Maximum Likelihood Estimation of Sum-Difference Time Series Models Using the EM algorithm

William S. Cleveland and Chuanhai Liu

Bell Laboratories, Lucent Technologies

Murray Hill, NJ 07974

E-mail: wsc, liu@research.bell-labs.com

Abstract

Integrated moving average (IMA) processes, especially the first-order moving average process IMA$(1, 1)$, are useful for modeling time series data from economics and industrial control. The sequence of the first-order differences of an observed IMA(1,1) process can be expressed as the sum of two independent processes: the first-order differences of a white noise process and the first-order sums of another white noise process. The corresponding spectrum decomposition provides a simple and useful tool for model building. Moreover, the decomposition leads to a simple implementation of the EM algorithm for maximum likelihood estimation for an IMA$(1, 1)$ Gaussian process. In the frequency domain, the decomposition leads to a new class of IMA processes, called integrated sum-difference (ISD) processes. Maximum likelihood estimation for ISD processes is straightforward via the EM algorithm and its extensions. An example is provided.

Key Words: Integrated moving average processes; The Levinson-Durbin algorithm; Power spectra; Variance components.

1 Introduction

Integrated Gaussian moving average (IMA) processes (Box and Jenkins, 1970, p. 103-114) in general and IMA$(1,1)$ processes in particular are popular tools for modeling time series, especially, in economics and industrial control (*i.e.*, Vander Wiel (1996) and Clark et al. (1999)). Maximum likelihood (ML) estimation of Gaussian IMA processes, which is often difficult, is discussed by Box and Jenkins (1970), Kohn and Ansley (1986), and Anderson and Mentz (1993). We show in this paper that the ML estimate of the parameters of a IMA$(1,1)$ process can be obtained easily using the EM algorithm (Dempster, Laird, and Rubin, 1977).

Box and Jenkins (1970, p. 121-124) noticed that an IMA(1,1) process with added white noise is still an IMA(1,1) process. Although it is questionable to claim that any IMA(1,1) process can be written as a random walk buried in white noise (Box and Jenkins, 1970, p.123), any IMA(1,1) process can be viewed as a *smoothest possible* IMA(1,1) process buried in white noise. This view motivates our orthogonal decomposition of IMA(1,1) Gaussian processes. More precisely, we can write an IMA(1,1) Gaussian process $\{y_t\}$ as $y_t = x_t + e_t$ and $x_t - x_{t-1} = a_t + a_{t-1}$, where $\{a_t\}$ and $\{e_t\}$ are independent white noise with $\sigma_a^2 = \mathrm{var}(a_t)$ and $\sigma_e^2 = \mathrm{var}(e_t)$ and $\{x_t\}$ is the *smoothest possible* IMA(1,1) process. The observed IMA(1,1) time series is then a simple linear combination (with known coefficients) of two independent white noise processes, which leads to a simple implementation of the EM algorithm for finding the maximum (conditional) likelihood estimates of the parameters.

Using the backward operator $\mathbf{B}$ (*e.g.*, Box and Jenkins, 1970), we can re-write the IMA(1,1) process as $(1-\mathbf{B})y_t = (1+\mathbf{B})a_t + (1-\mathbf{B})e_t$. The spectral density of $(1-\mathbf{B})y_t$ can be decomposed as $8\sigma_a^2\cos^2(\pi f) + 8\sigma_e^2\sin^2(\pi f) \quad (0 \le f \le 1/2)$. The last decomposition motivates a more general class of IMA processes: $(1-\mathbf{B})^d y_t = \sum_k \phi_k(\mathbf{B})\varepsilon_t^{(k)}$, where $\{\varepsilon_t^{(k)}\}$ are white noise processes and $\phi_k(\mathbf{B})$ are polynomial functions of $\mathbf{B}$ with *known* coefficients. The known functions $\phi_k(\mathbf{B})$ typically represent prior knowledge. To be specific, we discuss, but will not be limited to, the class of IMA processes that can be represented as $(1-\mathbf{B})^d y_t = \sum_{k=0}^q (1-\mathbf{B})^k(1+\mathbf{B})^{q-k}\varepsilon_t^{(k)}$, where $\{\{\varepsilon_t^{(k)}\} : k = 0, ...q\}$ are independent white noise with $\varepsilon_t \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_k^2)$. We call this process an *integrated sum-difference process*, ISD(d,q). The frequency distribution of the ISD$(0,q)$ component can be written as $2^{2q+1}\sum_{k=0}^q \sigma_k^2 \sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f) \quad (0 \le f \le 1/2)$. The ECME algorithm (Liu and Rubin, 1994), or more generally, the AECM algorithm (Meng and van Dyk, 1997), can be implemented for finding the maximum (conditional) likelihood estimates of the parameters, which are the variance components $\{\sigma_k^2 : k = 0, ..., q\}$ of ISD(d, q).

In Section 2, we describe the orthogonal decomposition in more detail and the EM algorithm for ML estimation for a Gaussian ISD(1,1) process, which is an IMA(1,1) process, with a linear trend. Section 3 extends the orthogonal decomposition to the more general ISD(d, q) processes. Section 4 shows how EM-type algorithms can be used to obtain maximum likelihood estimates of the parameters of ISD(d, q) processes. Section 5 illustrates the use of ISD(d, q) processes. Section 6 has a few remarks showing the advantages of ISD processes and their multivariate versions for model building and multiple imputation.

2 The Gaussian IMA(1,1) processes

2.1 An orthogonal decomposition of the Gaussian IMA(1,1) processes

The IMA(1,1) process with a linear trend for a time series is defined by $y_t - y_{t-1} = \delta + \varepsilon_t + \theta\varepsilon_{t-1}$, where the scalar δ is the slope of the linear trend, $-1 < \theta < 1$, and $\varepsilon_t \overset{\text{ind}}{\sim} \text{N}(0, \sigma_\varepsilon^2)$ (Box and Jenkins, 1970). As Clark et al. (1999) showed, any IMA$(1,1)$ process can be written as

$$x_t - x_{t-1} = \delta + a_t + a_{t-1}, \quad \text{and} \quad y_t = x_t + e_t, \tag{1}$$

where the random shock $a_t \overset{\text{ind}}{\sim} \text{N}(0, \sigma_a^2)$, the white noise $e_t \overset{\text{ind}}{\sim} \text{N}(0, \sigma_e^2)$, and $\{a_t\}$ and $\{e_t\}$ are independent.

Equation (1) decomposes an IMA(1,1) time series $\{y_t\}$ into two components: the smoothest possible IMA(1,1) component $\{x_t\}$ (the one with $\theta = 1$) and a white noise component $\{e_t\}$. Comparing the variances and first-order auto-covariances of the first-order difference of $\{y_t\}$ shows that $(1 + \theta^2)\sigma_\varepsilon^2 = 2(\sigma_a^2 + \sigma_e^2)$ and $\theta\sigma_\varepsilon^2 = \sigma_a^2 - \sigma_e^2$, which implies that $\theta = (\sigma_a - \sigma_e)/(\sigma_a + \sigma_e)$ and $\sigma_\varepsilon = \sigma_a + \sigma_e$. Taking $\sigma_e^2 = \sigma_a^2$ gives $\theta = 0$. Thus, a *random walk* corresponds to decomposition (1) with $\sigma_e^2 = \sigma_a^2$, and an IMA(1,1) that is noisier than a *random walk* can be thought of as a *random walk* buried in *white noise.*

2.2 Maximum likelihood estimation

Anderson and Mentz (1993) have used Newton-Raphson algorithm for maximum likelihood estimation of a Gaussian IMA$(1, 1)$ process. In general, however, the EM algorithm is simpler to implement than Newton-Raphson and it converges monotonically in the sense that each iteration of EM increases the likelihood. Here we show how the EM algorithm can be used to find the ML estimate of an IMA$(1,1)$ Gaussian process.

Using the decomposition (1), the log-likelihood function (more exactly, the conditional log-likelihood given y_1) obtained from the first-order differences of $\{y_t\}$ can be written as

$$L(\delta, \sigma_a^2, \sigma_e^2 | Y_{\text{obs}}) = -\frac{1}{2} z W^{-1} z' - \frac{1}{2} \ln |W| + \text{constant}, \tag{2}$$

where $z = (y_2 - y_1 - \delta, ..., y_n - y_{n-1} - \delta)'$ and W is the $(n-1) \times (n-1)$ tridiagonal matrix with $2(\sigma_a^2 + \sigma_e^2)$ and $\sigma_a^2 - \sigma_e^2$ as its diagonal and off-diagonal elements, respectively.

Directly maximizing the log-likelihood function (2) is difficult because it is nonlinear in the parameters and W is high dimensional. Instead, we treat x_t in the decomposition (1) as missing, define the complete data as $Y_{\text{com}} = \{x_t, y_t : t = 1, ..., n\}$, and apply the EM algorithm. The complete-data log-likelihood is then $L(\delta, \sigma_a^2, \sigma_e^2 | Y_{\text{com}}) = -\frac{n-1}{2} \ln \sigma_a^2 - \frac{1}{2\sigma_a^2}(S_{\Delta x^2} - 2 S_{\Delta x \mathbf{1}} \delta + \mathbf{1}' G^{-1} \mathbf{1} \delta^2) - \frac{n}{2} \ln \sigma_e^2 - \frac{1}{2\sigma_e^2} S_{(y-x)^2} +$ constant, where

$$\begin{aligned}
S_{\Delta x \mathbf{1}} &= (x_2 - x_1, ..., x_n - x_{n-1}) G^{-1} \mathbf{1}, && (3)\\
S_{\Delta x^2} &= \text{trace}\left[G^{-1}(x_2 - x_1, ..., x_n - x_{n-1})'(x_2 - x_1, ..., x_n - x_{n-1})\right], && (4)\\
S_{(y-x)^2} &= \sum_{t=1}^{n} (y_t - x_t)^2, && (5)
\end{aligned}$$

and G is the $(n-1) \times (n-1)$ tridiagonal matrix with 2 and 1 as its diagonal and off-diagonal elements, respectively. The (i,j)-th element of G^{-1} is then $(-1)^{(i+j)}(n-i)j/n$ for $i = 1, ..., n-1$ and $j = 1, ..., i$. Given the complete-data sufficient statistics (3)—(5), the complete-data ML estimates of δ, σ_a^2, and σ_e^2 are simply

$$\widehat{\delta} = \frac{S_{\Delta x \mathbf{1}}}{\mathbf{1}' G^{-1} \mathbf{1}}, \quad \widehat{\sigma}_a^2 = \frac{1}{n-1}\left(S_{\Delta x^2} - \frac{S_{\Delta x \mathbf{1}}^2}{\mathbf{1}' G^{-1} \mathbf{1}}\right), \quad \text{and} \quad \widehat{\sigma}_e^2 = \frac{S_{(y-x)^2}}{n}. \tag{6}$$

Thus, the E-step and M-step at each iteration of the EM algorithm are as follows.

E-step: Compute the expected complete-data sufficient statistics $S_{\Delta x \mathbf{1}}$, $S_{\Delta x^2}$, and $S_{(y-x)^2}$ in Equations (3)—(5) given Y_{obs} and the current estimates of the parameters. The conditional distribution of $(x_1, ..., x_n)$ given $(Y_{\text{obs}}, \delta, \sigma_a^2, \sigma_e^2)$ is $\mathrm{N}((\widehat{x}_1, ..., \widehat{x}_n)', B^{-1})$, where

$$(\widehat{x}_1, ..., \widehat{x}_n)' = B^{-1}[\sigma_e^{-2}(y_1, ..., y_n)' + \sigma_a^{-2} \delta D(0, 1, ..., n-1)'], \tag{7}$$

$B = \sigma_e^{-2}\boldsymbol{I} + \sigma_a^{-2}D$, the i-th diagonal element of the $n \times n$ matrix D is $[(n-1)+4(i-1)(n-i)]/n$ for $i = 1, ..., n$, and the (i,j)-th element of D is $(-1)^{(i+j)}[-2n-1+2(i+j)+4j(n-i)]/n$ for $i = 2, ..., n$ and $j = 1, ..., i-1$.

M-step: Update the estimates of the parameters using the set of equations (6), replacing the sufficient statistics with the corresponding expected complete-data sufficient statistics, which were computed in the E-step.

Clearly, the E and M steps are simple to compute. Moreover, the EM algorithm provides estimates of the smoothest IMA(1,1) component $\{x_t\}$, the random shocks $\{a_t\}$, and the white noise $\{e_t\}$. The estimate $\{\widehat{x}_t : t = 1, ..., n\}$ is given by Equation (7). The estimate $\{\widehat{a}_t : t = 1, ..., n\}$ can be obtained by noting that $(a_1, ..., a_n)'|(Y_{\text{obs}}, \delta, \sigma_a^2, \sigma_e^2) \sim \text{N}\left((\widehat{a}_1, ..., \widehat{a}_n)', V^{-1}\right)$, where $V = \frac{1}{\sigma_a^2}\boldsymbol{I} + \frac{1}{\sigma_e^2}KU^{-1}K'$. Then $(\widehat{a}_1, ..., \widehat{a}_n)' = \sigma_e^{-2}V^{-1}KU^{-1}(y_2 - y_1 - \delta, \ldots, y_n - y_{n-1} - \delta)'$, where U is the $(n-1) \times (n-1)$ tridiagonal matrix with 2 and -1 as its diagonal and off-diagonal elements, and K is the $n \times (n-1)$ matrix with all non-zero elements $K_{j,j} = K_{j+1,j} = 1$ for $j = 1, ..., n-1$. The (i,j)-th element of U^{-1} is $(n-i)j/n$ for $i = 1, ..., n-1$ and $j = 1, ..., i$. When n is large, the inverse of these matrices can be computed efficiently using the Levinson-Durbin algorithm. Details are given in Appendix A.1.

3 The integrated sum-difference processes

3.1 Definition

As the decomposition (1) shows, an IMA(1,1) process with a linear trend can be written as $(1-\mathbf{B})y_t = \delta + (1+\mathbf{B})a_t + (1-\mathbf{B})e_t$, where $\mathbf{B}$ is the backward operator and $\{a_t\}$ and $\{e_t\}$ are independent white noise processes. More generally, the decomposition can be extended to a class of IMA(d,q) processes by writing

$$(1-\mathbf{B})^d y_t = g(t,\beta) + \sum_{k=0}^{q}(1-\mathbf{B})^k(1+\mathbf{B})^{q-k}\varepsilon_t^{(k)}, \tag{8}$$

where $g(t,\beta)$ is a deterministic function with the unknown parameter β and $\{\varepsilon_t^{(k)} : t = 0, \pm1, ...; k = 1, ..., q\}$ are mutually independent with $\varepsilon_t^{(k)} \sim \text{N}(0, \sigma_k^2)$. because the process defined by (8) is a sum of accumulated sums and differences, we call it an *integrated sum-difference process* and denote it by ISD(d,q).

3.2 The power spectra and variances of an ISD(d, q) process

The spectrum density function $p_y(f)$ of the random component of $\{y_t\}$ of an ISD(p, q) process satisfies the equation

$$2^{d+1}\sin^{2d}(\pi f)\;\; p_y(f) = 2^{2q+1}\sum_{k=0}^{q}\sigma_k^2 \sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f) \qquad \left(0 \le f \le \frac{1}{2}\right). \tag{9}$$

The time domain component $\{\varepsilon_t^{(k)}\}$ in the orthogonal decomposition (8) corresponds to the frequency component $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$ for $k = 0, ..., q$ and $q > 0$. The frequency components $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$ are displayed in Figure 1 for six values of q. For example, the frequency component $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$ is a unimodal function of f for $0 \le f \le 1/2$ with mode at $f_{q,k} = \pi^{-1}\arctan\left(k/(q-k)\right)^{1/2}$. As is common for normal distributions, the square root of the inverse of the curvature of $-\ln[\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)]$ at the mode is used to scale the frequency distribution $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$. That is, the scale of the frequency distribution $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$ is $\pi^{-1}(2q)^{-1/2}\cos(\pi f_{q,k})$ for $k = 0$, $(2\pi)^{-1}k^{-1/2}\sin(\pi f_{q,k})$ for $0 < k < q$, and $\pi^{-1}(2q)^{-1/2}\sin(\pi f_{q,k})$ for $k = q$. Thus, if two pairs (q, k) have the same mode in $(0, 1/2)$ (for example, $(16, 4)$ and $(4, 1)$), then the components with larger k have smaller scales and are more concentrated around the mode.

Following Box and Jenkins (1970, p40), the portion of the variance of $(1-\mathbf{B})^d y_t$ that can be explained by the component $(1-\mathbf{B})^k(1+\mathbf{B})^{q-k}\varepsilon_t^{(k)}$, henceforth the (q, k) component, can be written as

$$\operatorname{var}\left((1-\mathbf{B})^k(1+\mathbf{B})^{q-k}\varepsilon_t^{(k)}\Big|\,\sigma_k^2\right) = \frac{2^{2q}}{\pi}\mathrm{Beta}\left(\frac{2k+1}{2}, \frac{2(q-k)+1}{2}\right)\sigma_k^2.$$

Hence $\operatorname{var}\left((1-\mathbf{B})^d y_t\right) = \pi^{-1}2^{2q}\sum_{k=0}^{q}\mathrm{Beta}\left(k+1/2, (q-k)+1/2\right)\sigma_k^2$. For example, when $q = 1$ $\operatorname{var}\left((1-\mathbf{B})^d y_t\right) = 2\sigma_0^2 + 2\sigma_1^2$, which can be easily computed using expression (8).

3.3 The ISD(d, q) processes as a class of IMA(d, q) processes

The class of ISD$(1, 1)$ processes is the same as the class of IMA$(1, 1)$ processes. Here, we explore the relationship between ISD(d, q) and IMA(d, q) processes, especially ISD$(d, 2)$ and IMA$(d, 2)$. The results are summarized into the following three propositions whose proofs are given in Appendix A.2. Proposition 1 states that the class of ISD(d, q) processes contains the class of IMA(d, q) processes that have real characteristic roots. Proposition 2 states that the class of ISD(d, q) processes can be larger than the class of IMA(d,q) processes with real characteristic roots. Proposition 3 says that ISD(d,q) processes are nested by q. This nesting is useful for model fitting.

Proposition 1. *Any Gaussian IMA(d,q) process*

$$(1-\mathbf{B})^d y_t = \phi(\mathbf{B}) a_t = (1+\phi_1\mathbf{B}+...+\phi_q\mathbf{B}^q)a_t, \tag{10}$$

where $a_t \overset{\text{ind}}{\sim} \mathrm{N}(0,\sigma_a^2)$ *for* $t=0,\pm1,...$ *has a unique decomposition of the form (8) if all the roots of the characteristic equation* $\phi(\mathbf{B})=0$ *are real.*

Proposition 2. *An ISD*$(0,2)$ *process is also an IMA*$(0,2)$ *process with real coefficients and real characteristic roots if and only if* $\sigma_1^2 - 2\sigma_0\sigma_2 \geq 0$.

Proposition 3. *Any ISD*(d,q_1) *process* y_t *is also an ISD*(d,q_2) *process for* $q_2 > q_1$.

The relationship between ISD$(0,2)$ and IMA$(0,2)$ is illustrated in Figure 2. It is obtained as follows. Let $\alpha_0 = 3\sigma_0^2/(3\sigma_0^2+\sigma_1^2+3\sigma_2^2)$, $\alpha_1 = \sigma_1^2/(3\sigma_0^2+\sigma_1^2+3\sigma_2^2)$, and $\alpha_2 = 3\sigma_2^2/(3\sigma_0^2+\sigma_1^2+3\sigma_2^2)$, so $\alpha_1 = 1-\alpha_0-\alpha_2$, $\alpha_0 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_0+\alpha_2 \geq 0$. In terms of the first two autocorrelation coefficients ρ_1 and ρ_2, we have $\rho_1 = \frac{2}{3}\alpha_0 - \frac{2}{3}\alpha_2$ and $\rho_2 = \frac{2}{3}\alpha_0 + \frac{2}{3}\alpha_2 - \frac{1}{2}$. Thus, the first two autocorrelations of an ISD$(0,2)$ process lie within the area bounded by $\rho_2-\rho_1 = -\frac{1}{2}$, $\rho_2+\rho_1 = -\frac{1}{2}$, and $\rho_2 = \frac{1}{6}$. The corresponding area for an IMA$(0,2)$ process is given in Box and Jenkins (1970, p.71).

3.4 Reduced ISD(d,q) process

The spectrum decomposition of ISD(d,q) processes provides a way to model time series data in the frequency domain. A more general class of models of the form (8) that may also be useful for modeling can be obtained by selecting a set of components $\mathcal{C} \subset \Omega = \{(q,k) : 0 \leq k \leq q \text{ and } q = 1,2,...\}$:

$$(1-\mathbf{B})^d y_t = \sum_{(q,k)\in\mathcal{C}} (1-\mathbf{B})^k (1+\mathbf{B})^{q-k} \varepsilon_t^{(q,k)}, \tag{11}$$

where $\{\varepsilon_t^{(q,k)}\}$ are independent white noise with $\varepsilon_t^{(q,k)} \sim \mathrm{N}\left(0,\sigma_{q,k}^2\right)$ for $(q,k)\in\mathcal{C}$. Any processes ISD(d, C) process in (11) is called a *regular integrated sum-difference* process. For example,

1) ISD$(d, \mathcal{C}_q)$ with $\mathcal{C}_q = \{(q,0),(q,1),...,(q,q)\}$ gives the unconstrained class of ISD(d,q) processes (8),

2) ISD$(d, \mathcal{C}_{q[k_1,k_2,...,k_m]})$ with $\mathcal{C}_{q[k_1,k_2,...,k_m]} = \{(q,k_1),(q,k_2),...,(q,k_m)\}$ specifies the restricted class of ISD(d,q) processes for which $\sigma^2_{q,k} = 0$ for all $(q,k) \in \mathcal{C}_q \setminus \mathcal{C}_{q[k_1,k_2,...,k_m]}$, and

3) ISD$(d, \mathcal{C}_{q_1,q_2[k_1,k_2,...,k_m]})$ with $\mathcal{C}_{q_1,q_2[k_1,k_2,...,k_m]} = \mathcal{C}_{q_1} \cup \mathcal{C}_{q_2[k_1,k_2,...,k_m]}$ and $q_1 < q_2$ gives the class of ISD processes that consists of all the components of ISD(d, q_1) and all the components (q_2,k_1), (q_2,k_2), ..., and (q_2,k_m) of ISD(d,q_2).

The power spectrum of $(1-\mathbf{B})^d y_t$ in (11) is $\sum_{(q,k)\in\mathcal{C}} 2^{2q+1}\sigma^2_{q,k} \sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$. It is clear that ISD$(d,\mathcal{C})$ provides a flexible class of ISD processes for modeling time series $\{y_t\}$ whose power spectra can be decomposed into the form $\sum_{(q,k)\in\mathcal{C}} 2^{2q-d+1}\sigma^2_{q,k}\sin^{2(k-d)}(\pi f)\cos^{2(q-k)}(\pi f)$, where $\mathcal{C} \subset \{(q,k) : 0 \le k \le q \text{ and } q = 1,2,...\}$. Although ISD$(d,\mathcal{C}) \subset$ ISD$(d, \max_{(q,k)\in\mathcal{C}} q)$ (Proposition 3), the number of unknown components to be estimated in ISD$(d,\mathcal{C})$ can be dramatically smaller than that in ISD$(d, \max_{(q,k)\in\mathcal{C}} q)$. For convenience, we introduce the concepts of *representable*, *closure*, and *minimal cover*. We say a component (q,k) is *representable* by $\mathcal{C}$ if there exists $a_{m,j} \ge 0$ for $(j,m) \in \mathcal{C}$ such that

$$\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f) = \sum_{(j,m)\in\mathcal{C}} a_{m,j}\sin^{2j}(\pi f)\cos^{2(m-j)}(\pi f) \tag{12}$$

for all $f \in [0,1/2]$. For example, from the proof of Proposition 3, (q,k) is representable by $\{(q+1,k),(q+1,k+1)\}$ for any $0 \le k \le q$ and $q \ge 1$. By the *closure* of $\mathcal{C}$, denoted by $\mathcal{C}^+$, we mean the set of all components that are representable by $\mathcal{C}$. We call a subset of $\mathcal{C}$ the *minimal cover* of $\mathcal{C}$, denoted by $\mathcal{C}^-$, if 1) the closure of $\mathcal{C}$ is the same as that of $\mathcal{C}^-$, that is, $\mathcal{C}^+ = (\mathcal{C}^-)^+$, and 2) the closure of $\mathcal{C}$ is not the same as the closure of any proper subset of $\mathcal{C}^-$. Minimal cover is closely related to the cover of identifiable components. For estimation, we replace ISD$(d,\mathcal{C})$ with ISD$(d,\mathcal{C}^-)$. Finding the minimum cover set $\mathcal{C}^-$ can be difficult in general and deserves further investigation. In practice, we use an *ad hoc* method that finds the least-squares solution for the coefficients $\{a_{j,m}\}$ in Equation (12) from the corresponding components evaluated over a sufficiently large set of equally spaced frequencies in the interval $[0,1/2]$. When an error term is small, the corresponding component is omitted. The process is then repeated to find the next candidate to remove from the remaining components. Finding minimal cover is less critical for model building if we use components with small value of q to explain correlated errors that are not of interest and a small set of components with large q that have scientific interpretations.

As an application, ISD$(d, \mathcal{C})$ can be used for *smoothing*. In particular, suppose we are interested in removing white noise from the original series $\{y_t\}$. More precisely, assume that y_t can be decomposed into $y_t = s_t + e_t$, where e_t is white noise and s_t follows an ISD$(d, \mathcal{C}_s)$ without white noise component. So the minimum of the power spectra of $\{s_t\}$ is zero. Then estimate $\{s_t\}$ given the observed sequence $\{y_t\}$. Because $\{y_t\}$ follows ISD$(d, \mathcal{C}_y)$, where $\mathcal{C}_y = \mathcal{C}_s \cup \{(d,d)\}$, with the component (d,d) representing the white noise component in the original scale, we see that (d,d) is not representable by $\mathcal{C}_s$. Thus, a simple way to estimate $\{s_t\}$ is to estimate $\{e_t\}$ by $\{\widehat{e}_t\}$, and then use

$$\widehat{s}_t = y_t - \widehat{e}_t \tag{13}$$

as an estimate of $\{s_t\}$. An estimate of $\{e_t\}$ is given in the next section.

4 Maximum likelihood using the EM algorithm and its extensions

Section 2 describes an EM algorithm for ML estimation of a Gaussian IMA(1,1) or ISD(1,1) process by treating the smoothest IMA(1,1) component $\{x_t\}$ as missing data. As a by-product, the algorithm also gives estimates of $\{x_t\}$, which itself might be of interest. As an alternative, the ECME algorithm (Liu and Rubin, 1994) or more generally the AECM algorithm (Meng and van Dyk, 1997) can be used as follows.

1) Maximize the constrained likelihood given the current estimates of the variance components $\sigma_a^2 = \widehat{\sigma}_a^2$ and $\sigma_e^2 = \widehat{\sigma}_e^2$ to update δ, the slope of the linear trend.

2) Use a conditional EM-step given the current estimates of δ and σ_e^2 to update the estimate of σ_a^2 with the complete data being $\{y_t\}$ and $\{a_t\}$.

3) Use a conditional EM-step given the current estimates of δ and σ_a^2 to update the estimate of σ_e^2 with the complete data being $\{y_t\}$ and $\{e_t\}$.

A nice feature of this version is that the components $\{a_t\}$ and $\{e_t\}$ are treated symmetrically. This section extends this algorithm to ML estimation of an ISD$(d, \mathcal{C})$.

Suppose that $y_1, ..., y_n$ are observed and that an ISD$(d, \mathcal{C})$ process with $g(t, \beta) = x_t'\beta$ is to be fit to $\{y_t : t = 1, ..., n\}$. The ML estimates of the parameters are found by maximizing the log-likelihood function derived from

$$Dy = X\beta + \sum_{(k,q)\in\mathcal{C}} H_{q,k}\varepsilon^{(q,k)}, \tag{14}$$

with known $y = (y_1, ..., y_n)'$, where D is the $((n-d)\times n)$ matrix whose i-th row is $(0, ..., 0, c_{d,d}, 0, ..., 0)$ with $i-1$ leading zeros, $H_{q,k}$ is the $((n-d)\times(n-d+q))$ matrix whose i-th row is $(0, ..., 0, c_{q,k}, 0, ..., 0)$ with $i-1$ leading zeros, $c_{q,k}$ is the vector of the $(q+1)$ coefficients of $(\mathbf{B}^q, ..., \mathbf{B}, 1)$ in the expansion of $(1-\mathbf{B})^k(1+\mathbf{B})^{q-k}$, $X = (x_{d+1}, ..., x_n)'$, and $\varepsilon^{(q,k)} = (\varepsilon^{(q,k)}_{1+d-q}, ..., \varepsilon^{(q,k)}_n)' \sim \mathrm{N}(0, \sigma^2_{q,k}\boldsymbol{I}_{n-d+q})$ for $(q, k) \in \mathcal{C}$. Equation (14) implies that the ML estimates can be computed by conditioning on $(y_1, ..., y_d)$.

Treating $\{\varepsilon^{(q,k)}\}$ as missing data leads to the complete data $\{y, \varepsilon^{(q,k)}\}$. The complete-data ML estimate of $\sigma^2_{q,k}$ is then $\widehat{\sigma}^2_{q,k} = \frac{1}{n-d+q}\sum_{t=1+d-q}^{n}(\varepsilon^{(q,k)}_t)^2$ for $(q, k) \in \mathcal{C}$. We use the ECME algorithm (Liu and Rubin, 1994) that updates $\widehat{\sigma}^2_{q,k}$ using the complete data $\{y, \varepsilon^{(q,k)}\}$ (rather than $\{y, \varepsilon^{(q,k)} : (q, k) \in \mathcal{C}\}$) and updates $\widehat{\beta}$ by maximizing the actual/incomplete-data log-likelihood. That is, $\widehat{\beta} = \left[X'V^{-1}X\right]^{-1} X'V^{-1}Dy$, where $V = \sum_{(q,k)\in\mathcal{C}} H_{q,k}H'_{q,k}\sigma^2_{q,k}$.

Let $\theta = \{\beta, \sigma_{q,k} : (q, k) \in \mathcal{C}\}$. The expectation of $\sum_{t=1+d-q}^{n}(\varepsilon^{(q,k)}_t)^2$ is computed from the conditional distribution of $\varepsilon^{(q,k)}|(y, \theta)$, which is

$$\mathrm{N}\left(\left[H'_{q,k}V^{-1}_{-(q,k)}H_{q,k} + \frac{1}{\sigma^2_{q,k}}\boldsymbol{I}\right]^{-1} H'_{q,k}V^{-1}_{-(q,k)}(Dy - X\beta), \left[H'_{q,k}V^{-1}_{-(q,k)}H_{q,k} + \frac{1}{\sigma^2_{q,k}}\boldsymbol{I}\right]^{-1}\right), \tag{15}$$

where $V_{-(q,k)} = \sum_{(m,j)\in\mathcal{C}} H_{m,j}H'_{m,j}\sigma^2_{m,j} - H_{q,k}H'_{q,k}\sigma^2_{q,k}$. Noticing that $H_{q,k}H'_{q,k}\sigma^2_{q,k}$ and hence V and $V_{-(q,k)}$ are Toeplitz matrices for $(q, k) \in \mathcal{C}$, the Levinson-Durbin algorithm (see Appendix A.1) is used to compute V^{-1} and $V^{-1}_{-(q,k)}$ for $(q, k) \in \mathcal{C}$. To compute mean vectors and covariance matrices in Equation (15) for $\varepsilon^{(q,k)}$, we use the equality $\left[H'_{q,k}V^{-1}_{-(q,k)}H_{q,k} + \sigma^{-2}_{q,k}\boldsymbol{I}\right]^{-1} = \sigma^2_{q,k}I - \sigma^4_{q,k}H'_{q,k}V^{-1}H_{q,k}$. for $(q, k) \in \mathcal{C}$. The log-likelihood function is evaluated using (14), that is, $-\frac{n-d}{2}\ln(2\pi) - \frac{1}{2}\ln|V| - \frac{1}{2}(Dy - X\beta)'V^{-1}(Dy - X\beta)$.. Thus, this algorithm can be summarized as follows.

CML-step: Update β by maximizing the constrained actual log-likelihood.

CMQ-steps: For each $(q, k) \in \mathcal{C}$, compute the expectation $S_{q,k} = \mathrm{E}_\theta\left(\sum_{t=1+d-q}^{n}(\varepsilon^{(q,k)}_t)^2\right)$ using Equation (15) and update $\sigma_{q,k}$: $\widehat{\sigma}^2_{q,k} = \frac{S_{q,k}}{n-d+q}$.

After the algorithm converges, we obtain the ML estimates of β and $\{\sigma^2_{q,k} : (q,k) \in \mathcal{C}\}$ and the estimates of $\{\varepsilon^{(q,k)} : (q,k) \in \mathcal{C}\}$ as the conditional expectation of $\{\varepsilon^{(q,k)} : (q,k) \in \mathcal{C}\}$ given $\{y_t\}$ and $\sigma^2_{q,k} = \widehat{\sigma}^2_{q,k}$ for $(q,k) \in \mathcal{C}$. That is, $\widehat{\varepsilon}^{(q,k)} = \left[H'_{q,k}V^{-1}_{-(q,k)}H_{q,k} + \frac{1}{\sigma^2_{q,k}}\boldsymbol{I}\right]^{-1} H'_{q,k}V^{-1}_{-(q,k)}(Dy - X\beta)$ for $(q,k) \in \mathcal{C}$. Moreover, the contribution of $\widehat{\varepsilon}^{(q,k)}$ to $(1-\mathbf{B})^d y_t$ can be computed as $H_{q,k}\widehat{\varepsilon}^{(q,k)}$. Note that Equation (20) in Appendix A.1 can be used to set the starting values of the parameters for fitting an ISD$(d, q+1)$ process if the ML estimates of the parameters of ISD(d,q) are available.

5 An example

In this section we consider readings of an "uncontrolled" chemical concentration process from Series A of Box and Jenkins (1970, p.525). The series is displayed in Figure 3(a) with a LOESS trend. Its first-order difference is displayed in Figure 3(b). Tapered periodograms for both the original series and its first-order differences are displayed in Figures 4(a) and 4(b).

We fit an IMA(1,1) or ISD$(1,1)$ model with a linear trend to the series. The ML estimates of the parameters are $\widehat{\delta} = 0.004045$, $\widehat{\sigma}_a = \widehat{\sigma}_0 = 0.04687$, and $\widehat{\sigma}_e = \widehat{\sigma}_1 = 0.2702$. The parameters in the conventional IMA(1,1) are then $\widehat{\delta} = 0.004045$, $\widehat{\theta} = -0.7044$, and $\widehat{\sigma}_\varepsilon = 0.3171$.

We also fit four ISD(1,q) models to the series with $q = 2$, 3, 4, and 16, respectively. The corresponding estimates of the power spectrum are displayed in Figures 4(c) and 4(d). The AIC model selection criterion (Akaike, 1974) suggests the model ISD(1,2). However, from the fit of ISD$(1,16)$ we see that the power spectral density has modes near $f = 0.16, 3.0$, and 4.5, corresponding to the periodicities of $T = 12$, 6, and 4 hours. In particular, the periodicity $T = 12$ hours might reflect a possible half-day effect. Accordingly, one may also suspect that there exists a possible daily T=24 effect.

To illustrate de-noising we consider removing the white noise component from the original series using the ISD$(1,\mathcal{C})$ model with $\mathcal{C} = \{(1,1),\ (2,0),\ (2,1),\ (15,1),\ (16,4)\}$, where the modes of the components $(15,1)$ and $(16,4)$ correspond to approximate periodicities of $T = 24$ and $T = 12$ hours, respectively. The use of a value of q close to 16 is based on consideration of the scales near the modes, as discussed in Section 3.2, along with the periodograms in Figures 4(a) and 4(b). The other two components $(2,0)$ and $(2,1)$ are chosen so that $\{(1,1),\ (2,0),\ (2,1)\}$ approximate the ISD$(1,2)$, as suggested by the AIC criterion. The component $(1,1)$ is not representable by

$\mathcal{C} \setminus \{(1,1)\}$. That is, the white noise component added to the original series is identifiable for this model and the estimate of $\sigma^2_{(1,1)}$ is unique. This white noise component is estimated as a by-product of the ECME algorithm. The de-noised series is obtained using (13). Figure 5 displays the de-noised and original series and a LOESS trend.

6 Discussion

This paper introduces a class of IMA models that we call integrated sum-difference (ISD) processes. The ISD class has many nice features, including 1) its models can be decomposed into additive independent moving average processes, each having a simple expression, 2) the decomposition allows the components to be estimated using EM-type algorithms, which are simple to implement and converge monotonically in terms of likelihood, and 3) ISD components can be selected using knowledge from the underlying science or from the available data. With minor changes, the algorithms presented in this paper also apply to the class of models $\psi(\mathbf{B})y_t = g(X_t, \beta) + \sum_k \phi_k(\mathbf{B})\varepsilon_t^{(k)}$, where, $\psi(\mathbf{B})$ and $\phi_k(\mathbf{B})$ are polynomial functions of $\mathbf{B}$ with known coefficients, $g(X, \beta)$ is a deterministic function with known $\{X_t\}$ and unknown parameters β, and $\varepsilon_t^{(k)}$ are independent white noise.

Constructing models using specific operators, such as $(1-\mathbf{B})^k(1+\mathbf{B})^{q-k}$, is not new. Box and Jenkins (1970, p. 302-303), for example, suggested the operator $1-\sqrt{3}\mathbf{B}+\mathbf{B}^2$ for monthly data and $1-\mathbf{B}^s$ for seasonal data. As Box and Jenkins (1970, p. 302) claimed, choosing the right operator is not a mathematical problem, but a question about how the world tends to behave. Decomposition is especially important because it makes interpreting a model in real-word terms easier. Unlike Box and Jenkins (1970), our model components are additive rather than multiplicative. Moreover, our model components can be easily estimated using the EM algorithm, which is both simple and stable.

The class of models discussed in this article can be useful for modeling complex data. Clark, et al (1999) provides such an example. Although we focused on univariate ISD processes, extensions to multivariate time series are straightforward. Hopke, Liu, and Rubin (2001) demonstrate that multivariate versions of ISD time series models are useful for multiple imputation of incomplete multivariate time series data with or without seasonal effects.

Acknowledgment

We thank Scott Vander Wiel for helpful discussion, and Diane Lambert, the Editor, the Associate Editor, and two referees for their insightful comments.

Appendix

A.1 The Levinson-Durbin algorithm

Given the auto-covariance function $r_0, r_1, ...$ of a stationary process $\{x_t\}$, the Levinson-Durbin algorithm is designed to compute $a_{p,1}, a_{p,2}, ..., a_{p,p}$ and σ_p^2 for $p = 1, 2, ...$ such that $a_{p,1}X_{t-1} + ... + a_{p,p}X_{t-p}$ is the minimum variance (one-step ahead) predictor of X_t from the p past values $(x_{t-1}, ..., x_{t-p})$ with the minimum variance σ_p^2. Starting with $a_{1,1} = r_1/r_0$ and $\sigma_1^2 = r_0(1 - a_{1,1}^2)$, the Levinson-Durbin algorithm (Levinson, 1946; Durbin, 1960; Dempster, Lecture notes) proceeds recursively as follows. Given $a_{p,1}, ...a_{p,p}$ and σ_p^2, $a_{p+1,p+1} = \frac{1}{\sigma_p^2}\left(r_{p+1} - \sum_{k=1}^{p} a_{p,k}r_{p-k+1}\right)$, $a_{p+1,k} = a_{p,k} - a_{p+1,p+1}a_{p,p-k+1}$ for $k = 1, ..., p$, and $\sigma_{p+1}^2 = \sigma_p^2\left(1 - a_{p+1,p+1}^2\right) = r_0 - \sum_{k=1}^{p} a_{p,k}r_k$.

The Levinson-Durbin algorithm can be used to compute the inverse of the Toeplitz matrix Ψ_t with the first row the autocovariance function $r_0, r_1, ...r_{t-1}$ of a Gaussian stationary series. Let A_t be the $(t \times t)$ lower triangular matrix with the elements $A_t(i,i) = 1$ for $i = 1, ..., t$ and $A_t(i,j) = -a_{i-1,i-j}$ for $i > j$. From the Levinson-Durbin's algorithm we have $A_t(x_1, ..., x_t)' \sim \mathrm{N}(0, \Lambda_t)$, where $\Lambda_t = \mathrm{Diag}(\sigma_0^2, ..., \sigma_{t-1}^2)$ and $\sigma_0^2 = r_0$. This implies the following equality $A_t\Psi_tA_t' = \Lambda_t$ or $\Psi_t^{-1} = A_t'\Lambda_t^{-1}A_t$, which provides an efficient way of computing the inverse of Toeplitz matrices.

A.2 Proofs of Propositions

A.2.1 *Proof of Proposition 1*

The spectrum of the Gaussian IMA(d,q) process in (10) is

$$\begin{aligned} 2^{d+1}\sin^{2d}(\pi f) \;\; p_y(f) &= 2\sigma_a^2 \prod_{k=1}^{q} (1 + \lambda_k e^{-i2\pi f})(1 + \lambda_k e^{+i2\pi f}) \\ &= 2\sigma_a^2 \prod_{k=1}^{q} \left[(1+\lambda_k)^2\cos^2(\pi f) + (1-\lambda_k)^2\sin^2(\pi f)\right]. \qquad (16) \end{aligned}$$

The proof is accomplished by noticing that (16) is a homogeneous polynomial of degree q in $\sin^2(\pi f)$ and $\cos^2(\pi f)$ with positive coefficients and that (9) represents the class of all the homogeneous polynomials of degree q in $\sin^2(\pi f)$ and $\cos^2(\pi f)$ with positive coefficients.

A.2.2 *Proof of Proposition 2*

A moving average process with real coefficients and real characteristic roots can be represented as $x_t = \theta_1 e_{t-2} + \theta_2 e_{t-1} + \theta_3 e_t$, where θ_1, θ_2, and θ_3 are real and $\theta_2^2 - 4\theta_1\theta_3 \geq 0$. Comparing the autocovariance function, we have

$$\theta_1^2 + \theta_2^2 + \theta_3^2 = 6\sigma_0^2 + 2\sigma_1^2 + 6\sigma_2^2, \tag{17}$$

$$\theta_2(\theta_1 + \theta_3) = 4\sigma_0^2 - 4\sigma_2^2, \tag{18}$$

$$\theta_1\theta_3 = \sigma_0^2 - \sigma_1^2 + \sigma_2^2. \tag{19}$$

From (17)+2×(19) and (18), we obtain $\theta_2^2 = 4(\sigma_0 \pm \sigma_2)^2$. Because $(\theta_1, \theta_2, \theta_3)$ and $(-\theta_1, -\theta_2, -\theta_3)$ are equivalent in the sense that they lead to the same moving average process, without losing generality we let $\theta_2 = 2(\sigma_0 \pm \sigma_2)$, that is, (I) $\theta_2 = 2(\sigma_0 + \sigma_2)$ or (II) $\theta_2 = 2(\sigma_0 - \sigma_2)$.

For case (I), we have $\theta_2 = 2(\sigma_0 + \sigma_2)$ and $\theta_1, \theta_3 = \sigma_0 - \sigma_2 \pm \sqrt{\sigma_1^2 - 2\sigma_0\sigma_2}$, which implies that the condition $\sigma_1^2 - 2\sigma_0\sigma_2 \geq 0$ must be satisfied to ensure that the coefficients θ_1, θ_2, and θ_3 are real. For case (II), we have $\theta_2 = 2(\sigma_0 - \sigma_2)$ and $\theta_1, \theta_3 = \sigma_0 + \sigma_2 \pm \sqrt{\sigma_1^2 + 2\sigma_0\sigma_2}$, which guarantees that the coefficients θ_1, θ_2, and θ_3 are real. The corresponding condition under which the characteristic equation has real roots is $\theta_1^2 - 4\theta_1\theta_3 = 4(\sigma_1^2 - 2\sigma_0\sigma_2) \geq 0$.

It is also easy to show that the condition $\sigma_1^2 - 2\sigma_0\sigma_2 > 0$ implies that both the coefficients θ_1, θ_2, and θ_3 and the characteristic roots are real.

A.2.3 *Proof of Proposition 3*

The power spectrum of the (q, k) component of $(1 - \mathbf{B})^d y_t$ can be written as follows:

$$\begin{aligned}
& 2^{2q+1}\sigma_{q,k}^2 \sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f) \\
= \ & 2^{2q+1}\sigma_{q,k}^2 \sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)\left(\sin^2(\pi f) + \cos^2(\pi f)\right) \\
= \ & 2^{2q+1}\sigma_{q,k}^2 \sin^{2(k+1)}(\pi f)\cos^{2((q+1)-(k+1))}(\pi f) + 2^{2q+1}\sigma_{q,k}^2 \sin^{2k}(\pi f)\cos^{2((q+1)-k)}(\pi f).
\end{aligned}$$

Thus, the power spectrum of $(1-\mathbf{B})^d y_t$ can be written as: $2^{2q+1}\sum_{k=0}^{q}\sigma_{q,k}^2\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f) = 2^{2(q+1)+1}\sum_{k=0}^{q+1}\sigma_{q+1,k}^2\sin^{2k}(\pi f)\cos^{2((q+1)-k)}(\pi f)$, where

$$\sigma_{q+1,k}^2 = \begin{cases} 2^{-2}\sigma_{q,0}^2, & \text{if } k=0; \\ 2^{-2}(\sigma_{q,k-1}^2+\sigma_{q,k}^2), & \text{if } k=1,...,q; \\ 2^{-2}\sigma_{q,q}^2, & \text{if } k=q+1. \end{cases} \tag{20}$$

Therefore, we proved that $y_t \sim \text{ISD}(d, q+1)$ if $y_t \sim \text{ISD}(d,q)$. Proposition 3 follows by induction.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, **AC-19**, 716-722.

Anderson, T. W. and Mentz, R. P. (1993). A note on maximum likelihood estimation in the first-order Gaussian moving average model. *Statistics & Probability Letters*, **16**, 205-211.

Box, G. E. P. and Jenkins G. M. (1970). *Time Series Analysis: forecasting and control.* Holden-Day.

Clark, L. A., Cleveland, W. S., Denby, L., and Liu, C. (1999). Modeling customer survey data (with discussion). In *Case Studies in Bayesian Statistics* Vol 4 (eds. C. Gatsonis, R. E. Kass, B. P. Carlin, Carriquiry, A. A. Gelman, Verdinelli, I., and Mike W.), 3-57.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **39**, 1-38.

Durbin, J. (1960). The fitting of time series models, *Rev. Inst. Internat. Statist.*, **28**, 233-244.

Hopke, P. K., Liu, C. H., and Rubin, D. B. (2001). Multiple imputation for multivariate data with missing and below-threshold measurements: Time-series concentrations of pollutants in the Arctic *Biometrics*, **57**, 22-33.

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.

Levinson, N. (1946). The Wiener RMS (root mean square) error criterion in filter design and prediction, *J. of Math. Phys.* **25**, 261-278.

Liu, C. H. and Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**, 633-648.

Meng, X. L., and van Dyk, D. (1997). The EM algorithm — an old folk song sung to fast new tune (with discussion). *Journal of the Royal Statistical Society,* **59**, 511-567.

Vander Wiel S. A. (1996). Statistical process monitoring using integrated moving averages, *Technometrics*, 38:1.

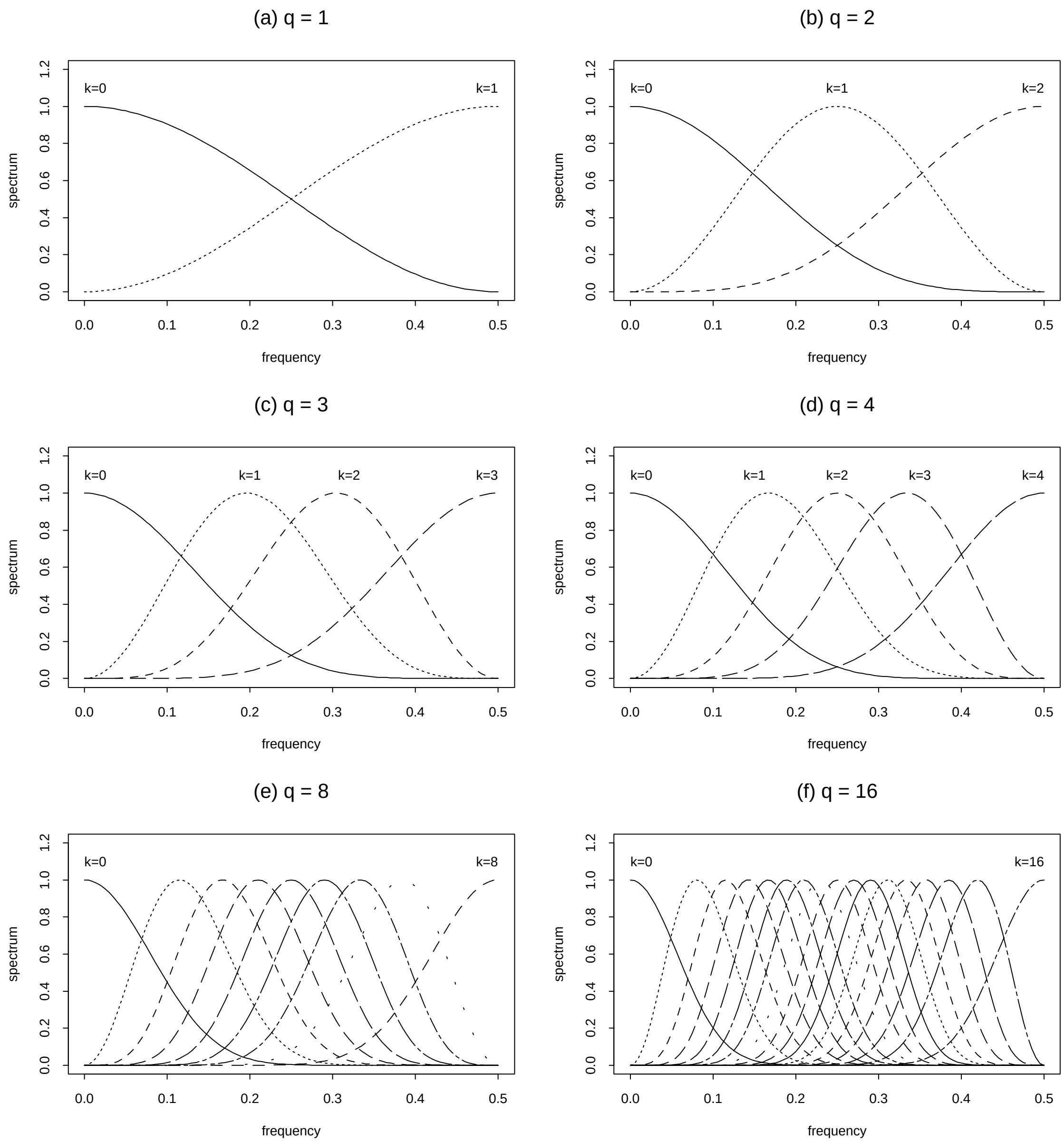

Figure 1: The frequency components $\sin^{2k}(\pi f)\cos^{2(q-k)}(\pi f)$ for $k = 0, ..., q$; (a) $q = 1$, (b) $q = 2$, (c) $q = 3$, (d) $q = 4$, (e) $q = 8$, and (f) $q = 16$. Each frequency component is scaled so that its maximum is one.

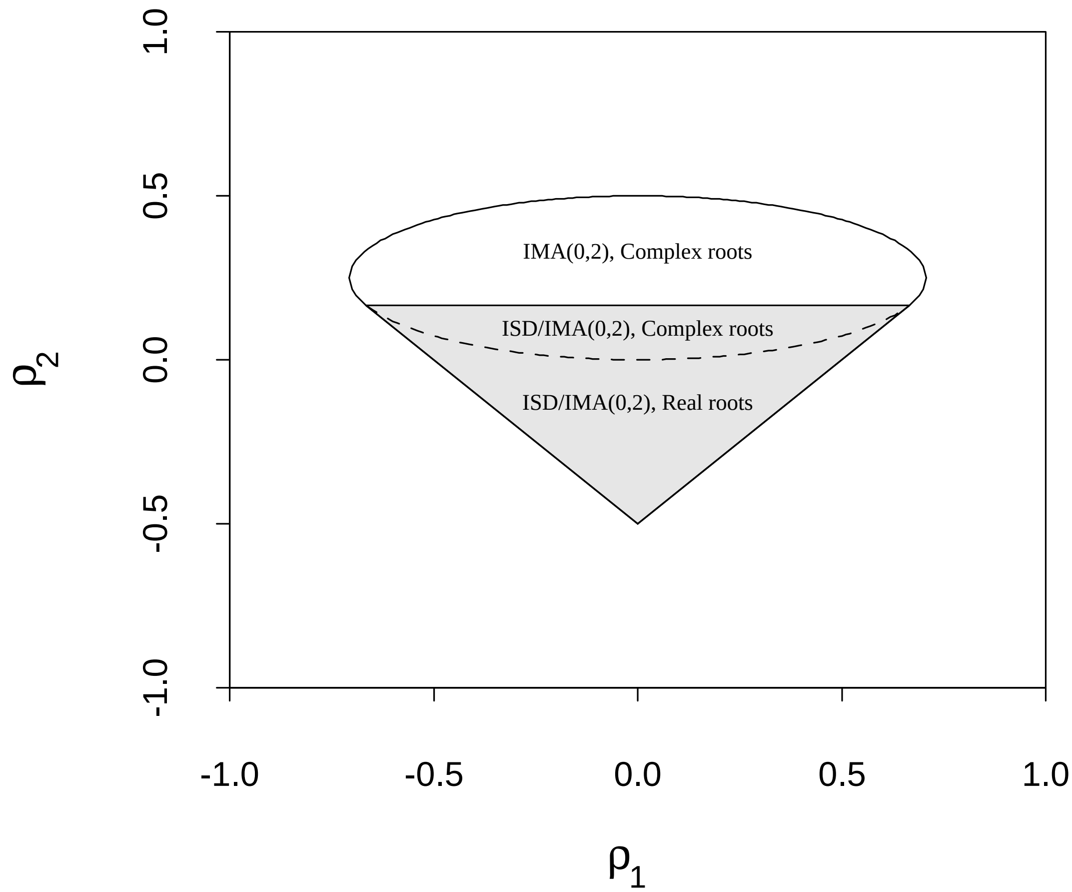

Figure 2: The areas of the first two autocorrelations ρ_1 and ρ_2 of the ISD$(0,2)$ and IMA$(0,2)$ processes.

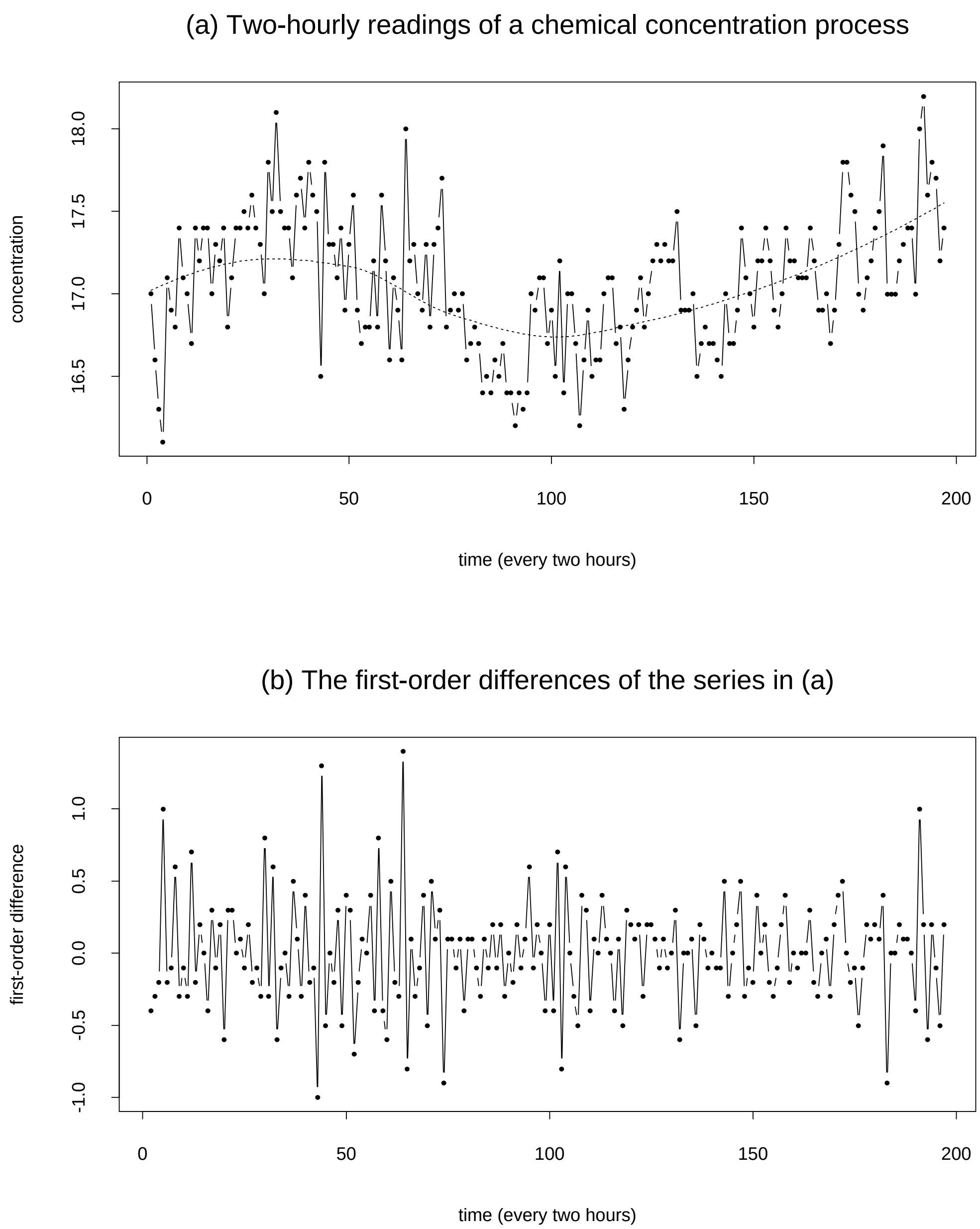

Figure 3: (a) Readings of a chemical concentration process (Box and Jenkins, 1970, p. 525), where the dotted line is the smooth trend from LOESS fit. (b) the first-differences of the series in (a).

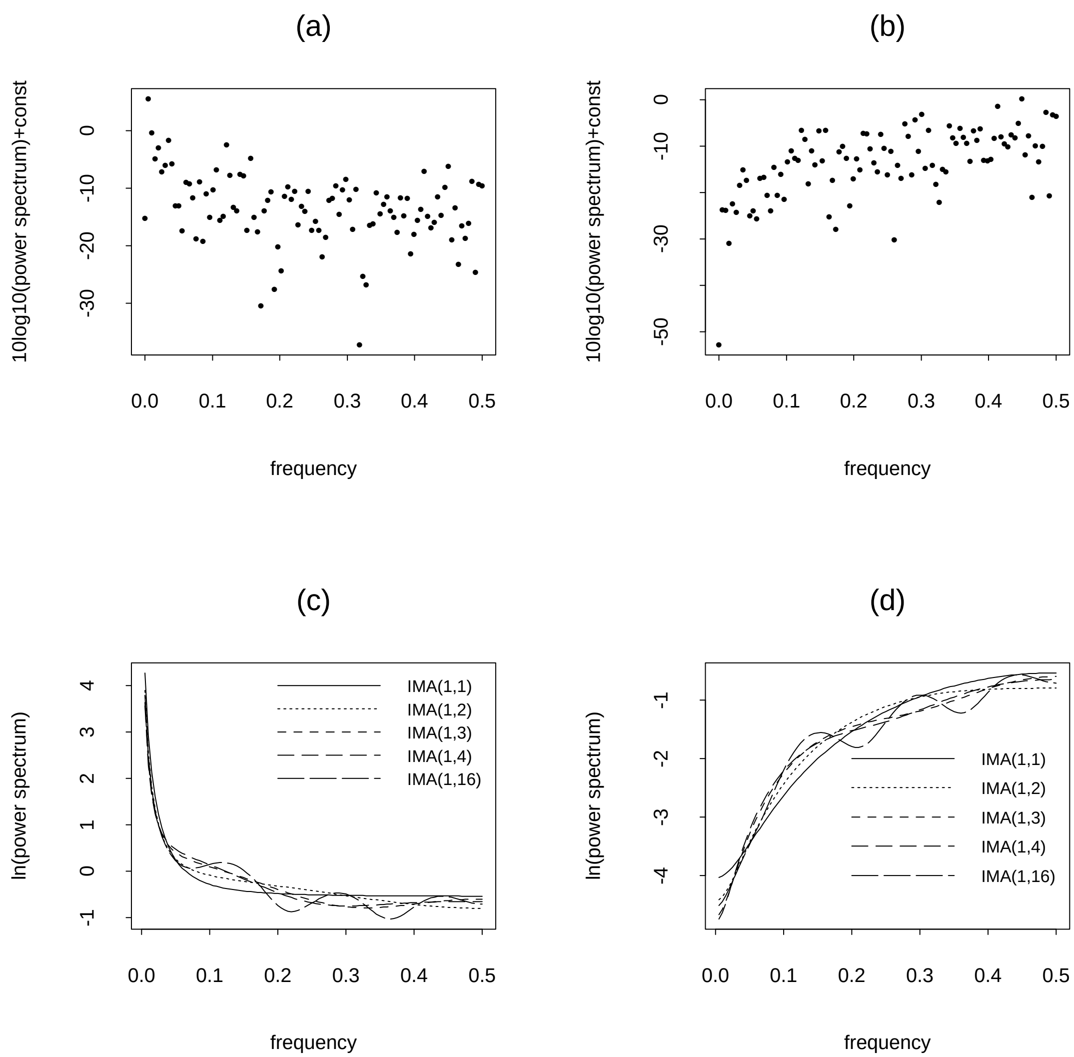

Figure 4: (a) and (b) are tapered periodograms of the original series in the example and its first-order differences, respectively. (c) and (d) are the corresponding estimates based on a set of IMA processes.

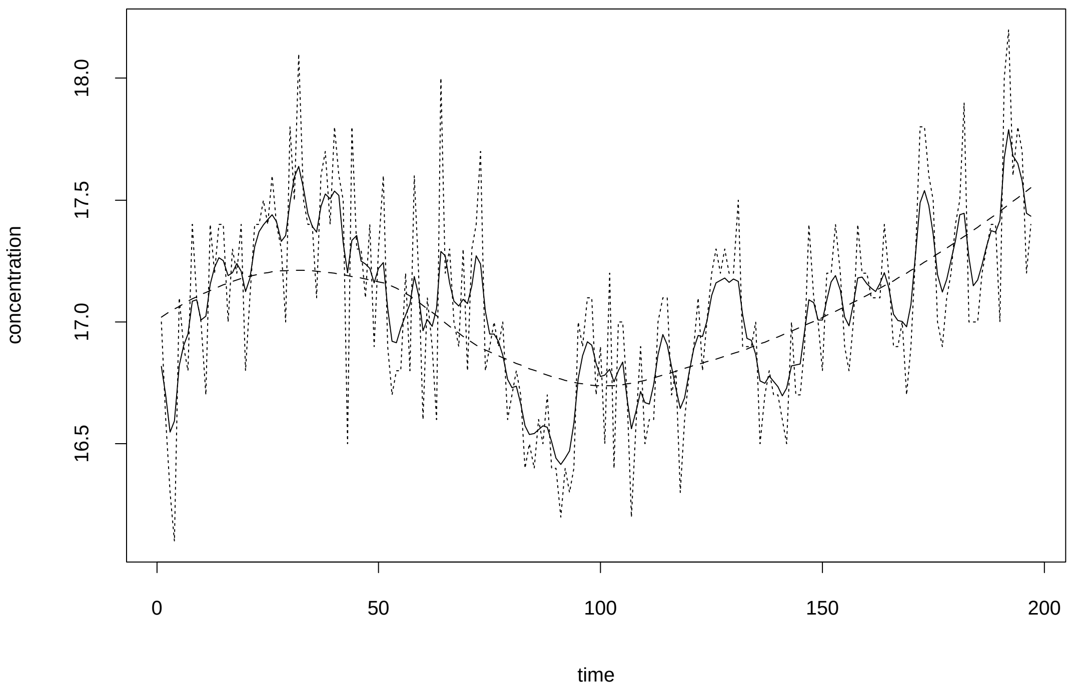

Figure 5: The original series in the example (dotted line), the de-noised series based on $\text{IMA}(1, \{(1,1),(2,0),(2,1),(15,1),(16,4)\})$ (solid line), and a LOESS trend (dashed line).