Selecting optimal ARMA order by a minimum spectrum distance criterion

Y. M. Zhang† and Q. M. Zhu‡§

In general, an optimal model which has the best effect in a specific application among the different estimated models may not necessarily have the same order as the actual model. The objective here is to propose a criterion for selecting the optimal order model which gives the minimum ARMA spectral estimation error. The model spectral distance (MSD) is proposed to measure the difference between two models in terms of their spectral characteristics. In order to select the optimal order model, the modelling error is decomposed into two components using the measurements of MSD. One component is caused by the parameter estimation (PE ME), and the other is due to order insufficiency (OIME). Based on the error decomposition, the model spectral distance criterion (MSDC) is proposed to select the order by comparing the changes in the PE ME and OIME. Theoretically, orders selected by the MSDC are optimal in minimizing the MSD from the estimated model to the actual model. However, in implementation, approximations must be made. To minimize their influence on the order selection, the approximations have been designed so that the selected models are optimal or nearly optimal. Simulations have been conducted to demonstrate the new approach.

1. Introduction

Order selection is a fundamental step in identifying autoregressive moving average (ARMA) models. A strong interest in order selection methods has been maintained for two decades, partially due to the importance of order selection in modelling, but also because of the lack of satisfactory criteria to meet various applications.

Numerous approaches have been proposed to select the order of an ARMA model. The AIC selects orders based on the concept of entropy (Akaike 1974). The approach proposed by Fuchs (1987) is based on the rank determination of estimated covariance matrices. Ribeiro and Moura (1988, 1991) determine the order by minimizing a functional $d$ that measures the mismatch of the assumed model to the data. This functional is evaluated from the estimated reflection coefficient sequence associated with the process. Hall (1991) and Burshtein and Weinstein (1991) proposed approaches based on an instrumental variable and a Wald statistic, respectively. Although the principles of these criteria may differ, a similarity does exist. In fact, the model goodness has been measured based on a single property associated with the ARMA process, without any direct consideration of the properties of the process which are related to the model application. The percentage of times in which the order selected is the same as in the actual models is calculated in simulations to show the superiority of the proposed approaches (Fuchs 1987, Burshtein and Weinstein 1991, Ribeiro and Moura 1988, 1991), even if the sample sizes are small. However, the estimated model with the same order as the actual model may not be the most appropriate when the samples are limited (see section 4).

As a matter of fact, each individual property associated with the process is only a reflection of model goodness from each specific point of view. A full description of model goodness cannot be made through a single property. A model selected by one criterion may not be selected by another criterion. Thus, application-oriented optimal order models should be selected from different estimated models.

In this paper the spectrum-oriented order selection of ARMA models is addressed. Among the existing order
selection criteria, the final prediction error (FPE) criterion (Akaike 1969) and the criterion for autoregressive transfer function (CAT) (Parzen 1974) can be regarded as application-oriented. However, they work only for autoregressive (AR) models. Although numerous approaches have been proposed to obtain linear ARMA spectral estimates (Cadzow 1982, 1983, Fassois 1990, Park and Gerharde 1989), few authors have been involved in order selection. Cadzow (1982, 1983) proposed selection of the order by examining the normalized ratio $\psi(k)$ of the extended order autocorrelation matrix estimate using a singular value decomposition. In this method, the order of the AR part is set equal to the smallest $k$ for which $\psi(k)$ is deemed adequately close to one. However, the description adequately close to one is subjective. Also, no relationship between the normalized ratio and spectral estimation has been studied. Thus, in this paper we will propose a new objective criterion for order selection for ARMA spectral estimation.

2. Model spectral distance

Suppose that $y_t$ is described by a stationary-invertible ARMA model (Box and Jenkins 1976):

$$
\phi(B)y_t = \theta(B)e_t,
$$

where $B$ is the backshift operator, $e_t \sim N(0, \sigma_e^2)$, and

$$
\phi(B) := 1 - \sum_{j=1}^{p} \phi_j B^j, \quad \theta(B) := 1 - \sum_{j=1}^{q} \theta_j B^j,
$$

where $\phi_j$, $j = 1, \ldots, q$, and $\theta_j$, $j = 1, \ldots, q$, are real, $|\phi_j| < 1$, $j = 1, \ldots, p$, and $|\theta_j| < 1$, $j = 1, \ldots, q$, and $\phi(B)$ and $\theta(B)$ are coprime. We also use the vectors

$$
\phi = (\phi_1, \ldots, \phi_p)^T, \quad \theta = (\theta_1, \ldots, \theta_q)^T.
$$

(3)

to denote the parameters of the ARMA model.

The power spectral density of $y_t$ can be written as (Box and Jenkins 1976):

$$
p(\omega) = |H(\omega)|^2 \sigma_e^2 = \frac{\sigma_e^2}{2\pi} \left| \frac{\phi(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2 = \frac{\sigma_e^2}{2\pi} u^2(\omega) + \frac{\sigma_e^2}{2\pi} v^2(\omega),
$$

(4)

where

$$
H(\omega) = \frac{1}{\sqrt{2\pi} \phi(e^{-i\omega})} = \frac{1}{\sqrt{2\pi} a(\omega) + ib(\omega)},
$$

(5)

and

$$
da(\omega) = 1 - \sum_{j=1}^{p} \phi_j \cos(j\omega),
$$

$$
b(\omega) = \sum_{j=1}^{p} \phi_j \sin(j\omega),
$$

$$
u(\omega) = 1 - \sum_{j=1}^{q} \theta_j \cos(j\omega),
$$

$$
\nu(\omega) = \sum_{j=1}^{q} \theta_j \sin(j\omega).
$$

(6)

Consider $H^{(1)}(\omega)$ and $H^{(2)}(\omega)$:

$$
H^{(k)}(\omega) = \frac{u^{(k)}(\omega) + iv^{(k)}(\omega)}{\sqrt{2\pi} a^{(k)}(\omega) + ib^{(k)}(\omega)}, \quad k = 1, 2.
$$

(7)

Assume that the corresponding $\phi^{(k)}(B)$ and $\theta^{(k)}(B)$, $k = 1, 2$, are stationary. The relative error between these two functions at $\omega$ can be described by

$$
d(H^{(2)}(\omega) \rightarrow H^{(1)}(\omega)) := \frac{H^{(2)}(\omega) - H^{(1)}(\omega)}{H^{(1)}(\omega)}.
$$

(8)

Here $H^{(1)}$ is used as the reference for generating the relative error. By averaging the relative error for different frequencies, one can define

$$
D(H^{(2)} \rightarrow H^{(1)}) := \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |d(\omega)|^2 d\omega \right]^{1/2}.
$$

(9)

as the model spectral distance (MSD) from $H^{(2)}$ to $H^{(1)}$, where the square root is used to acquire the distance. It is evident that the MSD can be used to measure the difference for two different models based on the spectral characteristics. In the time domain, a similar concept, i.e. the model distance, has been proposed and used in a number of modelling applications (Kovacevic and Zhang 1992, Kovacevic et al. 1993, 1995).

3. Modelling error decomposition

Assume that an ARMA$(p,q)$ model (1) is given, where $\phi_p \neq 0$ and $\theta_q \neq 0$. Assume that ARMA$(n,m)$ is an estimate of ARMA$(p,q)$, and thus is denoted by $\hat{H}(\omega) = \frac{\hat{u}(e^{-i\omega})}{\sqrt{2\pi} \phi(e^{-i\omega})}$ or by the parameter vectors $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_n)^T$ and $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$. The distance from $\hat{H}$ to $H$ is

$$
D(\hat{H} \rightarrow H) = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |d(\omega)|^2 d\omega \right]^{1/2}.
$$

(10)

where

$$
d = \frac{\hat{H}(\omega) - H(\omega)}{H(\omega)}.
$$

(11)
Because $|d|^2 \geq 0 \quad \forall \omega$, the necessary and sufficient condition for $D(\hat{H} \to H) = 0$ is

$$d = 0, \quad \forall \omega. \quad (12)$$

In addition, the spectral density of a stationary process is not infinite. Thus, the necessary and sufficient condition becomes

$$\hat{H}(\omega) = H(\omega), \quad \forall \omega. \quad (13)$$

It can further be shown that this condition is equivalent to

$$\begin{align*}
\hat{\phi}_j &= \phi_j, \quad j = 1, \ldots, p, \\
\hat{\theta}_j &= \theta_j, \quad j = 1, \ldots, q, \\
\hat{\phi}_j &= 0, \quad n \geq j \geq p + 1, \\
\hat{\theta}_j &= 0, \quad n \geq j \geq +1.
\end{align*} \quad (14)$$

The necessary condition for satisfying (14) is $n \geq p, \quad m \geq q$. Otherwise

$$D(\hat{H} \to H) > 0, \quad \forall \phi \in R^p, \forall \theta \in R^q. \quad (15)$$

That is, if the used order is not sufficient so that $n < p$ and/or $m < q$, the modelling error must exist despite the size of the samples and the algorithm used for parameter estimation. For the given $ARMA(p, q)$ and $(n, m)$, the minimum $D(\hat{H} \to H)$ can be defined as

$$D_{\min}(\hat{H} \to H) = \min_{\phi \in R^p} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |d|^2 \, d\omega \right]^{1/2}. \quad (16)$$

Assume that the corresponding parameter vectors which minimize the distance are

$$\begin{align*}
\hat{\phi}^* &= (\hat{\phi}_1^*, \hat{\phi}_2^*, \ldots, \hat{\phi}_n^*)^T, \\
\hat{\theta}^* &= (\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_m^*)^T.
\end{align*} \quad (17)$$

Thus $\hat{\phi}^*$ and $\hat{\theta}^*$ define an optimal $ARMA(n, m)$ approximation of the given $ARMA(p, q)$ in the sense of minimizing the MSD. We define

$$D(\hat{H}^* \to H) := D_{\min}(\hat{H} \to H), \quad (18)$$

as OME.

In the modelling, $H$ is unknown and $\hat{H}$ must be estimated from the samples. In general, the number of samples is not infinite so it cannot be expected that $\hat{H} = \hat{H}^*$. The actual MSD from $\hat{H}$ to $H$ must be larger than $D(\hat{H}^* \to H)$. In order to decompose the modelling error, consider

$$d = \frac{\hat{H}(\omega) - H(\omega)}{H(\omega)} = \frac{(\hat{H}(\omega) - \hat{H}^*(\omega)) + (\hat{H}^*(\omega) - H(\omega))}{H(\omega)}. \quad (19)$$

Denote

$$\begin{align*}
\hat{H}(\omega) &= \hat{H}_p(\omega) + i\hat{H}_q(\omega), \\
\hat{H}^*(\omega) &= \hat{H}^*_p(\omega) + i\hat{H}^*_q(\omega), \\
H(\omega) &= H_p(\omega) + iH_q(\omega),
\end{align*} \quad (20)$$

where $\hat{H}_p(\omega), \hat{H}_q(\omega), \hat{H}^*_p(\omega), \hat{H}^*_q(\omega), H_p(\omega)$ and $H_q(\omega)$ are real. Thus

$$|d|^2 = \frac{(\hat{H}_p(\omega) - \hat{H}^*_p(\omega))^2 + (\hat{H}_q(\omega) - H_q(\omega))^2 + (\hat{H}^*_q(\omega) - H_q(\omega))^2 + (\hat{H}_q(\omega) - H_q(\omega))^2}{|H(\omega)|^2} \quad (21)$$

In practical modelling, $\hat{H}$ is known, and $H$ and $\hat{H}^*$ are deterministic although they are unknown. In analysis, the error $\hat{H}^* - H$ is regarded as being deterministic whereas the error $\hat{H} - \hat{H}^*$ is random. Hence, the expectation of $|d|^2$ is

$$E(\hat{H}_p(\omega) - \hat{H}^*_p(\omega))^2 + E(\hat{H}_q(\omega) - H_q(\omega))^2 + E(\hat{H}^*_q(\omega) - H_q(\omega))^2 + (\hat{H}_q(\omega) - H_q(\omega))^2$$

$$\frac{E(\hat{H}_p(\omega) - \hat{H}^*_p(\omega))^2 + (\hat{H}^*_q(\omega) - H_q(\omega))^2 + (\hat{H}_q(\omega) - H_q(\omega))^2}{|H(\omega)|^2} \quad (22)$$

It can be shown that

$$\sqrt{2\pi}(\hat{H}(\omega) - \hat{H}^*(\omega))$$

$$\frac{\hat{\phi}(e^{-i\omega}) + \Delta \phi(e^{-i\omega})}{\hat{\theta}(e^{-i\omega}) + \Delta \theta(e^{-i\omega})}$$

$$\frac{\hat{\phi}(e^{-i\omega}) - \hat{\phi}(e^{-i\omega})\Delta \theta(e^{-i\omega})}{\hat{\theta}(e^{-i\omega})\Delta \phi(e^{-i\omega})} \quad (23)$$

where
\[
\Delta \phi (e^{-j\omega}) = \hat{\phi} (e^{-j\omega}) - \phi^*(e^{-j\omega}),
\]
\[
\Delta \theta (e^{-j\omega}) = \hat{\theta} (e^{-j\omega}) - \theta^*(e^{-j\omega}).
\]

From (6)
\[
\Delta \phi (e^{-j\omega}) = - \sum_{j=1}^{n} \Delta \phi_j [\cos (j\omega) - i \sin (j\omega)],
\]
\[
\Delta \theta (e^{-j\omega}) = - \sum_{j=1}^{m} \Delta \theta_j [\cos (j\omega) - i \sin (j\omega)],
\]
where
\[
\Delta \phi_j = \hat{\phi}_j - \phi^*_j,
\]
\[
\Delta \theta_j = \hat{\theta}_j - \theta^*_j,
\]
is the error of the parameter estimation. When an unbiased parameter estimation algorithm is used, so that
\[
E \Delta \phi_j = 0,
\]
\[
E \Delta \theta_j = 0,
\]
the following holds:
\[
E \Delta \phi (e^{-j\omega}) = 0,
\]
\[
E \Delta \theta (e^{-j\omega}) = 0.
\]

Because \( \hat{H} = \hat{\theta}(e^{-j\omega})/\sqrt{2\pi} \hat{\phi}(e^{-j\omega}) \) and \( \hat{H}^* = \hat{\theta}^*(e^{-j\omega})/\sqrt{2\pi} \hat{\phi}^*(e^{-j\omega}) \) are known or deterministic
\[
E (\hat{H}(\omega) - \hat{H}^*(\omega))
\]
\[
= \frac{1}{\sqrt{2\pi}} \frac{E [\hat{\theta}(e^{-j\omega}) \Delta \phi (e^{-j\omega}) - \hat{\phi}^*(e^{-j\omega}) \Delta \theta (e^{-j\omega})]}{\hat{\theta}(e^{-j\omega}) \hat{\theta}^*(e^{-j\omega})}
\]
\[
= \frac{1}{\sqrt{2\pi}} \frac{\hat{\theta}(e^{-j\omega}) E \Delta \phi (e^{-j\omega}) - \hat{\phi}^*(e^{-j\omega}) E \Delta \theta (e^{-j\omega})}{\hat{\theta}(e^{-j\omega}) \hat{\theta}^*(e^{-j\omega})} = 0.
\]

Thus
\[
E[d_1^2] = E \left| \frac{\hat{H}(\omega) - \hat{H}^*(\omega)}{H(\omega)} \right|^2 + \left| \frac{\hat{H}^*(\omega) - H(\omega)}{H(\omega)} \right|^2.
\]

Therefore
\[
ED^2 (\hat{H} \rightarrow H) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ E \left| \frac{\hat{H}(\omega) - \hat{H}^*(\omega)}{H(\omega)} \right|^2 + \left| \frac{\hat{H}^*(\omega) - H(\omega)}{H(\omega)} \right|^2 \right\} d\omega
\]
\[
+ D^2 (\hat{H} \rightarrow H).
\]

It can be seen that the first term corresponds to the difference between \( \hat{H} \) and \( \hat{H}^* \), and is actually a description of the modelling error component caused by the parameter estimation. Thus, the modelling error, measured by \( D^2 (\hat{H} \rightarrow H) \), can be composed into OI ME and PE ME. In this paper we denote
\[
D_H (\hat{H} \rightarrow \hat{H}^*) = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{H}(\omega) - \hat{H}^*(\omega)}{H(\omega)} \right|^2 d\omega \right]^{1/2},
\]
as the modified model spectral distance (MMSD) from \( \hat{H} \) to \( \hat{H}^* \) with \( H \) as the reference. For convenience of discussion, the MSD defined in (8) and (9) can also be regarded as the standard model spectral distance (SMSD). It can be seen that both MMSD and SMSD can be used to describe the difference between two models. However, for the MMSD case the reference for computation of the relative error can be a third model, whereas in the MSD case this reference must be one of the two models whose difference is being measured. Thus
\[
ED^2 (\hat{H} \rightarrow H) = ED_H^2 (\hat{H} \rightarrow \hat{H}^*) + D^2 (\hat{H} \rightarrow H).
\]

That is, the total modelling error, measured by \( ED^2 (\hat{H} \rightarrow H) \), is the sum of the PE ME and OI ME which are measured by \( ED_H^2 (\hat{H} \rightarrow \hat{H}^*) \) and \( D^2 (\hat{H} \rightarrow H) \), respectively.

### 4. Modelling errors and orders

It is known that the regression residuals decrease as the order increases. However, the residuals in the regression equations are not exactly the same as the modelling errors. The modelling errors are actually measured by the new residuals produced when the estimated model is used. Higher order models can acquire smaller regression residuals so that the existing data can be fit better, but the resultant models could be invalid and larger new residuals could be produced when applied.

As the order increases, the OI ME component decreases. This decrease is caused by the increased flexibility of the parameters. However, for the parameter estimation, the error tends to increase as the number of the parameters used increases. In fact parameter estimation for an ARMA model involves identifying the model parameters from the outputs of a noise-driving process. The accuracy of the parameter estimation depends on the sample size and the number of parameters. In order to improve the accuracy, the sample size to parameter number ratio should be increased so that the stochastic influence of the process is reduced. When the sample size is given, the estimation error increases as the parameter number increases.

Because of the distinctive trends, an increase in the order will produce opposite influences on OI ME and
PE ME. In addition, when the order is lower, the decrease in OI ME will be dominant, whereas the increase in PE ME will be gradual. Once the order becomes sufficient OI ME will be very small or zero. The resultant decrease in the OI ME therefore becomes slight or zero, while PE ME continues to increase. Thus, theoretically, there must be a specific order above which the decrease in OI ME is less then the increase in PE ME.

In order to show the above characteristics associated with the OI ME and PE ME, theoretical results will be derived in this section. To simplify the theoretical derivation, time domain error measurement and autoregressive-average (AR) models will be used.

Assume that \( y_t \) is an AR\((p)\) process:

\[
y_t = \phi^t Y_t + \epsilon_t, \tag{34}
\]

where \( \epsilon_t \sim N(0, \sigma^2) \)

\[
\phi = (\phi_1, \ldots, \phi_p)^T, \quad \phi_p \neq 0,
\]

\[
Y_t = (y_{t-1}, \ldots, y_{t-p})^T.
\]

For any \( n < p \), the parameter vector can be partitioned into

\[
\phi = (\phi^T(1:n)\phi^T(n+1:p))^T, \tag{36}
\]

where

\[
\phi(1:n) = (\phi_1, \ldots, \phi_n)^T,
\]

\[
\phi(n+1:p) = (\phi_{n+1}, \ldots, \phi_p)^T.
\]

4.1. Modelling error

Assume that an AR\((n)\) \((n < p)\) is used to approximate the AR\((p)\) process in (34), where the parameters consist of an \( n \times 1 \) vector \( \hat{\phi}(1:n) \). For derivation convenience, denote

\[
\hat{\phi} = (\phi^T(1:n)\phi^T(n+1:p))^T,
\]

\[
\Delta\phi = \phi - \hat{\phi},
\]

\[
\Delta\phi(1:n) = \phi(1:n) - \hat{\phi}(1:n),
\]

where \( O(n+1:p) \) is a \( (p-n) \times 1 \) null vector.

When the AR\((n)\) is used, the one-step-ahead prediction is

\[
\hat{y}_t = \phi^t Y_t. \tag{39}
\]

The prediction error (residual) is

\[
\hat{\epsilon}_t := y_t - \hat{y}_t = \Delta\phi^t Y_t + \epsilon_t \tag{40}
\]

Thus

\[
E(\hat{\epsilon}_t^2) = E(\hat{\epsilon}_t \hat{\epsilon}_t^T) = E(\Delta\phi^t Y_t Y_t^T \Delta\phi) + E(\epsilon_t^2) + 2E(\Delta\phi^T \epsilon_t) \\
= \Delta\phi^T E(Y_t Y_t^T) \Delta\phi + \sigma^2 + 2\Delta\phi^T E(Y_t \epsilon_t) \\
= \Delta\phi^T R\Delta\phi + \sigma^2, \tag{41}
\]

where \( E(Y_t \epsilon_t) = 0 \) if due to the causality, and \( R \) is the covariance matrix:

\[
R = \begin{bmatrix} r(0) & r(1) & \ldots & r(p-1) \\
               r(1) & r(0) & \ldots & r(p-2) \\
               \vdots & \vdots & \ddots & \vdots \\
               r(p-1) & r(p-2) & \ldots & r(0) \end{bmatrix}, \tag{42}
\]

and

\[
r(j) = E(y_{t+j} y_{t-1}). \tag{43}
\]

It is known that \( \sigma^2 \) is the minimum variance of the onestep-ahead prediction error for the given process \( y_t \). When the AR\((n)\) model is used, additional prediction errors are induced so that

\[
E(\hat{\epsilon}_t^2) - \sigma^2 = \Delta\phi^T R\Delta\phi > 0, \quad \forall n < p. \tag{44}
\]

In fact, the \( p \times p \) covariance matrix \( R \) of an AR\((p)\) process is positive definite. Thus, the necessary and sufficient condition for \( \Delta\phi^T R\Delta\phi = 0 \) is \( \Delta\phi = 0 \). However, it has been assumed that \( \Delta\phi \neq 0 \), \( \forall n < p \). Hence, an additional error must be caused when a lower model is used. In the time domain, this additional error can be used to measure the modelling error.

4.2. OI ME

For a given lower order, the additional error can be minimized. From (44) it can be shown that the additional error, measured using \( A(\hat{\phi}(1:n) \rightarrow \phi) \), can be calculated by

\[
A(\hat{\phi}(1:n) \rightarrow \phi) := E(\hat{\epsilon}_t^2) - \sigma^2 \\
= \Delta\phi^T (1:n) R_{11} \Delta\phi(1:n) \\
+ 2\Delta\phi^T (1:n) R_{12} \phi(n+1:p) \\
+ \phi^T (n+1:p) R_{22} \phi(n+1:p), \tag{45}
\]

where \( R_{11}, R_{12} \) and \( R_{22} \) are defined by

\[
R = \begin{bmatrix} R_{11} & R_{12} \\
               R_{12}^T & R_{22} \end{bmatrix}. \tag{46}
\]

The dimensions are \( n \times n \), \( n \times (p-n) \) and \( (p-n) \times (p-n) \), respectively. Define the optimal parameter vector \( \hat{\phi}(1:n) \) as an \( n \times 1 \) real vector which minimizes the additional error
\[ A(\hat{\phi}^*(1:n) \rightarrow \phi) = \min_{\hat{\phi}(1:n) \in R} A(\hat{\phi}(1:n) \rightarrow \phi). \]  

(47)

Thus, OI ME can be measured using \( A(\hat{\phi}^*(1:n) \rightarrow \phi) \). It is known that

\[ \frac{\partial A(\hat{\phi}(1:n) \rightarrow \phi)}{\partial \hat{\phi}(1:n)} \bigg|_{\hat{\phi}(1:n) = \hat{\phi}^*(1:n)} = 2R_{11} \Delta \hat{\phi}^*(1:n) + 2R_{12} \hat{\phi}(n+1:p) = 0, \quad n \geq 1, \]

(48)

where \( \Delta \hat{\phi}^*(1:n) := (1:n) - \hat{\phi}^*(1:n) \). As \( R_{11} > 0 \), \( \Delta \hat{\phi}^*(1:n) \) has a unique solution:

\[ \Delta \hat{\phi}^*(1:n) = -R_{11}^{-1} R_{12} \hat{\phi}(n+1:p). \]

(49)

Hence

\[ A(\hat{\phi}^*(1:n) \rightarrow \phi) = \hat{\phi}(n+1:p)R_{12} \hat{\phi}(n+1:p) \]

\[ - \hat{\phi}(n+1:p)R_{12}^T R_{11}^{-1} R_{12} \hat{\phi}(n+1:p). \]

(50)

That is, when an AR(\( n \)), \( n < p \), is used to approximate the given AR(\( p \)) model, OI ME can be exactly determined.

The following is apparent:

\[ \phi^T(n+1:p)R_{12} \hat{\phi}(n+1:p) = A(\hat{\phi}(1:n) \rightarrow \phi) \big|_{\hat{\phi}(1:n) = \phi^*(1:n)}. \]

(51)

Because the quadratic

\[ \phi^T(n+1:p)R_{12}^T R_{11}^{-1} R_{12} \hat{\phi}(n+1:p) \]

\[ = (R_{12} \hat{\phi}(n+1:p))^T R_{11}^{-1} (R_{12} \hat{\phi}(n+1:p)), \]

(52)

is positive definite, the optimal lower order approximation of a higher order model cannot be acquired by simply letting \( \hat{\phi}(1:n) = \phi^*(1:n) \).

In order to examine how \( A(\phi^*(1:n) \rightarrow \phi) \) varies with \( n \), this can be compared with \( A(\phi^*(1:n+1) \rightarrow \phi) \). It is known that

\[ A(\hat{\phi}^*(1:n+1) \rightarrow \phi) = \min_{\hat{\phi}(1:n+1) \in R^{n+1}} A(\hat{\phi}(1:n+1) \rightarrow \phi) \]

\[ \leq A\left( \begin{bmatrix} \phi^*(1:n) \\ 0 \end{bmatrix} \rightarrow \phi \right) \]

\[ = A(\phi^*(1:n) \rightarrow \phi). \]

(53)

Also

\[ A(\hat{\phi}^*(1:n) \rightarrow \phi) > 0, \quad \forall n < p, \]

\[ A(\hat{\phi}^*(1:n) \rightarrow \phi) = 0, \quad \forall n \geq p. \]

(54)

Thus, \( A(\phi^*(1:n) \rightarrow \phi) \), and therefore OI ME, decreases as the order increases until \( n \geq p \).

### 4.3. PE ME

Assume that the least square algorithm (Pandit and Wu 1983) is used to estimate the parameters. In order to decouple PE ME from OI ME, let \( n = p \). The parameter estimate is

\[ (\hat{\phi}_s, \ldots, \hat{\phi}_p)^T = [\Phi^T \Phi]^{-1} \Phi^T Y, \]

(55)

where

\[ \Phi = \begin{bmatrix} y_p, \ldots, y_1 \\ y_{p+1}, \ldots, y_2 \\ \vdots \\ y_{N-1}, \ldots, y_{N-p} \end{bmatrix}, \]

(56)

\[ Y = [y_p, \ldots, y_1]^T, \]

(57)

and \( N \) is the number of samples. In this case, the additional error is only caused by the parameter estimation. Thus, the PE ME can be directly measured using the additional error

\[ A(\hat{\phi}(1:p) \rightarrow \phi) = \Delta \phi^T \Delta \phi = \sum_{i=1}^{p} \sum_{j=1}^{p} (\Delta \phi_i \Delta \phi_j) \tau(|i-j|), \]

(58)

where

\[ \Delta \phi = \phi - \hat{\phi}, \]

\[ \Delta \phi_j = \phi_j - \hat{\phi}_j, \quad j = 1, 2, \ldots, 0. \]

(59)

It is known that the parameter estimates are stochastic and will vary depending on which segment of the data is employed. The used segment of the data is only one realization of the stochastic process. In order to study the statistics of PE ME, the expectation of the additional error should be taken for different estimates which are acquired using different realizations of the process. This expectation is

\[ EA(\hat{\phi}(1:p) \rightarrow \phi) = \sum_{i=1}^{p} \sum_{j=1}^{p} E(\Delta \phi_i \Delta \phi_j) \tau(|i-j|). \]

(60)

It is known that the covariance matrix of the parameter estimation errors is (Ljung 1987):

\[ V = [\Phi^T \Phi]^{-1} \sigma^2 \approx \frac{\sigma^2}{N} R^{-1}, \]

(61)

where the \( i \)th row \( j \)th column element of \( V \), denoted as \( V(i,j) \), is

\[ V(i,j) := E(\Delta \phi_i \Delta \phi_j). \]

(62)

Thus
\[ E A(\phi(1 : p) \rightarrow \phi) = \sum_{i=1}^{p} \sum_{j=1}^{p} V(i, j) R(i, j) \]

\[ \approx \frac{\sigma^2}{N} \sum_{i=1}^{p} \sum_{j=1}^{p} R^{-1}(i, j) R(i, j) = \frac{p}{N} \sigma^2. \]

When a lower order model AR(n), \( n < p \), is used, the parameter estimation algorithm tries to acquire the best lower order approximation. In this case, the parameter estimation error vector should be defined between the estimated lower order parameter vector \( \hat{\phi}(1 : n) \) and the optimal lower order parameter vector \( \phi^*(1 : n) \). In this case, PE ME should be measured by

\[ EA(\hat{\phi}(1 : n) \rightarrow \phi^*(1 : n)) \approx \frac{n}{N} \sigma^2, \quad n < p. \]  

When a higher order model AR(n), \( n > p \), is used, the parameter estimation error vector should be defined between the estimated higher order parameter vector and an extended parameter vector, i.e. \( (\phi^T, 0, \ldots, 0)^T \) where the number of the zeros equals \( (n - p) \). Denote the extended parameter vector as \( \phi(1 : n) := (\phi^T, 0, \ldots, 0)^T \). In this case, PE ME should be measured by

\[ EA(\hat{\phi}(1 : n) \rightarrow \phi(1 : n)) \approx \frac{n}{N} \sigma^2, \quad n > p. \]  

It can be seen that the expectation of the PE ME increases as the number of the used parameters increases, whereas the OI ME decreases. In addition, the following can be shown:

\[ EA(\hat{\phi}(1 : n) \rightarrow \phi) = EA(\hat{\phi}(1 : n) \rightarrow \phi^*) \]

\[ + A(\hat{\phi}^*(1 : n) \rightarrow \phi), \quad n < p. \]  

In fact, when \( \hat{\phi}(1 : n) \) and \( \phi^*(1 : n) \) are used, the one-step-ahead predictions are \( \hat{y}_n = \hat{\phi}^T(1 : n) Y_n(1 : n) \) (see (39)) and \( \hat{y}_n^* = (\phi^*(1 : n))^T Y_n(1 : 1 : n) \), respectively. Here the vector \( Y_n(1 : n) \) is defined as \( Y_n(1 : n) := (y_{n-1}, y_{n-2}, \ldots, y_{n-n})^T \). Thus

\[ \hat{e}_n := \hat{y}_n - y_n = (\hat{y}_n - \hat{y}_n^*) + (\hat{y}_n^* - y_n) \]

\[ = (\hat{\phi}(1 : n) - \phi^*(1 : n))^T Y_n(1 : n) + (\phi^* - \phi)^T Y_n + \epsilon_n, \quad (67) \]

where

\[ \phi^* = \begin{pmatrix} \phi^*(1 : n) \\ O(n + 1 : p) \end{pmatrix}. \]

Here the \( (p-n) \times 1 \) null vector \( O(n + 1 : p) \) has been defined before. From the definition given in (45), the following can be shown:

\[ A(\phi(1 : n) \rightarrow \phi) = E(\hat{\epsilon}^2 - \sigma^2) \]

\[ = (\hat{\phi}(1 : n) - \phi^*(1 : n))^T R_{11}(\hat{\phi}(1 : n) - \phi^*(1 : n)) \]

\[ - \phi^*(1 : n))^T + (\phi^* - \phi)^T R\phi^* - \phi \)

\[ + 2(\hat{\phi}(1 : n) - \phi^*(1 : n))^T (R_{11} R_{12}) \]

\[ \times (\phi^* - \phi) \]

\[ = A(\phi(1 : n) \rightarrow \phi^*(1 : n)) \]

\[ + A(\phi^*(1 : n) \rightarrow \phi) + 2(\hat{\phi}(1 : n) - \phi^*(1 : n))^T (R_{11} R_{12})(\phi^* - \phi), \quad (69) \]

where \( R, R_{11} \) and \( R_{12} \) were defined in (42) and (46). Thus

\[ EA(\hat{\phi}(1 : n) \rightarrow \phi) = EA(\hat{\phi}(1 : n) \rightarrow \phi^*(1 : n)) \]

\[ + A(\phi^*(1 : n) \rightarrow \phi) \]

\[ + 2\{E(\hat{\phi}(1 : n) - \phi^*(1 : n))^T \}

\[ \times (R_{11} R_{12})(\phi^* - \phi) \]

\[ = EA(\hat{\phi}(1 : n) \rightarrow \phi^*(1 : n)) \]

\[ + A(\phi^*(1 : n) \rightarrow \phi). \quad (70) \]

That is, the modelling error can also be decomposed into PE ME and OI ME for the AR model when the time domain error measurements are used.

Figure 1 illustrates the correlation between the errors (OI ME, expectation of PE ME, and modelling error) and the order. It can be seen that for a given AR(\( p \)) process and sample size, there is an optimal order which produces the minimum modelling error. When the sample size is small, the expected optimal order could be smaller than the actual order \( p \). In addition to the sample size, the characteristics of the process also have a significant influence on the expected optimal order.

The above results are acquired based on AR processes and time domain modelling error measurements. When the ARMA processes and MSD are used, the above analytic results will not hold exactly. For example, for ARMA processes, nonlinear estimation algorithms must be used. No analytic form can be provided for the covariance matrix of the parameter estimation errors \( \epsilon \) using the data before the iteration is completed. Theoretically, analytic solutions could be acquired for OI ME. However, in order to minimize the MSD, a complex frequency function must be integrated to generate an analytic formula. Then the integral is minimized with respect to the model parameters. This seems a complex issue. Consequently, in this paper a strict theor-
Assumethat"etrical analysis will not be performed for the ARMA models using the MSD. It is expected that the MSD measurements of PE ME and OI ME of ARMA models will tend to increase and decrease, respectively, as the order increases. Thus, the optimal orders can be acquired by examining the changes in PE ME and OI ME.

5. Model spectral distance criterion
Assume that \( ARMA(n + \Delta n, m + \Delta m) \), \( \Delta n \geq 0, \Delta m \geq 0 \), is a higher order ARMA model than \( ARMA(n, m) \). The mission of the order selection is to determine whether or not the order increase from \((n, m)\) to \((n + \Delta n, m + \Delta m)\), \(\Delta n \geq 0, \Delta m \geq 0\), is accepted. The proposed criterion selects the orders which can produce a lower MSD than the actual model. This criterion can be described as follows.

If \( \Delta_p < \Delta_o \) where

\[
\Delta_p := ED_{\hat{H}}^2(H(n + \Delta n, m + \Delta m))
\]

\[
\rightarrow \hat{H}^*(n + \Delta n, m + \Delta m)
\]

\[
\rightarrow \hat{H}^*(n, m)
\]

\[
\Delta_o := D^2(\hat{H}^*(n, m) \rightarrow H(p, q)) - D^2(\hat{H}^*(n + \Delta n, m + \Delta m) \rightarrow H(p, q))
\]

accept the order increase. Otherwise, the order increase is not accepted.

The above order selection criterion for ARMA models is referred to as the model spectral distance criterion (MSDC). There \( \Delta_p \) and \( \Delta_o \) specify the possible decrease and increase of OI ME and PE ME, respectively, when the order changes from \((n, m)\) to \((n + \Delta n, m + \Delta m)\), \(\Delta n \geq 0, \Delta m \geq 0\). Because of the strict mathematical decomposition (33), theoretically the better model is selected by the MSDC. However, in order to implement the proposed MSDC, \( \Delta_p \) and \( \Delta_o \) must be estimated. Thus, in addition to the estimation of \( \Delta_p \) and \( \Delta_o \), the analysis of the resultant error is also an important issue.

6. Implementation and error analysis
6.1. Estimation of PE ME for \( \Delta_p \)
In order to express the modelling error as the sum of PE ME and OI ME, PE ME has been measured using

\[
ED_{\hat{H}}^2(H \rightarrow \hat{H}^*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \hat{H}(\omega) - \hat{H}^*(\omega) \right|^2 d\omega.
\]

(71)

However, this equation cannot be directly used to calculate PE ME because both \( H \) and \( \hat{H}^* \) are unknown during practical modelling. An approximation must be made to estimate the PE ME. It can be shown that \( \Delta_o \) is much larger than \( \Delta_p \) when the order is lower. In this case, precise estimation of \( \Delta_p \) is not mandatory for this application. However, when the order becomes sufficient, \( \Delta_o \) must be precisely estimated. Thus, assume that \( D_{\hat{H}}^2(H \rightarrow \hat{H}^*) \) is an estimation of \( D_{\hat{H}}^2(H \rightarrow \hat{H}^*) \), the following conditions must be satisfied:

\[
\lim_{\parallel \Delta_p \parallel \rightarrow q} D_{\hat{H}}^2(H \rightarrow \hat{H}^*) = D_{\hat{H}}^2(H \rightarrow \hat{H}^*), \quad \forall n < p \text{ or } m < q,
\]

\[
D_{\hat{H}}^2(H \rightarrow \hat{H}^*) = D_{\hat{H}}^2(H \rightarrow \hat{H}^*), \quad \forall n \geq p, m \geq q.
\]

(72)

Consider

\[
D_{\hat{H}}^2(H \rightarrow \hat{H}^*) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \hat{H}(\omega) - \hat{H}^*(\omega) \right|^2 d\omega
\]

(73)

Because

\[
\frac{\partial \hat{H}^* - H}{\partial p} \leq 0,
\]

\[
\frac{\partial \hat{H}^* - H}{\partial q} \leq 0
\]

(74)

\[
\hat{H}^* = H, \quad \forall n \geq p, m \geq q,
\]
the following can be shown:
\[
\begin{aligned}
\varnothing[D_H^2(\hat{H} \to \hat{H}^*) - D_H^2(\hat{H} \to \hat{H}^*)] &= 0, \\
\varnothing[D_H^2(\hat{H} \to \hat{H}^*) - D_H^2(\hat{H} \to \hat{H}^*)] &= 0, \\
D_H^2(\hat{H} \to \hat{H}^*) &= D_H^2(\hat{H} \to \hat{H}^*), \quad \forall n \geq p, m \geq q.
\end{aligned}
\]
(75)

Thus
\[
D_H^2(\hat{H} \to \hat{H}^*) D_H^2(\hat{H} \to \hat{H}^*),
\]
(76)
satisfies (72) and is therefore an acceptable approximation for the order selection.

In addition to the satisfaction of (72) and zero errors when \( n \geq p \) and \( m \geq q \), it can also be shown that the approximation error of (76) is small when \( n < p \) and \( m < q \). In fact, both \( D_H^2(\hat{H} \to \hat{H}^*) = D_H^2(\hat{H} \to \hat{H}^*) \) and \( D_H^2(\hat{H} \to \hat{H}^*) \) are measurements of the difference between \( \hat{H} \) and \( \hat{H}^* \). The only distinction is the weight used for normalizing the error scale along the frequency axis in order to form the relative error. Because \( \hat{H}^* \) is the optimal frequency-domain approximation of \( H \) for the used order \((n,m)\), the error of the approximation
\[
\int_{-\pi}^{\pi} \frac{1}{|H|^2} d\omega = \int_{-\pi}^{\pi} \frac{1}{|\hat{H}^*|^2} d\omega,
\]
(77)
must be very small. Thus, although a strict mathematical proof is difficult to derive, it is believed that the error of the approximation (76) is not significant when \( n < p \) and \( m < q \).

Assume that
\[
\frac{\hat{\theta}(B)}{\theta(B)} = \frac{1 - \sum_{j=1}^{m} \hat{\theta}_j B^j}{1 - \sum_{j=1}^{n} \theta_j B^j} = \prod_{j=1}^{m} (1 - \hat{\theta}_j B),
\]
(78)
is the estimated ARMA\((n,m)\). The optimal ARMA\((n,m)\) approximation is
\[
\frac{\hat{\phi}^*(B)}{\phi^*(B)} = \frac{1 - \sum_{j=1}^{m} \hat{\phi}_j B^j}{1 - \sum_{j=1}^{n} \phi_j B^j} = \prod_{j=1}^{m} (1 - \hat{\phi}_j B).
\]
(79)

Thus
\[
\hat{H}(\omega) = (\hat{a}(\omega) + i\hat{b}(\omega)) / \sqrt{2\pi a(\omega) + i\hat{b}(\omega)),}
\]
\[
\hat{H}^*(\omega) = (\hat{a}^*(\omega) + i\hat{b}^*(\omega)) / \sqrt{2\pi a^*(\omega) + i\hat{b}^*(\omega)),
\]
(80)

\[
\hat{a}(\omega) = 1 - \sum_{j=1}^{n} \hat{\phi}_j \cos(j\omega),
\]
\[
\hat{a}^*(\omega) = 1 - \sum_{j=1}^{n} \hat{\phi}_j \cos(j\omega),
\]
\[
\hat{b}(\omega) = \sum_{j=1}^{n} \hat{\phi}_j \sin(j\omega),
\]
\[
\hat{b}^*(\omega) = \sum_{j=1}^{n} \hat{\phi}_j \sin(j\omega),
\]
\[
\hat{u}(\omega) = 1 - \sum_{j=1}^{m} \hat{\theta}_j \cos(j\omega),
\]
\[
\hat{u}^*(\omega) = 1 - \sum_{j=1}^{m} \hat{\theta}_j \cos(j\omega),
\]
\[
\hat{v}(\omega) = \sum_{j=1}^{m} \hat{\theta}_j \sin(j\omega),
\]
\[
\hat{v}^*(\omega) = \sum_{j=1}^{m} \hat{\theta}_j \sin(j\omega).
\]
(81)

Define
\[
\Delta a(\omega) := \hat{a}(\omega) - a(\omega) = -\sum_{j=1}^{n} \Delta \phi_j \cos(j\omega),
\]
\[
\Delta b(\omega) := \hat{b}(\omega) - b(\omega) = \sum_{j=1}^{n} \Delta \phi_j \sin(j\omega),
\]
\[
\Delta u(\omega) := \hat{u}(\omega) - u(\omega) = -\sum_{j=1}^{n} \Delta \theta_j \cos(j\omega),
\]
\[
\Delta v(\omega) := \hat{v}(\omega) - v(\omega) = \sum_{j=1}^{n} \Delta \theta_j \sin(j\omega).
\]
(82)

It can be shown that (see (24) for the definition)
\[
\Delta a(e^{-i\omega}) = \Delta a(\omega) + i\Delta b(\omega),
\]
\[
\Delta \theta(e^{-i\omega}) = \Delta a(\omega) + \imath \Delta b(\omega).
\]
(83)

Denote
\[
|\hat{d}|^2 = \left| \frac{\Delta a(e^{-i\omega})}{a(e^{-i\omega})} - \frac{\Delta a(e^{-i\omega})}{a(e^{-i\omega})} \right|^2,
\]
\[
\delta = \frac{|a(e^{-i\omega})|^2}{|\hat{a}(e^{-i\omega})|^2} - 1 = (\hat{u} + \hat{u}^*) \Delta u + (\hat{v} + \hat{v}^*) \Delta v.
\]
(84)

Because the parameter estimates \( \hat{\phi}_j \) (\( 0 \leq j \leq n \)) and \( \hat{\theta}_j \) (\( j \leq m \)) are known, \( \hat{u} \) and \( \hat{v} \) are known. In addition, for the given ARMA\((p,q)\) and \((n,m)\), \( \hat{u}^* \) and \( \hat{v}^* \) are certain and deterministic. The only random variables in \( \delta \) are \( \Delta u \) and \( \Delta v \). Therefore

Selecting optimal ARMA order

231
\[ E\delta = \frac{(\hat{u} + \hat{u}^*)E(\Delta u) + (\hat{v} + \hat{v}^*)E(\Delta v)}{(\hat{u}^*)^2 + (\hat{v}^*)^2}. \] (85)

Assume that the used parameter estimation algorithm is unbiased so that the expectation of the parameter estimation error is zero. Thus, \( E(\Delta u) = E(\Delta v) = 0 \), and hence
\[ E\delta = 0. \] (86)

Since
\[ |\hat{d}|^2 + \delta |\hat{d}|^2 = \frac{\hat{\theta}^2}{\hat{\theta}^*} \frac{|\Delta \theta - \Delta \phi|}{\theta^*} = \frac{|\hat{\theta}(\hat{\theta} - \hat{\theta}^*) - \hat{\phi}(\hat{\phi} - \hat{\phi}^*)|^2}{\theta^*} \]
\[ = \frac{|\hat{\theta}(\hat{\phi} - \hat{\phi}^*)|^2}{\hat{\theta}^*} = |\hat{H} - \hat{H}^*|^2 = |d|^2, \] (87)
the relative error of the approximation equation \( \hat{d} \approx \hat{d}^* \) is
\[ \left| \frac{|d|^2 - |\hat{d}|^2}{|\hat{d}|^2} \right| = \delta. \] (88)

Because \( E\delta = 0 \), \( |\hat{d}|^2 \) is an unbiased approximation of \( |d|^2 \). In addition, in our order selection procedure, the order \((n, m)\) increases until it is sufficient. As the parameter estimation errors are roughly proportional to the parameter number to sample size ratio, the parameter estimation errors of lower order models encountered in the order selection procedure will be small so that \( |\hat{\theta}/\hat{\phi}| \approx 1 \). Thus, the amplitude of \( \delta \) is small (Fig. 2).

Hence, the following equation can be used:
\[ D^2(\hat{H} \rightarrow \hat{H}^*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} E \left| \frac{\Delta \theta}{\theta} - \frac{\Delta \phi}{\phi} \right|^2 \, d\phi \] (89)

because one can show that the relative error
\[ \Delta = \frac{\int_{-\pi}^{\pi} |\delta| |d|^2 \, d\phi}{\int_{-\pi}^{\pi} |\hat{d}|^2 \, d\phi}, \] (90)
is small from the above discussion of \( \delta \) and \( E\delta = 0 \). In the example shown in Fig. 2, \( \Delta = 3.5\% \). More examples of the relative computation error \( \Delta \) are given in Table 1.

It can be seen that the accuracy of the computation equation (89) is sufficient. Thus
\[ ED(\hat{H} \rightarrow \hat{H}^*) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} E \left| \frac{\Delta \theta}{\theta} - \frac{\Delta \phi}{\phi} \right|^2 \, d\phi \]
\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} E \left( \frac{(\hat{\Delta}u + \hat{\Delta}v)}{\hat{\theta}^2 + \hat{\phi}^2} - \frac{\Delta u + i\Delta b}{\hat{\theta} + i\hat{\phi}} \right)^2 \, d\phi \]
\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} E \left\{ \left( \frac{(\hat{\Delta}u + \hat{\Delta}v)}{\hat{\theta}^2 + \hat{\phi}^2} - \frac{\Delta u + i\Delta b}{\hat{\theta} + i\hat{\phi}} \right)^2 + \left( \frac{\hat{\Delta}v - \nu \Delta u}{\hat{\theta}^2 + \hat{\phi}^2} \right)^2 \right\} \, d\phi. \] (91)

It can be seen that the function to integrate and perform the mathematical expectation is a quadratic in the random variables \((\Delta u, \Delta v, \Delta a, \Delta b)\). Thus, if we denote

<p>| Table 1. Relative error ( \Delta ) in calculating PE ME |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>( \phi )</th>
<th>( \theta )</th>
<th>( N )</th>
<th>( \hat{\phi} )</th>
<th>( \hat{\theta} )</th>
<th>( \Delta(%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0, -0.09</td>
<td>0, -0.81</td>
<td>50</td>
<td>0.98, -0.015</td>
<td>-0.073, -0.773</td>
<td>4.09</td>
</tr>
<tr>
<td>100</td>
<td>1.082, -0.105</td>
<td>-0.024, -0.788</td>
<td>3.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.975, -0.027</td>
<td>-0.017, -0.813</td>
<td>2.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9, -0.18</td>
<td>-0.7</td>
<td>50</td>
<td>1.65, -0.079</td>
<td>-0.87</td>
<td>6.70</td>
</tr>
<tr>
<td>100</td>
<td>1.187, -0.234</td>
<td>-0.593</td>
<td>12.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.901, -0.132</td>
<td>-0.702</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ \Delta x_1 = \Delta a, \quad \Delta x_2 = \Delta b, \quad \Delta x_3 = \Delta u, \quad \Delta x_4 = \Delta v, \]

(92)

then

\[ \frac{E D^2 \hat{H}(\hat{H} \rightarrow \hat{H}^*)}{2\pi} \approx \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{4} \lambda_{ij}(\omega) E(\Delta x_i \Delta x_j) d\omega, \]

(93)

where

\[ E(\Delta x_1 \Delta x_2) = - \sum_{i=1}^{n} \sum_{j=1}^{4} \frac{E(\Delta t_i \Delta t_j) \cos(i\omega) \sin(j\omega))}{(i-j)^2} \]

\[ E(\Delta x_2 \Delta x_3) = - \sum_{i=1}^{n} \sum_{j=1}^{4} \frac{E(\Delta t_i \Delta t_j) \sin(i\omega) \sin(j\omega))}{(i-j)^2} \]

(94)

and \( \lambda_{ij}(\omega), \ i = 1, \ldots, 4, \ j = 1, \ldots, 4, \) can be calculated using \( \hat{a}(\omega), \hat{b}(\omega), \hat{u}(\omega) \) and \( \hat{v}(\omega); \) and \( E(\Delta t_i \Delta t_j), E(\Delta t_i \Delta t_j), E(\Delta t_i \Delta t_j) \) are the elements of the covariance matrix of the parameter estimation errors which are available from either the Least Squares or Maximum Likelihood estimate (Box and Jenkins 1976). Thus, PE ME can be calculated.

6.2. Estimation of \( \Delta_o \)

\( \Delta_o \) is used to measure the possible accuracy improvement when the order increases. The direct calculation of this based on the definition is difficult, because both \( H \) and \( \hat{H}^* \) are unknown during practical modelling. However, it can be seen that the possible accuracy improvement when the order increases can be evaluated using the difference between ARMA \((n,m)\) and ARMA \((n+\Delta n,m+\Delta m)\). Thus, the following has been used to estimate \( \Delta_o \):

\[ \Delta_o \approx D^2(\hat{H}(n,m) - \hat{H}(n+\Delta n,m+\Delta m)). \]

(95)

In the following, the order selection error caused by the estimated \( \Delta_o \) and \( \Delta_p \) will be analysed.

6.3. Error analysis

Because the error decomposition (33) is strict, the better model of two given estimated models is selected by the MSDC so that the selected order is optimal in the sense of minimizing the MSD. However, because of the lack of knowledge of \( H \) and \( H^* \), approximations must be made to estimate \( \Delta_o \) and \( \Delta_p \). Assume that the limits of the relative errors of these estimates are \( \delta_o \) and \( \delta_p \); that is

\[ (\Delta_o - \hat{\Delta}_o)/\Delta_o < \delta_o, \]

\[ (\Delta_p - \hat{\Delta}_p)/\Delta_p < \delta_p. \]

(96)

Assume that the orders of the two given models for the \( k \) th order selection are \((n(k-1), m(k-1))\) and \((n(k), m(k))\), respectively. The order increase will be denied if the following is satisfied:

\[ \hat{\Delta}_o(k) < \hat{\Delta}_p(k). \]

(97)

The maximum \( \Delta_o(k) \) which could cause an order increase denial is

\[ \max(\Delta_o(k)) = \frac{1+\delta}{1-\delta} \Delta_p(k) := (1+\lambda(\delta_o,\delta_p))\Delta_p(k), \]

(98)

where

\[ \lambda(\delta_o,\delta_p) := \frac{\delta_o + \delta_p}{1 - \delta_o}. \]

(99)

It is known that \( \Delta_o \) is the decrease in the OI ME and \( \Delta_p \) is the increase in the PE ME. Thus, when the increase in the parameter number \( n(k) - n(k-1) + m(k-1) \) is given, \( \Delta_p \) is roughly maintained at the same level as the order increases. However, \( \Delta_o \) undergoes a significant change as the order increases. When the order is lower, \( \Delta_o \) is much larger than \( \Delta_p \). As the order increases, \( \Delta_o \) decreases. Once \( \Delta_o \) drops to the level of \( \Delta_p \), the increase in the order becomes less beneficial in the sense of decreasing the MSD. If (97) is satisfied, the order increase will not be accepted.

Ideally, the condition for denying the order increase is

\[ \Delta_o(k) < \Delta_p(k). \]

(100)

The maximum \( \Delta_o(k) \) which can cause the order increase denial should be

\[ \max(\Delta_o(k)) = \Delta_p(k). \]

(101)

However, because of the estimation error, the actual maximum \( \Delta_o(k) \) is \( \Delta_p(k) + \lambda(\delta_o,\delta_p)\Delta_p(k) \). The possible accuracy loss is therefore \( \lambda(\delta_o,\delta_p)\Delta_p(k) \). That is, due to the use of the estimated \( \Delta_o \) and \( \Delta_p \), the order increase could be denied even when the order increase can still decrease the modelling error. However, the increase of the modelling error caused by the incorrect order selection is bounded by

\[ \lambda(\delta_o,\delta_p)\Delta_p(k). \]

(102)

Here \( \Delta_p(k) \) is the increase in PE ME when the order increases from \( (n(k-1), m(k-1)) \) to \((n(k), m(k)) \). \((n(k), m(k)) \) can be selected so that \( n(k) - n(k-1) + m(k) - m(k-1) = 1 \).

Thus, \( \Delta_p(k) \) is the increase in PE ME when the number of parameters increases by 1. This is the resolution of the parameter estimation. Hence, the error bound in (102) is in the resolution level of parameter estimation. The ratio \( \lambda \) depends on the maximum relative errors of the estimated \( \Delta_o \) and \( \Delta_p \). If \( \delta_p = 20\% \) and \( \delta_o = 40\% \), the ratio
When there are two possible ranges of $\lambda = 1$. In this case, the error bound is given by the resolution of the parameter estimation. Because this resolution is the best accuracy which can be expected from the modelling, the error bound is small.

The error bound (102) is only the maximum possible error caused by the order selection. In the practical modelling, the change in $\Delta_o$ is discrete. Assume that

$$\begin{cases} 
\Delta_o(k - 1) > (1 + \lambda)\Delta_p(k - 1), \\
\Delta_o(k) < (1 + \lambda)\Delta_p(k).
\end{cases} \tag{103}$$

There are two possible ranges of $\Delta_o(k)$:

1. $\Delta_p(k) < \Delta_o(k) < (1 + \lambda)\Delta_p(k)$,
2. $\Delta_o(k) \leq \Delta_p(k)$.

When $\Delta_o(k)$ falls into the first range, the order selection is incorrect and the resultant error is bounded by (102).

Because of the small error bound, the difference between the given two models for the order selection is small. The selected model can be regarded as nearly optimal. Otherwise, $\Delta_o(k)$ falls into the second range. The order selection is correct and the selected model is optimal.

7. Numerical examples

Basically, the order selection criterion focuses on the comparison between two estimated models. A detailed discussion of how the order change trials should be performed is beyond the scope of this paper. The modelling steps used in our simulations are shown in the flow chart of Fig. 3. In the figure, $(n(0), m(0))$ is the initial order. ARMA estimation is performed by using a standard nonlinear least squares subroutine. The order increase $\Delta n > 0$ ($\Delta m = 0$) or $\Delta m > 0$ ($\Delta n = 0$) can be selected between 1 and $M_{\text{step}}$, where $M_{\text{step}}$ is the maximum allowable order increase step, which is selected to be two because of the possible conjugate roots.

The F-test (Pandit and Wu 1983) has been widely used in selecting ARMA orders. However, the results depend on the confidence factor. For the purpose of comparison, an objective criterion may be more appropriate. AIC is a well-known objective criterion (Fuchs 1987). Although studies have been performed to investigate its inconsistency, it appears that no better alternatives have been widely accepted. In this section, comparisons will be made between the present criterion and AIC. The sample size varies from 50 to 500 in order to cover a wide range of samples. A variety of processes listed in Table 2 with different spectrum characteristics are considered. The white driving noise $\{\varepsilon_t\}$ is assumed to obey $N(0, 1)$ in all the examples. The parameters are estimated using a nonlinear least squares routine, using a standard program NLSE or STAT/LIBRARY, IMSL. The modelling error is measured in the terminology of $D^2(\hat{H} \rightarrow H)$, where $H$ is the actual model and $\hat{H}$ is the estimated model.

**Table 2. Processes for order determination simulations**

<table>
<thead>
<tr>
<th>Group</th>
<th>No.</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$y_t + 0.9y_{t-1} + 0.81y_{t-2} + 0.729y_{t-3} = \varepsilon_t - 0.9\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$y_t + 0.9y_{t-1} + 0.81y_{t-2} + 0.729y_{t-3} = \varepsilon_t - 0.7\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$y_t + 0.9y_{t-1} + 0.81y_{t-2} + 0.729y_{t-3} = \varepsilon_t - 0.5\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$y_t - 0.7y_{t-1} - 0.1y_{t-2} = \varepsilon_t + 0.5\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$y_t - 0.7y_{t-1} - 0.2y_{t-2} = \varepsilon_t + 0.5\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$y_t - 10y_{t-1} - 0.09y_{t-2} = \varepsilon_t + 0.81\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$y_t - 0.95y_{t-1} - 0.45y_{t-2} = \varepsilon_t + 0.81\varepsilon_{t-1}$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$y_t - 0.65y_{t-1} + 0.81y_{t-2} - 0.526y_{t-3} = \varepsilon_t - 0.6\varepsilon_{t-1}$</td>
</tr>
</tbody>
</table>
All the methods used in the simulations are defined in Table 2.

Example 1: Process group 1: This example considers small samples, the sample size is \(N = 50\). In this example, complex poles are encountered.

In the case of process 1–1, the optimal model which produces the minimum spectral estimation error is ARMA(4, 1). The MSDC selected the optimal model, while the AIC selected ARMA(3, 1) which has a larger error. The corresponding spectral estimates are plotted in Fig. 4. From these plots it can also be seen that the model selected by the present criterion produces a better spectral estimate than the model selected by the AIC. The order selection procedure is listed in Table 3.

For process 1–2, the optimal model is the ARMA(3, 1) with \(D^2(\hat{H} \rightarrow H) = 0.2571\). The proposed criterion selected the optimal model. However, the ARMA(3, 0) with \(D^2(\hat{H} \rightarrow H) = 0.7376\) has been selected by the AIC criterion. For process 1–3, both the present the AIC criteria selected the optimal model, ARMA(4, 0). See Table 3 for the details of order selection.

Example 2: Process Group 2: In this example, small samples, \(N = 50\), are considered again for a simple spectrum configuration.

In the case of process 2–1, the optimal model is ARMA(3, 1). However, the AIC selected ARMA(3, 0) with a larger error, while the MSDC selected the optimal model ARMA(3, 1). The estimated power spectral densities are depicted in Fig. 5. Once again, a much better estimate has been obtained using the new criterion. In the case of process 2–2, the optimal model, ARMA(1, 1), was selected by the new criterion and the AIC criterion. See Table 4 for order selection details.

Example 3: Process Group 3: In this example, the sample sizes are 100. Complex zeros are encountered. The spectrums are characterized by a valley.

First consider process 3–1. In this case, the optimal model is ARMA(3, 2) with \(D^2(\hat{H} \rightarrow H) = 0.1085\). The model selection is listed in Table 3.

Table 3. Order determination of process group 1

<table>
<thead>
<tr>
<th>No.</th>
<th>Transfer trial</th>
<th>MSDC</th>
<th>Selection comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–1</td>
<td>(3, 0) to (3, 1)</td>
<td>(3, 1) to (4, 1)</td>
<td>Order (3, 0) (3, 1) (4, 1)</td>
</tr>
<tr>
<td></td>
<td>Estimated (d_e)</td>
<td>0.0224</td>
<td>–0.0007 AIC 11.67 5.86</td>
</tr>
<tr>
<td></td>
<td>Estimated (d_{oa})</td>
<td>0.1568</td>
<td>0.0158 (\rho^2(\hat{H} \rightarrow H)) 3.396 0.597 0.449</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
<td>Yes Selection A B, M</td>
</tr>
</tbody>
</table>

| 1–2 | (3, 0) to (3, 1) | (3, 1) to (4, 1) | Order (3, 0) (3, 1) (4, 1) |
|     | Estimated \(d_e\) | 0.0732 | 0.0324 AIC –1.331 –0.583 1.904 |
|     | Estimated \(d_{oa}\) | 0.1023 | 0.0160 \(\rho^2(\hat{H} \rightarrow H)\) 0.738 0.257 0.305 |
|     | Trial acceptance | Yes | No Selection A B, M |

| 1–3 | (2, 0) to (3, 0) | (3, 0) to (4, 0) | Order (2, 0) (3, 0) (4, 0) |
|     | Estimated \(d_e\) | 0.0278 | 0.0222 AIC 29.20 7.200 –3.372 |
|     | Estimated \(d_{oa}\) | 0.1095 | 0.0406 \(\rho^2(\hat{H} \rightarrow H)\) 1.254 0.397 0.262 |
|     | Trial acceptance | Yes | Yes Selection B, A, M |

Only primary steps of order selection are listed in the table. A: selected by AIC. B: best model. M: selected by MSDC.
model selected by the MSDC was ARMA(2, 2) with $D^2(\hat{H} \rightarrow H) = 0.1270$, while the model selected by the AIC criterion is ARMA(1, 2) with $D^2(\hat{H} \rightarrow H) = 0.2569$. It is seen that the optimal model has not been selected by the MSDC. However, the difference in accuracy is very slight. On the other hand, this difference between the optimal model and the AIC selected model is much larger. The resulting spectral estimates are given in Fig. 6.

In the case of process 3–2, the optimal model is ARMA(3, 2) associated with $D^2(\hat{H} \rightarrow H) = 0.1700$. This optimal model has been selected by the MSDC. The model selected by the AIC criterion is ARMA(1, 2) with $D^2(\hat{H} \rightarrow H) = 0.3168$.

The order selection procedures can be seen in Table 5.

Example 4: Process Group 4: This example considers larger samples, $N = 500$. For the process 4–1, the optimal model is ARMA(2, 1) with $D^2(\hat{H} \rightarrow H) = 0.0113$. This model has been selected by the MSDC and AIC. Refer to Table 6 for order selection details.

Table 4. Order determination of process group 2

<table>
<thead>
<tr>
<th>No.</th>
<th>MSDC</th>
<th>Selection comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–1</td>
<td>Transfer trial</td>
<td>(1, 1) to (2, 1)</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0207</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.6406</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
</tr>
<tr>
<td>2–2</td>
<td>Transfer trial</td>
<td>(1, 0) to (1, 1)</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0183</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0396</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Only primary steps of order selection are listed in the table. A: selected by AIC. B: best model. M: selected by MSDC.

Table 5. Order determination of process group 3

<table>
<thead>
<tr>
<th>No.</th>
<th>MSDC</th>
<th>Selection comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>3–1</td>
<td>Transfer trial</td>
<td>(1, 2) to (2, 2)</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>-0.0049</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
</tr>
<tr>
<td>3–2</td>
<td>Transfer trial</td>
<td>(1, 2) to (2, 2)</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>-0.0019</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Only primary steps of order selection are listed in the table. A: selected by AIC. B: best model. M: selected by MSDC.

Table 6. Order determination of process group 4

<table>
<thead>
<tr>
<th>No.</th>
<th>MSDC</th>
<th>Selection comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>4–1</td>
<td>Transfer trial</td>
<td>(2, 0) to (2, 1)</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>Estimated $A_2$</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>Trial acceptance</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Only primary steps of order selection are listed in the table. A: selected by AIC. B: best model. M: selected by MSDC.
It can be seen that the optimal models have frequently been selected by the MSDC. In the case of failure to select the optimal models, the differences between the selected models and the optimal models are very small. These models are regarded as nearly optimal.

8. Conclusions
The modelling error can be decomposed into OI ME and PE ME. The proposed MSDC selects the order based on a comparison between the PE ME increase and OI ME decrease as the order increases. The maximum modelling error caused by the order selection is bounded by the resolution level of parameter estimation. The orders selected by the MSDC are optimal or nearly optimal. Simulations show that the MSDC does select the optimal or nearly optimal orders.

Acknowledgments
The authors are grateful to Professor S. A. Billings for his helpful suggestions, and to the anonymous reviewers for their comments and constructive suggestions with regard to this paper.

References