsilon(\cdot)$ is the cumulative distribution function of ε_t.

We have the following result on the MMAELI, $\hat{y}_T^{L1}$, of y_T for the AR(1) model.

PROPOSITION 3.1. *If the i.i.d. innovation process ε_t has a non-degenerate distribution, $F_\varepsilon(\cdot)$, whose density function, $p_\varepsilon(\cdot)$, exists and has a first absolute moment with mean 0, then*

$$\hat{y}_T^{L1} = d_0[y_{T+1} - \phi^2 y_{T-1}] + \frac{\phi}{1+\phi^2}[y_{T+1} + y_{T-1}] = \frac{\phi a_0 y_{T-1} + b_0 y_{T+1}}{1+\phi^2}. \tag{3.15}$$

Here $d_0 = 0$ if $\phi = 0$, and d_0 is the solution to (3.14) with $a_0 = a(d_0)$ and $b_0 = b(d_0)$ if $\phi \ne 0$.

Remark 3. (a) We conjecture that d_0 is unique under mild conditions. Firstly, the minimizer of $MSE(d) = E|v_T - du_T|^2$ is unique and equals 0. This suggests that the minimizer d_0 of (3.9) might be unique similarly. Secondly, d_0 is unique under mild conditions when $\phi = 0$. In fact, when $\phi = 0$, from (3.12) and (3.13) together with $a_0 = 1$ and $b_0 = d_0$, (3.12) reduces to $A(d_0) = 0$, where $A(d_0) = E[\varepsilon_t F_\varepsilon(d_0\varepsilon_t)]$. If the derivative with respect to d_0 and the expectation in $A(d_0)$ are exchangeable, then $A'(d_0) = E[\varepsilon_t^2 p_\varepsilon(d_0\varepsilon_t)] > 0$. Hence $d_0 = 0$ is the unique solution to $A(d_0) = 0$. Thirdly, our computational experience in Sections 4 and 5 for non-zero ϕ also indicates that our conjecture might be true. Since the general case $\phi \neq 0$ leads to a complex equation (3.14), this conjecture remains open.

(b) Note that $|\phi| < 1$ is not required in the derivation. The assumption is that the density $p_0(y)$ of y_0 exists. Proposition 3.1 applies to the non-stationary ($\phi = \pm 1$) series. Furthermore, $p(v_T \mid u_T) = p_\varepsilon(u_T/2 - v_T)p_\varepsilon(u_T/2 + v_T)$ is a symmetric function of v_T when $\phi = 1$. Hence $d_0 = 0$ correspondingly. Also $a_0 = 1 + 2d_0$ and $b_0 = -1 + 2d_0$ when $\phi = -1$. It follows from (3.14) that

$$-E\left[\varepsilon_t F_\varepsilon\left(\frac{1+2d_0}{-1+2d_0}\varepsilon_t\right)\right]\operatorname{sgn}(-1+2d_0) - E\left[\varepsilon_t F_\varepsilon\left(\frac{-1+2d_0}{1+2d_0}\varepsilon_t\right)\right]\operatorname{sgn}(1+2d_0) = 0,$$

and $d_0 = 0$ is the solution. This is the reason why we get back to MMSELI when $\phi = \pm 1$. Note that $p_\varepsilon(\cdot)$ is not assumed to be symmetric in this remark and Proposition 3.1.

COROLLARY 3.1. *Under the condition of Proposition* 3.1,

$$MAE(\hat{y}_T^{L1}) = 2h\left(\frac{a_0}{b_0}\right)\operatorname{sgn}(b_0) \quad \textit{and} \tag{3.16}$$

$$MAE(\hat{y}_T^{L2}) = \frac{2}{1+\phi^2}\left[h\left(\frac{1}{\phi}\right)\operatorname{sgn}(\phi) + \phi h(\phi)\right],$$

where $h(\alpha) = E[\varepsilon_t F_\varepsilon(\alpha\varepsilon_t)]$. *If* ε_t *has a finite second order moment, then*

$$MSE(\hat{y}_T^{L1}) = \frac{1+(1+\phi^2)^2 d_0^2}{1+\phi^2}\sigma_\varepsilon^2 \quad \textit{and} \quad MSE(\hat{y}_T^{L2}) = \frac{\sigma_\varepsilon^2}{1+\phi^2}. \tag{3.17}$$

PROOF. The result follows from

$$y_T - \hat{y}_T^{L1} = v_T - d_0 u_T = \frac{a_0\varepsilon_T - b_0\varepsilon_{T+1}}{1+\phi^2} \quad \text{and} \quad y_T - \hat{y}_T^{L2} = v_T = \frac{\varepsilon_T - \phi\varepsilon_{T+1}}{1+\phi^2}.$$

COROLLARY 3.2. *In Proposition* 3.1, *if* $p_\varepsilon(\cdot)$ *is further assumed to be symmetric and* d_0 *is a real solution of* (3.14) *for* ϕ, *then* $-d_0$ *is the real solution of* (3.14) *with* $-\phi$ *replaced by* ϕ.

PROOF. This is clear by noting that

$$a_0 = 1 - \phi(1+\phi^2)d_0 = 1 - (-\phi)(1+\phi^2)(-d_0), \qquad -b_0 = (-\phi) + (1+\phi^2)(-d_0),$$

and $\phi\operatorname{sgn}(b_0) = (-\phi)\operatorname{sgn}(-b_0)$ as well as $F_\varepsilon(-x) = 1 - F_\varepsilon(x)$.

Therefore, only the calculation of d_0 for $\phi > 0$ is required when the innovation variable ε_t is symmetric.

4. Some typical non-normality innovations

In the case of non-normality, the calculation of the MMAELI depends on the determination of d_0 in Proposition 3.1. If $d_0 \neq 0$, the MMAELI differs from the MMSELI. Once the innovation distribution is assumed, d_0 can be determined from (3.14). In this section, we derive specific solution equations of d_0 for some non-normal distributions.

4.1 *Mixed normal innovations*

We first consider the AR(1) model with innovations that have a mixed normal distribution

$$F_\varepsilon(x) = (1-\delta)\Phi\left(\frac{x-\mu_1}{\sigma_1}\right) + \delta\Phi\left(\frac{x-\mu_2}{\sigma_2}\right), \tag{4.1}$$

where $(1-\delta)\mu_1 + \delta\mu_2 = 0$. If $\mu_1 = \mu_2 = 0$, then (4.1) is the contaminated normal distribution.

From (3.14), the first step in determining d_0 is to calculate the expectation

$$h(\alpha) = E[\varepsilon_t F_\varepsilon(\alpha\varepsilon_t)]. \tag{4.2a}$$

For the mixed normal in (4.1), ε_t has second order moment. Then (4.2a) can be calculated in the following way with $h(0) = E[\varepsilon_t F_\varepsilon(0)] = 0$.

$$h(\alpha) = \int_0^\alpha h'(u)du, \qquad \text{and} \qquad h'(u) = E[\varepsilon_t^2 p_\varepsilon(u\varepsilon_t)]. \tag{4.2b}$$

From (4.1), $p_\varepsilon(x) = \frac{1-\delta}{\sigma_1}\varphi(\frac{x-\mu_1}{\sigma_1}) + \frac{\delta}{\sigma_2}\varphi(\frac{x-\mu_2}{\sigma_2})$, where $\varphi(x)$ is the standard normal density function. (4.2a) and (4.2b) give

$$\begin{aligned} h'(u) = &\left(\frac{1-\delta}{\sigma_1}\right)^2 g(u;\mu_1,\sigma_1,\mu_1,\sigma_1) \\ &+ \frac{(1-\delta)\delta}{\sigma_1\sigma_2}[g(u;\mu_1,\sigma_1,\mu_2,\sigma_2) + g(u;\mu_2,\sigma_2,\mu_1,\sigma_1)] \\ &+ \left(\frac{\delta}{\sigma_2}\right)^2 g(u;\mu_2,\sigma_2,\mu_2,\sigma_2), \end{aligned} \tag{4.3a}$$

where

$$\begin{aligned} &g(u;\mu_1,\sigma_1,\mu_2,\sigma_2) \\ &= \frac{(\sigma_1^2+\mu_1^2)\sigma_2^4u^2 + 2\sigma_1^2\sigma_2^2\mu_1\mu_2u + (\sigma_2^2+\mu_2^2)\sigma_1^4}{\sqrt{2\pi}(\sigma_1^2+\sigma_2^2u^2)^{5/2}} e^{-(\mu_1-\mu_2u)^2/2(\sigma_1^2+\sigma_2^2u^2)}. \end{aligned} \tag{4.3b}$$

Thus it follows from (4.2) that

$$\begin{aligned} h(\alpha) = &\left(\frac{1-\delta}{\sigma_1}\right)^2 G(\alpha;\mu_1,\sigma_1,\mu_1,\sigma_1) \\ &+ \frac{(1-\delta)\delta}{\sigma_1\sigma_2}[G(\alpha;\mu_1,\sigma_1,\mu_2,\sigma_2) + G(\alpha;\mu_2,\sigma_2,\mu_1,\sigma_1)] \end{aligned} \tag{4.4a}$$

$$+\left(\frac{\delta}{\sigma_2}\right)^2 G(\alpha;\mu_2,\sigma_2,\mu_2,\sigma_2),$$

where

$$G(\alpha;\mu_1,\sigma_1,\mu_2,\sigma_2)=\int_0^\alpha g(u;\mu_1,\sigma_1,\mu_2,\sigma_2)du. \tag{4.4b}$$

Now combining Proposition 3.1 with (3.14) and (4.4), we have the following result.

THEOREM 4.1. *For the $AR(1)$ model with innovations having a mixed normal distribution* (4.1), *the MMAELI of y_T is given in* (3.16) *with d_0 satisfying the equation*

$$\phi h\left(\frac{a_0}{b_0}\right)\operatorname{sgn}(b_0)-h\left(\frac{b_0}{a_0}\right)\operatorname{sgn}(a_0)=0,$$

where $h(\cdot)$ is defined in (4.4) *and a_0 and b_0 are given in Proposition* 3.1.

Note that $\mu_1=\mu_2=0$ gives

$$h'(u)=\left(\frac{1-\delta}{\sigma_1}\right)^2 g\left(\frac{u}{\sigma_1},\frac{1}{\sigma_1}\right)+\frac{(1-\delta)\delta}{\sigma_1\sigma_2}\left[g\left(\frac{u}{\sigma_1},\frac{1}{\sigma_2}\right)+g\left(\frac{u}{\sigma_2},\frac{1}{\sigma_1}\right)\right] \tag{4.5a}$$

$$+\left(\frac{\delta}{\sigma_2}\right)^2 g\left(\frac{u}{\sigma_2},\frac{1}{\sigma_2}\right),$$

where $g(\frac{u}{a},\frac{1}{b})=\frac{1}{\sqrt{2\pi}}(\frac{u^2}{a^2}+\frac{1}{b^2})^{-3/2}$. We have

$$\int_0^\alpha g\left(\frac{u}{a},\frac{1}{b}\right)du=\frac{1}{\sqrt{2\pi}}\frac{b^2\alpha}{\sqrt{\dfrac{\alpha^2}{a^2}+\dfrac{1}{b^2}}}. \tag{4.5b}$$

It follows from (4.2b) and (4.5) that

$$h(\alpha)=\frac{\alpha}{\sqrt{2\pi}}\left\{\left(\frac{1-\delta}{\sigma_1}\right)^2\frac{\sigma_1^2}{\sqrt{\dfrac{\alpha^2}{\sigma_1^2}+\dfrac{1}{\sigma_1^2}}}+\frac{(1-\delta)\delta}{\sigma_1\sigma_2}\left[\frac{\sigma_2^2}{\sqrt{\dfrac{\alpha^2}{\sigma_1^2}+\dfrac{1}{\sigma_2^2}}}+\frac{\sigma_1^2}{\sqrt{\dfrac{\alpha^2}{\sigma_2^2}+\dfrac{1}{\sigma_1^2}}}\right]\right.$$

$$\left.+\left(\frac{\delta}{\sigma_2}\right)^2\frac{\sigma_2^2}{\sqrt{\dfrac{\alpha^2}{\sigma_2^2}+\dfrac{1}{\sigma_2^2}}}\right\}.$$

Hence

$$h\left(\frac{a_0}{b_0}\right)=a_0\operatorname{sgn}(b_0)H(a_0,b_0)\quad\text{and}\quad h\left(\frac{b_0}{a_0}\right)=b_0\operatorname{sgn}(a_0)H(b_0,a_0), \tag{4.6a}$$

where

$$H(a,b) = \frac{1}{\sqrt{2\pi}} \left\{ \frac{(1-\delta)^2\sigma_1}{\sqrt{a^2+b^2}} + (1-\delta)\delta \left[\frac{\sigma_2^2}{\sqrt{a^2\sigma_2^2+b^2\sigma_1^2}} + \frac{\sigma_1^2}{\sqrt{a^2\sigma_1^2+b^2\sigma_2^2}} \right] + \frac{\delta^2\sigma_2}{\sqrt{a^2+b^2}} \right\}. \tag{4.6b}$$

Combining Proposition 3.1 with (3.14) and (4.6b) gives the following result.

THEOREM 4.2. *For the $AR(1)$ model with contaminated normal innovations* (1.1), *the MMAELI of y_T is given in* (3.16) *with d_0 satisfying the equation*

$$\phi a_0 H(a_0, b_0) - b_0 H(b_0, a_0) = 0,$$

where a_0 and b_0 are specified in Proposition 3.1.

It follows from Corollary 3.1 that

$$MAE(\hat{y}_T^{L1}) = 2a_0 H(a_0, b_0), \qquad MAE(\hat{y}_T^{L2}) = \frac{2}{1+\phi^2}[H(1,\phi) + \phi^2 H(\phi, 1)],$$

and

$$MSE(\hat{y}_T^{L1}) = \frac{1+(1+\phi^2)^2 d_0^2}{1+\phi^2}\sigma_\varepsilon^2, \qquad MSE(\hat{y}_T^{L2}) = \frac{\sigma_\varepsilon^2}{1+\phi^2},$$

where $\sigma_\varepsilon^2 = (1-\delta)\sigma_1^2 + \delta\sigma_2^2$.

For contaminated normal innovations, numerical results for different ϕ, the ratio of σ_2 to σ_1, and δ, are tabulated in Table 1. Only d_0 corresponding to $\phi > 0$ is calculated due to symmetry. It can be seen that the difference between MMAELI and MMSELI becomes more and more significant with the increase of the ratio of σ_2 to σ_1 and the contaminated portion of δ. This difference is larger for $|\phi|$ close to 0.5 than for $|\phi|$ away from 0.5. Table 1 also shows that the increase in MSE between L_2 and L_1 is small compared with the decrease in MAE.

Table 1. Solutions of d for different ϕ with contaminated normal distribution ε_t.

$\frac{\sigma_2}{\sigma_1}$	δ	ϕ	d_0	ϕ	d_0	$\frac{SMSE(\hat{y}_T^{L1})-SMSE(\hat{y}_T^{L2})}{SMSE(\hat{y}_T^{L2})}$	$\frac{MAE(\hat{y}_T^{L2})-MAE(\hat{y}_T^{L1})}{MAE(\hat{y}_T^{L1})}$
2	0.1	0.1	−0.0104794	−0.1	0.0104794	0.005601101%	0.2414013%
		0.3	−0.0225983	−0.3	0.0225983	0.03033253%	1.57710%
		0.5	−0.0200576	−0.5	0.0200576	0.03142532%	2.497017%
		0.7	−0.0110414	−0.7	0.0110414	0.01353198%	2.184916%
		0.9	−0.00292215	−0.9	0.00292215	0.001398715%	0.8746421%
10	0.1	0.1	−0.079585442	−0.1	0.079585442	0.3225375%	4.574259%
		0.5	−0.246506865	−0.5	0.246506865	4.639682%	21.66369%
		0.9	−0.037129038	−0.9	0.037129038	0.2255615%	5.714213%
10	0.3	0.1	−0.082327435	−0.1	0.082327435	0.3451065%	5.325714%
		0.5	−0.283590239	−0.5	0.283590239	6.097201%	23.07139%
		0.9	−0.05514599	−0.9	0.05514599	0.4969095%	6.788403%

4.2 *Student's t innovations*

Here we consider the innovation process having a t distribution with k degrees of freedom. Note that neither MMAELI nor MMSELI applies for $k = 1$. However, MMAELI does apply while MMSELI does not for $k = 2$.

For $k = 2$, the cumulative distribution of t_2 is $F_2(x) = \int_{-\infty}^{x} f_2(u)du = \frac{1}{2}(1 + \frac{x}{\sqrt{2+x^2}})$. We have $h_2(\alpha) = E[\varepsilon_t F_2(\alpha\varepsilon_t)] = \alpha \int_0^\infty \frac{x^2}{\sqrt{2+\alpha^2 x^2}} f_2(x)dx$.

THEOREM 4.3. *For the* $AR(1)$ *model with* t_2 *innovations, the MMAELI of* y_τ *is given in* (3.15) *with* d_0 *satisfying the equation*

$$\phi a_0 H(a_0, b_0) - b_0 H(b_0, a_0) = 0, \tag{4.7}$$

where a_0 *and* b_0 *are specified in Proposition* 3.1, *and*

$$H(a,b) = \int_0^\infty \frac{x^2}{\sqrt{2b^2 + a^2 x^2}} f_2(x)dx. \tag{4.8}$$

The solution to (4.7) with (4.8) can be obtained numerically. Some results are reported in Table 2, where $Q(d_0)$ is the value of the LHS of (4.7). d_0 against ϕ is also plotted in Fig. 1.

If $k \geq 3$, then ε_t has second order moment. We calculate $h(\alpha)$ as in (4.2a,b).

Table 2. Solutions of d for different ϕ with $\varepsilon_t \sim t_2$.

ϕ	d_0	$Q(d_0)$	ϕ	d_0
0.1	−0.074725	-7.1252×10^{-7}	−0.1	0.074725
0.3	−0.1639724	1.22232×10^{-7}	−0.3	0.1639724
0.4	−0.17487	3.8705×10^{-7}	−0.4	0.17487
0.5	−0.1645972	-6.48049×10^{-7}	−0.5	0.1645972
0.7	−0.1020768	-2.6553×10^{-7}	−0.7	0.1020768
0.9	−0.0286801	1.00778×10^{-7}	−0.9	0.0286801

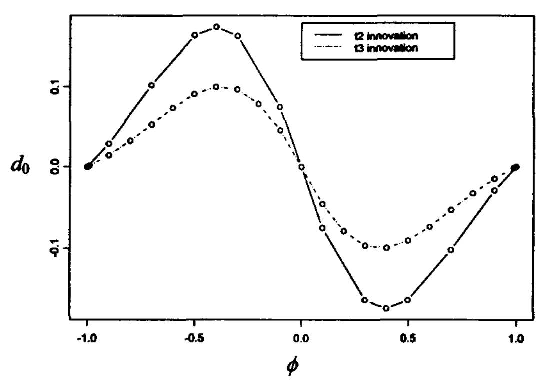

Fig. 1. d_0 against ϕ from Tables 2 and 3.

Table 3. Solutions of d for different ϕ with $\varepsilon_t \sim t_3$.

ϕ	d_0	ϕ	d_0
0.1	−0.0453601	−0.1	0.0453601
0.3	−0.0965324	−0.3	0.0965324
0.4	−0.0996965	−0.4	0.0996965
0.5	−0.0905989	−0.5	0.0905989
0.7	−0.052839	−0.7	0.052839
0.9	−0.0143892	−0.9	0.0143892

THEOREM 4.4. *For the* $AR(1)$ *model with* t_3 *innovations, the MMAELI of* y_τ *is given in* (3.15) *with* d_0 *satisfying the equation*

$$\phi a_0(|a_0| + 2|b_0|) - b_0(|b_0| + 2|a_0|) = 0, \tag{4.9}$$

where a_0 *and* b_0 *are specified in Proposition* 3.1.

The proof of this theorem is presented in Appendix A.

Numerical solutions of d_0 are given in Table 3 and are depicted in Fig. 1. It can be seen that d_0 against ϕ looks like the shape of a sine function. d_0's distinct from 0 emerges for $\phi \neq 0, \pm 1$ and are particularly marked especially for $|\phi|$ near 0.5. This phenomenon is also observed in Table 1 for contaminated normal innovations. Intuitively, the AR(1) process is mainly contributed to by the innovation for $\phi \approx 0$ and by the lag itself for $\phi \approx \pm 1$. Therefore, the combined contribution of the innovation and the process lag is comparatively less for $\phi \approx 0$ and ± 1 while it is much stronger for $\phi \approx \pm 0.5$. This is the intuitive reason why the difference between d_0 and 0 is more marked for $\phi \approx \pm 0.5$. Also d_0's that are different from 0 are more marked for the t_2 innovation than for the t_3 innovation.

5. An illustrative example

We consider the model presented in Wei ((1990), p. 107, Example 6.1) for the daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. For the 45 daily observations, Wei (1990) fitted an AR(1) model

$$(1 - 0.43B)(Z_t - 1.79) = a_t. \tag{5.1}$$

Here we examine the L_1 linear interpolation of the specified model.

The residuals, w1res, of the fitted model are shown in Fig. 2. Figure 3 gives the q-q plot of the standardized residuals, which indicates that the residual is not normally distributed. In Fig. 4, we depict the kernel densities of w1res using Scott and 1.2 times Scott bandwidths (c.f, Venables and Ripley (1994)). It shows that w1res seems to be distributed approximately as a mixed normal. Since the bimodal distribution is convenient for us considering the development in Section 4.1, the distribution of w1res is approximated by the mixed normal with $\delta = 0.0638615$, $\mu_1 = -0.0860311$, $\mu_2 = 1.26112$, $\sigma_1 = .324218$ and $\sigma_2 = .16677$. The density is shown in Fig. 4. This mixed normal density seems to fit the residuals, w1res, rather well. Figure 5 gives the q-q plot of w1res with respect to the samples of the fitted mixed normal density.

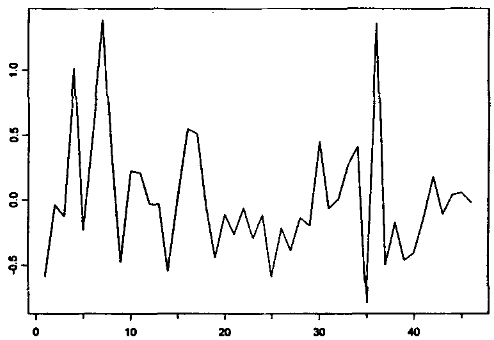

Fig. 2. Time series plot of w1res.

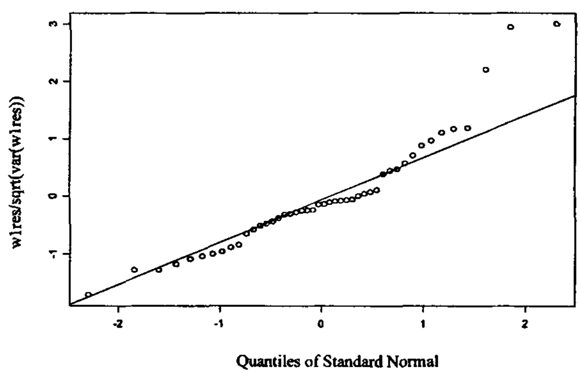

Fig. 3. qqnorm of w1res.

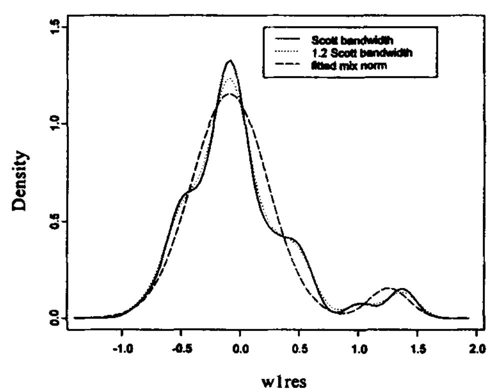

Fig. 4. Densities of w1res.

Based on the formula specified in Theorem 4.1, we calculate the solution for d_0 which equals -0.076745. Set $y_t = Z_t - 1.79$. The MMAELI of the model in (5.1) is

$$\begin{aligned}
\hat{Z}_T^{L1} &= 1.79 - 0.076745[y_{T+1} - 0.43^2 y_{T-1}] + \frac{0.43}{1+0.43^2}[y_{T+1} + y_{T-1}] \\
&= 1.79 + 0.3770892 y_{T-1} + 0.2861549 y_{T+1},
\end{aligned} \tag{5.2}$$

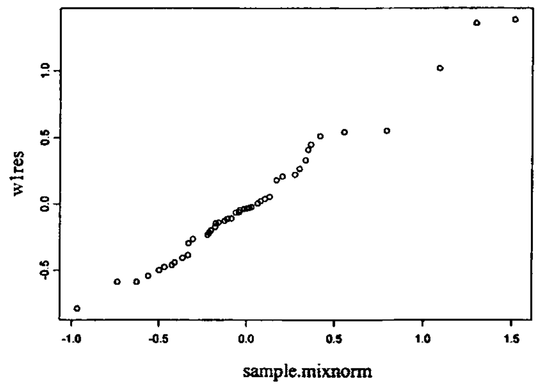

Fig. 5. qqplot of w1res w.r.t. mixed normal.

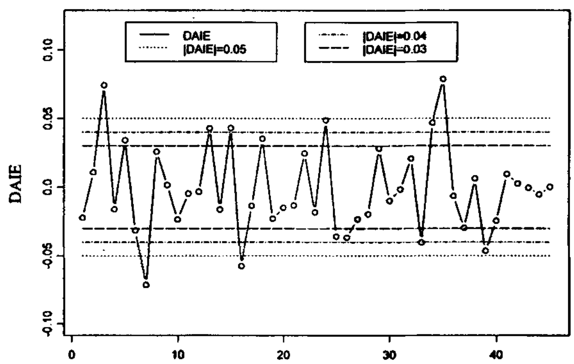

Fig. 6. Differences of absolute interpolation errors from (5.2) and (5.3).

Table 4. Comparison between MMSELI and MMAELI based on DAIE among 45 observations.

MMAELI better	Number of points	MMSELI better	Number of points
$DAIE \geq 0.03$	8	$DAIE \leq -0.03$	7
$DAIE \geq 0.04$	6	$DAIE \leq -0.04$	4
$DAIE \geq 0.05$	2	$DAIE \leq -0.05$	2
$DAIE \geq 0.06$	2	$DAIE \leq -0.06$	1
$DAIE \geq 0.07$	2	$DAIE \leq -0.07$	1
$DAIE \geq 0.072$	2	$DAIE \leq -0.072$	0

while the MMSELI is

$$\hat{Z}_\tau^{L2} = 1.79 + 0.3628998 y_{\tau-1} + 0.3628998 y_{\tau+1}. \tag{5.3}$$

We next consider the MMAELI and the MMSELI for Z_τ's, $\tau = 1, \ldots, 45$, and compare their interpolation residuals. It is assumed that $Z_0 = Z_{46} = 1.79$. Let $DAIE_\tau = |Z_\tau - \hat{Z}_\tau^{L2}| - |Z_\tau - \hat{Z}_\tau^{L1}|$. We plot the $DAIE_\tau$, $\tau = 1, \ldots, 45$, in Fig. 6. Based on $DAIE$, MMAELI beats MMSELI if $DAIE > 0$, and MMSELI is preferred if $DAIE < 0$. Table 4 gives the comparison between MMSELI and MMAELI for $|DAIE| \geq 0.03$. It is noted that the number of observations for which MMAELI is better is uniformly more than that for which MMSELI is better. This indicates that MMAELI outperforms MMSELI. Since a_t is not Gaussian, the process $\{Z_t\}$ in (5.1) is not time reversible from Corollary 4.3 of Tong ((1990), p. 196). This is why MMAELI is preferred to MMSELI, which completely ignores the time irreversibility of the process.

6. Extension to AR(p) model

We now extend the results for the AR(1) to the AR(p) model

$$y_t = \phi(B)y_t + \varepsilon_t = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-p} + \varepsilon_t, \tag{6.1}$$

where $1 - \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is a p-th order polynomial with all roots outside the unit circle, $\{\varepsilon_t\}$ is an i.i.d. innovation process with finite first absolute moment, and ε_t is independent of $\{y_s, s < t\}$. Assume that the observed samples are $y_1, \ldots, y_{\tau-1}, y_{\tau+1}, \ldots, y_n$ with a missing value at $t = \tau$, and the series mean $\mu = 0$. For simplicity, suppose that $p < \tau < n - p$. Set $Y_{\tau+p} = (y_{\tau+p}, \ldots, y_{\tau+1})'$, $Y_{\tau-1} = (y_{\tau-1}, \ldots, y_{\tau-p})'$, and a $p \times p$ matrix

$$\tilde{\phi} = \begin{pmatrix} \phi_{(p-1)} & \phi_p \\ I_{p-1} & \mathbf{0} \end{pmatrix}, \tag{6.2}$$

where $\phi_{(p-1)} = (\phi_1, \ldots, \phi_{p-1})$, $\mathbf{0} = (0, \ldots, 0)^\tau \in R^{p-1}$, and I_{p-1} is the identity matrix.

PROPOSITION 6.1. *If the i.i.d. innovation process ε_t has a non-degenerate distribution $F_\varepsilon(\cdot)$ and a density function $p_\varepsilon(\cdot)$ with a first absolute moment and zero mean, then*

$$\hat{y}_\tau^{L1} = D_0'[Y_{\tau+p} - \tilde{\phi}^{p+1} Y_{\tau-1}] + \sum_{i=1}^{p} (-\rho_i)[y_{\tau+i} + y_{\tau-i}]. \tag{6.3a}$$

Here $\tilde{\phi}$ is the matrix defined in (6.2) *and*

$$\rho_i = \frac{-\phi_i + \sum_{j=1}^{p-i} \phi_i \phi_{j+i}}{1 + \sum_{j=1}^{p} \phi_i^2}, \qquad i = 1, \ldots, p, \tag{6.3b}$$

are the inverse autocorrelations. $D_0 = (d_1, \ldots, d_p)'$ is the solution to the equations

$$\sum_{j=0}^{p} \phi_{i1}^{(p-j)} E[I_{\{a_0\varepsilon_\tau - \Sigma_{k=1}^{p} a_k \varepsilon_{\tau+k} > 0\}} \varepsilon_{\tau+j}] = 0, \ i = 1, \ldots, p \tag{6.3c}$$

with a_k's given by

$$a_0 = a_0(D_0) = 1 - \left(1 + \sum_{i=1}^{p} \phi_i^2\right) \left(\sum_{i=1}^{p} \phi_{i1}^{(p)} d_i\right), \tag{6.4a}$$

$$a_j = a_j(D_0) = \phi_j + \left(1 + \sum_{i=1}^{p} \phi_i^2\right) \left(\sum_{i=1}^{p} \phi_{i1}^{(p-j)} d_i\right), \ j = 1, \ldots, p, \tag{6.4b}$$

and $\phi_{i1}^{(\ell)}$*'s are calculated recursively from*

$$\text{(6.4c)}\quad \phi_{11}^{(\ell+1)} = \sum_{i=1}^{p} \phi_i \phi_{i1}^{(\ell)}, \quad \phi_{i1}^{(\ell+1)} = \phi_{i-1,1}^{(\ell)}, \quad i = 2, \ldots, p \text{ and } \ell = 0, 1, \ldots, p-1,$$

with $\phi_{11}^{(0)} = 1$ *and* $\phi_{i1}^{(0)} = 0$ *for* $i = 2, 3, \ldots, p$.

The proof of this proposition is given in Appendix B.

Remark 6.1. If $1 < \tau \leq p$ or $n - p \leq \tau < n$, then $\hat{y}_\tau^{L1}$ may be calculated by letting the unobserved y_i's in (6.3a) equal the series mean $\mu = 0$.

Remark 6.2. The series $\{y_t\}$ is allowed to be non-stationary in Proposition 6.1. That is, the root of $1 - \phi(B) = 0$ may be on the unit circle as long as the density of the initial series $y_{in} = (y_0, y_{-1}, \ldots, y_{-p+1})$ exists.

7. Discussions

More results on the MMAELI for different time series models are discussed in this section.

7.1 *Multiple missing values*

We extend Brubacker and Wilson (1976) and Beveridge (1992)'s idea of interpolating multiple missing values based on the MMSELI to the same setting using the MMAELI. The basic idea is to apply MMSELI to each missing value and replace the missing values on the RHS of (1.1) by their corresponding interpolators. The unobserved out-of-samples data is estimated by the series mean. We illustrate the idea using the following example.

Consider the MMAELI for the AR(1) process. Suppose the observed samples are y_1, y_4, y_5, y_7 and y_8. Here y_2, y_3 and y_6 are the missing values. Following Beveridge (1992), we have

$$\begin{aligned}
\hat{y}_2^{L1} &= d_0[\hat{y}_3^{L1} - \phi^2 y_1] + \frac{\phi}{1+\phi^2}[\hat{y}_3^{L1} + y_1] = \frac{\phi a_0 y_1 + b_0 \hat{y}_3^{L1}}{1+\phi^2}, \\
\hat{y}_3^{L1} &= d_0[y_4 - \phi^2 \hat{y}_2^{L1}] + \frac{\phi}{1+\phi^2}[\hat{y}_2^{L1} + y_4] = \frac{\phi a_0 \hat{y}_2^{L1} + b_0 y_4}{1+\phi^2}, \\
\hat{y}_6^{L1} &= d_0[y_7 - \phi^2 y_5] + \frac{\phi}{1+\phi^2}[y_7 + y_5] = \frac{\phi a_0 y_5 + b_0 y_7}{1+\phi^2}.
\end{aligned}$$

Thus solving these equations leads to

$$\hat{y}_2^{L1} = \frac{\phi a_0(1+\phi^2)y_1 + b_0^2 y_4}{(1+\phi^2)^2 - \phi a_0 b_0}, \quad \hat{y}_3^{L1} = \frac{b_0(1+\phi^2)y_1 + (\phi a_0)^2 y_4}{(1+\phi^2)^2 - \phi a_0 b_0}, \quad \hat{y}_6^{L1} = \frac{\phi a_0 y_5 + b_0 y_7}{1+\phi^2},$$

where a_0 and b_0 are defined in Proposition 3.1.

7.2 *ARMA model*

For the invertible ARMA(p,q) model

$$(1 - \phi(B))y_t = (1 - \theta(B))\varepsilon_t, \tag{7.1}$$

with all the roots of $1 - \theta(B) = 0$ outside the unit circle, it can be expressed as an AR(∞) model

$$y_t = \psi(B)y_t + \varepsilon_t = \sum_{j=1}^{\infty} \psi_j y_{t-j} + \varepsilon_t, \tag{7.2}$$

where $1 - \psi(B) = \frac{1-\phi(B)}{1-\theta(B)} = 1 - \sum_{j=1}^{\infty} \psi_j B^j$. Intuitively, the MMAELI for (7.2) should have a similar form obtained from letting $p \to \infty$ in Proposition 6.1.

$$\hat{y}_T^{L1} = \sum_{j=1}^{\infty}\{c_{-j}y_{T-j} + c_j y_{T+j}\} + \sum_{j=1}^{\infty} \rho_j\{y_{T-j} + y_{T+j}\}, \tag{7.3}$$

where the coefficients, c_j's, are determined by the infinite-dimension vector D_0, and the second summation equals the MMSELI in (1.1) with $\mu = 0$. The interpolation technique for AR process can be applied using an AR approximation to the ARMA model.

7.3 *Comparison of MMAELI with MMSELI*

Proposition 2.1 shows that $MAE(\hat{y}_T^{L1})$ is smaller than $MAE(\hat{y}_T^{L2})$, while $SMSE(\hat{y}_T^{L2}) \leq SMSE(\hat{y}_T^{L1})$. This implies that MMAELI, $\hat{y}_T^{L1}$, is better than MMSELI, $\hat{y}_T^{L2}$, in terms of MAE, but it is not the case in terms of SMSE.

(1) For the non-normal case, MMSELI is better than MMAELI in terms of SMSE. However, Table 1 shows that $(SMSE(\hat{y}_T^{L1}) - SMSE(\hat{y}_T^{L2}))/SMSE(\hat{y}_T^{L2})$ is very small compared to the large value of $(MAE(\hat{y}_T^{L2}) - MAE(\hat{y}_T^{L1}))/MAE(\hat{y}_T^{L1})$. Therefore, for an innovation with contaminated normal, MMAELI is a good approximation to MMSELI in terms of SMSE. MMAELI outperforms MMSELI in terms of MAE especially for serious contaminations.

(2) Subsection 4.2 shows that MMAELI exists, but MMSELI does not, for t_2 innovations. This illustrates that MMAELI is more applicable than MMSELI.

(3) (1.1) shows that the weightings of the observations after the missing value and before the missing value are symmetric. However, their contributions to the MMAELI are asymmetric in general, which capture the feature of asymmetry between y_{T+1} and y_{T-1} in the conditional density function of y_T given y_{ob} in (3.4).

(4) Note that MMAELI does not treat the missing observations as nuisance parameters to be estimated directly. MMAELI has good properties of (3) and (4) of the four criteria for the most useful technique suggested by Beveridge (1992).

Acknowledgements

We would like to express our deep gratitude to both referees and the editor for their invaluable comments and suggestions that greatly improved the presentation of this paper.

Appendix

A. Proof of Theorem 4.4.

Note that

$$h_k'(u) = E[\varepsilon_t^2 f_k(u\varepsilon_t)] = kW_k \left[g_k\left(\frac{k+1}{2}, \frac{k-1}{2}\right) - g_k\left(\frac{k+1}{2}, \frac{k+1}{2}\right)\right], \tag{A.1}$$

where $W_k = 2[\frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}]^2$ and $g_k(i,j) = \int_0^\infty (1+\frac{u^2x^2}{k})^{-i}(1+\frac{x^2}{k})^{-j}dx$. When $u^2 \neq 1$, $g_k(i,j)$ can be calculated recursively from

$$g_k(i,j) = \frac{1}{u^2-1}[u^2 g_k(i,j-1) - g_k(i-1,j)], \qquad \text{for} \quad i,j \geq 1. \tag{A.2}$$

When $u^2 = 1$, $g_k(i,j)$ reduces to $g_k(0, i+j)$. $g_k(i,0)$ and $g_k(0,j)$ can be calculated from

$$g_k(i,0) = \frac{\sqrt{k\pi}\Gamma\left(\frac{2i-1}{2}\right)}{2\Gamma(i)|u|} \quad \text{and} \quad g_k(0,j) = \frac{\sqrt{k\pi}\Gamma\left(\frac{2j-1}{2}\right)}{2\Gamma(j)}, \quad \text{for } i,j \geq 1. \tag{A.3}$$

Thus we have

$$\begin{aligned}
g_3(1,1) &= \frac{1}{u^2-1}(u^2 g_3(1,0) - g_3(0,1)) = \frac{\sqrt{3}\pi}{2}\frac{1}{|u|+1},\\
g_3(2,1) &= \frac{1}{u^2-1}(u^2 g_3(2,0) - g_3(1,1)) = \frac{\sqrt{3}\pi}{4}\frac{|u|+2}{(|u|+1)^2},\\
g_3(2,2) &= \frac{1}{u^2-1}\left[u^2 g_3(2,1) - \frac{1}{u^2-1}(u^2 g_3(1,1) - g_3(0,2))\right] = \frac{\sqrt{3}\pi}{4}\frac{u^2+3|u|+1}{(|u|+1)^3},
\end{aligned}$$

and

$$h_3'(u) = 3W_3[g_3(2,1) - g_3(2,2)] = 3W_3\frac{1}{(|u|+1)^3}.$$

It follows from (4.2b) that

$$h_3(\alpha) = \int_0^\alpha h_3'(u)du = 3W_3\frac{(2+|\alpha|)\alpha}{(|\alpha|+1)^2},$$

$$h_3\left(\frac{a_0}{b_0}\right) = a_0 \operatorname{sgn}(b_0) H_3(a_0,b_0) \quad \text{and} \quad h_3\left(\frac{b_0}{a_0}\right) = b_0 \operatorname{sgn}(a_0) H_3(b_0,a_0), \tag{A.4}$$

where $H_3(a,b) = \frac{|a|+2|b|}{(|a|+|b|)^2}$.

Theorem 4.4 follows from Proposition 3.1, (3.14) and (A.4).

B. Proof of proposition 6.1

For model (6.1), let $y_{in} = (y_0, y_{-1}, \ldots, y_{-p+1})$. The joint density function of (y_T, y_{ob}, y_{in}) is

$$p(y_T, y_{ob}, y_{in}) = p_{in}(y_{in}) \prod_{j=1}^n p_\varepsilon(\tilde{y}_j), \tag{B.1}$$

where $p_{in}(\cdot)$ is the density function of y_{in} and $\tilde{y}_j = y_j - \sum_{i=1}^{p} \phi_i y_{j-i}$ for $j = 1, \ldots, n$. The conditional density function of y_τ ($\tau > p$) given y_{ob} is

$$\text{(B.2)} \quad p(y_\tau \mid y_{ob}) = \frac{p(y_\tau, y_{ob})}{p(y_{ob})} = \frac{p_\varepsilon\left(y_\tau - \sum_{i=1}^{p} \phi_i y_{\tau-i}\right) \cdots p_\varepsilon\left(y_{\tau+p} - \sum_{i=1}^{p} \phi_i y_{\tau+p-i}\right)}{\int p_\varepsilon\left(y_\tau - \sum_{i=1}^{p} \phi_i y_{\tau-i}\right) \cdots p_\varepsilon\left(y_{\tau+p} - \sum_{i=1}^{p} \phi_i y_{\tau+p-i}\right) dy_\tau}.$$

Hence, conditional on y_{ob}, y_τ depends only on $y_{\tau,p} = (y_{\tau+p}, \ldots, y_{\tau+1}, y_{\tau-1}, \ldots, y_{\tau-p})$. Thus, $p(y_\tau \mid y_{ob}) = p(y_\tau \mid y_{\tau,p})$. The linear interpolator of y_τ defined in (2.1) now reduces to

$$\text{(B.3)} \qquad \hat{y}_\tau^{L1} = \sum_{i=1}^{p} c_i y_{\tau+i} + \sum_{i=1}^{p} c_{p+i} y_{\tau-i}$$

with c_i's being $2p$ real constants.

Define

$$\text{(B.4a)} \qquad V_\tau = y_\tau - \hat{y}_\tau^{L2}.$$

For AR(p) model, let ρ_i be the inverse autocorrelation. From Beveridge (1992),

$$\text{(B.4b)} \qquad \hat{y}_\tau^{L2} = -\sum_{i=1}^{p} \rho_i [y_{\tau+i} + y_{\tau-i}].$$

It follows from (B.4) that

$$\text{(B.5)} \qquad V_\tau = \left(\varepsilon_\tau - \sum_{i=1}^{p} \phi_i \varepsilon_{\tau+i}\right) \Big/ \left(1 + \sum_{i=1}^{p} \phi_i^2\right),$$

which is similar to v_τ in (3.7). Hence (B.4) is desired. We express the AR(p) model as

$$\text{(B.6)} \qquad Y_t = \tilde{\phi} Y_{t-1} + \mathcal{E}_t,$$

where $Y_t = (y_t, \ldots, y_{t-p+1})'$, $\mathcal{E}_t = (\varepsilon_t, 0, \ldots, 0)'$ are p-dimensional random vectors, and $\tilde{\phi}$ is defined in (6.2). From (B.3) and (B.6), consider the transformation

$$\text{(B.7)} \qquad U_\tau = Y_{\tau+p} - \tilde{\phi}^{p+1} Y_{\tau-1} = \mathcal{E}_{\tau+p} + \sum_{j=1}^{p} \tilde{\phi}^j \mathcal{E}_{\tau+p-j}.$$

Since (V_τ, U_τ) is independent of $Y_{\tau-1}$ from (B.5) and (B.7), we have

$$\text{(B.8)} \qquad p(V_\tau \mid y_{\tau,p}) = p(V_\tau \mid U_\tau, Y_{\tau-1}) = p(V_\tau \mid U_\tau).$$

Hence the MMAELI, $\hat{V}_\tau^{L1}$, of V_τ based on $y_{\tau,r}$ is of the form

$$\text{(B.9)} \qquad \hat{V}_\tau^{L1} = D_0' U_\tau,$$

where D_0 is a p-dimensional constant vector which minimizes

$$\widetilde{MAE}(D) = E[|V_\tau - D'U_\tau|]. \tag{B.10}$$

It follows from Pinkus ((1989), p. 14) that D_0 minimizing (B.10) is equivalent to

$$|E[\text{sgn}(V_\tau - D_0'U_\tau)U_{\tau,i}]| \le E[I_{\{V_\tau - D_0'U_\tau = 0\}}|U_{\tau,i}|], \qquad i = 1, \ldots, p, \tag{B.11}$$

where $U_{\tau,i}$ is the i-th element of the random vector U_τ.

Denote $D_0 = (d_1, \ldots, d_p)'$, $\tilde{\phi}^0 = I$ (unit matrix of order p) and $\phi_{i1}^{(j)}$ the i-th element of the first column of matrix $\tilde{\phi}^j$. Then

$$D_0'U_\tau = \sum_{j=0}^{p} D_0'\tilde{\phi}^j \mathcal{E}_{\tau+p-j} = \sum_{j=0}^{p}\sum_{i=1}^{p} d_i \phi_{i1}^{(j)} \varepsilon_{\tau+p-j} = \sum_{j=0}^{p}\sum_{i=1}^{p} d_i \phi_{i1}^{(p-j)} \varepsilon_{\tau+j}. \tag{B.12}$$

(B.5) and (B.7) together with (B.12) give

$$\begin{aligned} V_\tau - D_0'U_\tau &= \left(a_0\varepsilon_\tau - \sum_{j=1}^{p} a_j \varepsilon_{\tau+j}\right) \Big/ \left(1 + \sum_{j=1}^{p}\phi_j^2\right) \\ &= \left(a_0\varepsilon_\tau - \sum_{j=1}^{p} a_j \varepsilon_{\tau+j}\right) \Big/ \lambda_0. \end{aligned} \tag{B.13}$$

Here a_j's and $\phi_{i1}^{(\ell)}$ are defined in (6.4), and (6.4c) follows from $\tilde{\phi}^{\ell+1} = \tilde{\phi}\tilde{\phi}^\ell$ and (6.2). By (6.4),

$$a_0 = a_0(D_0) = 1 - \lambda_0 D_0'\tilde{\phi}^p\kappa, \qquad a_j = a_j(D_0) = \phi_j + \lambda_0 D_0'\tilde{\phi}^{p-j}\kappa, \qquad j = 1, \ldots, p,$$

where $\lambda_0 = 1 + \sum_{i=1}^{p}\phi_i^2$ and $\kappa = (1, 0, \ldots, 0)' \in R^p$.

Note that a_i's in (6.4) are not all equal to 0. Since the distribution of ε_τ is non-degenerate and ε_t's are independent, it follows from (B.13) that $P(V_\tau - D_0'U_\tau = 0) = 0$. Thus from (B.11), $E[\text{sgn}(V_\tau - D_0'U_\tau)U_\tau] = 0$, and

$$E\left[\text{sgn}\left(a_0\varepsilon_\tau - \sum_{j=1}^{p} a_j\varepsilon_{\tau+j}\right)\sum_{j=0}^{p}\phi_{i1}^{(p-j)}\varepsilon_{\tau+j}\right] = 0, \qquad i = 1, \ldots, p. \tag{B.14}$$

Proposition 6.1 follows from (B.4), (B.9) and (B.14). Here, (B.14) is equivalent to (6.3c).

References

Abraham, B. (1981). Missing observations in time series, *Comm. Statist. A. Theory Methods*, **10**, 1643–1653.

Beveridge, S. (1992). Least squares estimation of missing values in time series, *Comm. Statist. Theory Methods*, **21**, 3479–3496.

Brubacker, S. R. and Wilson, G. T. (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand, *J. Roy. Statist. Soc. Ser. C*, **25**, 107–116.

Dagum, E. B., Cholette, P. A. and Chen, Z. G. (1998). A unified view of signal extraction, benchmarking, interpolation and extrapolation of time series, *International Statistical Review*, **66**, 245–269.

Damsleth, E. (1980). Interpolating missing values in a time series, *Scand. J. Statist.*, **7**, 33–39.

Dunsmuir, W. and Robinson, P. M. (1981). Estimation of time series models in the presence of missing data, *J. Amer. Statist. Assoc.*, **76**, 560–568.

Ferreiro, O. (1987). Methodologies for the estimation of missing observations in time series, *Statist. Probab. Lett.*, **5**, 65–69.

Gómez, V. and Maravall, D. (1994). Estimation, prediction, and interpolation for nonstationary series with Kalman filter, *J. Amer. Statist. Assoc.*, **89**, 611–624.

Gómez, V., Maravall, A. and Pena, D. (1999). Missing observations in ARIMA models: Skipping approach versus additive outlier approach, *J. Econometrics*, **88**, 341–363.

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

Harvey, A. C. and Pierse, R. G. (1984). Estimating missing observations in economic time series, *J. Amer. Statist. Assoc.*, **79**, 125–131.

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

Kohn, R. and Ansley, C. F. (1986). Estimation, prediction, and interpolation for ARIMA models with missing data, *J. Amer. Statist. Assoc.*, **81**, 751–761.

Ljung, G. M. (1982). The likelihood function for a stationary Gaussian autoregressive-moving average process with missing observations, *Biometrika*, **69**, 265–268.

Ljung, G. M. (1989). A note on the estimation of missing values in time series, *Comm. Statist. Simulation comput.*, **18**, 459–465.

Luceño, A. (1997). Estimation of missing values in possibly partially nonstationary vector time series, *Biometrika*, **84**, 495–499.

Parzen, E. (1984). *Time Series Analysis of Irregularly Observed Data*, Springer, New York.

Peña, D. and Tiao, G.C. (1991). A note on likelihood estimation of missing values in time series, *Amer. Statist.*, **45**, 212–213.

Penzer, J. and Shea, B. (1997). The exact likelihood of an autoregressive-moving average model with incomplete data, *Biometrika*, **84**, 919–928.

Pinkus, A. M. (1989). *On L^1-approximation*, Cambridge University Press, London.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

Venables, W. N. and Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus*, Springer, New York.

Wei, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Redwood City, California.

Whittle, P. (1963). *Prediction and Regulation by Linear Least Square Methods*, English University Press, London.

Wincek, M. A. and Reinsel, G. C. (1986). An exact maximum likelihood estimation procedure for regression-ARMA time series models with possibly nonconsecutive data, *J. Roy. Statist. Soc. Ser. B*, **48**, 303–313.