

Research Article

Finite Sample FPE and AIC Criteria for Autoregressive Model Order Selection Using Same-Realization Predictions

Shapoor Khorshidi and Mahmood Karimi

Department of Electrical Engineering, Shiraz University, Shiraz 71348-51151, Iran

Correspondence should be addressed to Mahmood Karimi, karimi@shirazu.ac.ir

Received 26 May 2009; Revised 2 November 2009; Accepted 31 December 2009

Recommended by Abdelak Zoubir

A new theoretical approximation for expectation of the prediction error is derived using the same-realization predictions. This approximation is derived for the case that the Least-Squares-Forward (LSF) method (the covariance method) is used for estimating the parameters of the autoregressive (AR) model. This result is used for obtaining modified versions of the AR order selection criteria FPE and AIC in the finite sample case. The performance of these modified criteria is compared with other same-realization AR order selection criteria using simulated data. The results of this comparison show that the proposed criteria have better performance.

Copyright © 2009 S. Khorshidi and M. Karimi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Let a real AR process be given by

$$x(n) = a_1x(n-1) + \dots + a_px(n-p) + \varepsilon(n), \quad (1)$$

where p is the order of the process, a_1, \dots, a_p are the real coefficients of the process, and $\varepsilon(n)$ is an independent and identically distributed (i.i.d.) random process with zero mean and variance σ_ε^2 . The process $\varepsilon(n)$ is the input process to the AR model. It is assumed that $x(\cdot)$ is mean and covariance ergodic; so the poles of the AR model are inside the unit circle.

Suppose that $x(1), \dots, x(N)$ are consecutive samples of a sample function of the AR process $x(\cdot)$ given by (1). In addition, suppose that an arbitrary nonnegative integer q is considered as the true process order. The usual estimator for a nonrandom set of parameters is the maximum likelihood estimator (MLE) [1]. However, the exact solution of the MLE for the parameters of an AR(q) process is difficult to obtain [1]. If $N \gg q$, each of the AR estimation methods Least-Squares-Forward (LSF), Least-Squares-Forward-Backward (LSFB), and Burg and Yule-Walker [1–3] is an approximation

of the MLE. In this paper, we use the LSF method to estimate a model as

$$x(n) = \hat{a}_1x(n-1) + \dots + \hat{a}_qx(n-q) + \hat{\varepsilon}(n), \quad (2)$$

where \hat{a}_i 's are the LSF estimates of the AR parameters. Residual variance, denoted by $S^2(q)$, is a measure of the fitness of the above model to the data that have been used for estimating the parameters. It is defined as

$$S^2(q) = \frac{1}{N-q} \sum_{n=q+1}^N \left[x(n) - \sum_{i=1}^q \hat{a}_i x(n-i) \right]^2. \quad (3)$$

In the literature, two different kinds of predictions under model (1) are considered. For independent-realization predictions, the aim is to predict the future of another independent series which has exactly the same probabilistic structure as the observed one. One special feature of this type of prediction is that its mathematical analysis is relatively easy. The independent-realization prediction error is defined as follows:

$$PE(q) = E_y \left\{ \left[y(n) - \sum_{i=1}^q \hat{a}_i y(n-i) \right]^2 \mid x(1), \dots, x(N) \right\}, \quad (4)$$

where $y(\cdot)$ is another independent series which has exactly the same probabilistic structure as $x(\cdot)$, and $E_y\{\cdot\}$ is the expectation operator over this independent series. However, for the practitioner, the emphasis is usually placed on same-realization predictions, that is, on the prediction of $x(N+h)$, $h \geq 1$ given $x(1), \dots, x(N)$. The same-realization prediction error is defined as follows:

$$\text{PE}(q) = E \left\{ \left[x(N+h) - \sum_{i=1}^q \hat{a}_i x(N+h-i) \right]^2 \mid x(1), \dots, x(N) \right\}. \quad (5)$$

In (4) and (5), \hat{a}_i 's are the LSF estimates of a_i 's using $x(1), \dots, x(N)$. Note that in the expectation in (4) $y(n)$ is independent of the \hat{a}_i 's, but in the expectation in (5), $x(N+h)$ depends on the \hat{a}_i 's. This difference causes the prediction errors in (4) and (5) to be different. Almost all existing AR order selection criteria have been derived for independent-realization data. The independency is not a natural property for time series data, because when new observations of a time series become available, they are usually dependent on the previous data. So far, few time series model selection theories have been established without this unnatural assumption. In addition, the use of most of these model order selection criteria has not been justified for the same-realization case. Akaike information criterion (AIC) [4], final prediction error (FPE) criterion [5], Bayes information criterion (BIC) [6], minimum description length (MDL) criterion [7], Kullback information criterion (KIC) [8], corrected AIC (AIC_C) [9], and corrected KIC (KIC_C) [10] are examples of the criteria that have been derived for the large sample and independent-realization case. In [11, 12], Ing and Wei justified the use of FPE and AIC as AR order selection criteria in the large sample and same-realization case. Ing and Wei [11] presented a theoretical verification that AIC and FPE are asymptotically efficient (in the sense of the mean square prediction error) for same-realization predictions. When the underlying AR model is known to be stationary and of infinite order, Ing and Wei [11] showed that the values of the expectations of the squared prediction error for independent- and same-realization cases, with the order selected by FPE or AIC, have the same asymptotic expressions. Recently, Ing [12] removed the infinite-order assumption and verified the asymptotic efficiency of several information criteria, including FPE and AIC, for both finite- and infinite-order stationary AR models.

The order selection criteria obtained in the large sample and independent-realization case are not dependent on the method of estimation of the AR parameters. In other words, these criteria are identical for all AR parameter estimation methods. However, it is well known that, in the finite sample and independent-realization case, the performance of the order selection criteria depends on the parameter estimation method and it is necessary to derive different order selection criteria for each parameter estimation method [13–18].

In this paper, we derive a new estimate of the same-realization prediction error in the finite sample case and for the LSF parameter estimation method. We will use this new estimate to derive same-realization versions of FPE (FPEF) and AIC (AICF) in the finite sample case.

The remainder of this paper is organized as follows. In Section 2, a new theoretical approximation is derived for the expectation of same-realization prediction error in the case that the LSF method is used. Based on this approximation, the FPEF criterion is introduced. In Section 3, the AICF criterion is introduced. In Section 4, simulated data are used for comparing the performance of the proposed order selection criteria with that of the existing criteria. In Section 5, the conclusions of this paper are discussed.

2. Estimation of the Same-Realization Prediction Error

Suppose that we have N observation data of the AR model defined by (1) as $x(1), x(2), \dots, x(N)$. We define the following vectors and matrices for the case that the candidate order q is greater than or equal to the true order ($q \geq p$), and T stands for a transposed operation:

$$\boldsymbol{\beta}_A = [a_1, a_2, \dots, a_p, 0, \dots, 0]_{q \times 1}^T, \quad (6)$$

$$\mathbf{x}_i = [x(i), x(i+1), \dots, x(N-q+i-1)]^T, \quad (7)$$

$$\mathbf{x}(i) = [x(i), x(i-1), \dots, x(i-q+1)]^T, \quad (8)$$

$$\mathbf{X}_A = [\mathbf{x}_q \ \mathbf{x}_{q-1} \ \dots \ \mathbf{x}_1], \quad (9)$$

$$\boldsymbol{\varepsilon}_{q+1} = [\varepsilon(q+1), \dots, \varepsilon(N)]^T, \quad (10)$$

$$\mathbf{x} = [x(1), \dots, x(N)]. \quad (11)$$

It follows from (1), (6), (7), (9), and (10) that

$$\mathbf{x}_{q+1} = \mathbf{X}_A \boldsymbol{\beta}_A + \boldsymbol{\varepsilon}_{q+1}; \quad q \geq p. \quad (12)$$

We use the LSF method to obtain the least-squares estimate of $\boldsymbol{\beta}_A$ as follows:

$$\mathbf{b}_A = [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_q]^T = (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{x}_{q+1}. \quad (13)$$

The one-step same-realization prediction error is given by

$$\text{PE}(q) = E \left\{ [x(N+1) - \hat{x}_q(N+1)]^2 \mid \mathbf{x} \right\}, \quad (14)$$

where $\hat{x}_q(N+1)$ is the linear predictor of $x(N+1)$ given $x(1), \dots, x(N)$, that is,

$$\hat{x}_q(N+1) = \sum_{i=1}^q \hat{a}_i x(N+1-i) = \mathbf{x}^T(N) \mathbf{b}_A. \quad (15)$$

It follows from (1), (6), and (8) that

$$x(N+1) = \mathbf{x}^T(N) \boldsymbol{\beta}_A + \varepsilon(N+1). \quad (16)$$

Substituting (15) and (16) into (14), we can rewrite (14) as

$$PE(q) = E\left\{\left[\mathbf{x}^T(N)\boldsymbol{\beta}_A + \varepsilon(N+1) - \mathbf{x}^T(N)\mathbf{b}_A\right]^2 \mid \mathbf{x}\right\}. \quad (17)$$

Using (12) and (13), $\mathbf{x}^T(N)\mathbf{b}_A$ can be written as

$$\begin{aligned} \mathbf{x}^T(N)\mathbf{b}_A &= \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} \\ &= \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\mathbf{X}_A\boldsymbol{\beta}_A \\ &\quad + \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} \\ &= \mathbf{x}^T(N)\boldsymbol{\beta}_A + \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}; \quad q \geq p. \end{aligned} \quad (18)$$

Combining (17) and (18), we have

$$\begin{aligned} PE(q) &= E\left\{\left[\varepsilon(N+1) - \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2 \mid \mathbf{x}\right\} \\ &= E\left\{\left[\varepsilon(N+1)^2 \mid \mathbf{x}\right]\right\} \\ &\quad + E\left\{\left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2 \mid \mathbf{x}\right\} \\ &\quad - 2E\left\{\left[\varepsilon(N+1)\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} \mid \mathbf{x}\right]\right\}; \\ &\quad \quad \quad q \geq p. \end{aligned} \quad (19)$$

The value of the first term in the right-hand side of (19) is given as

$$E\left\{\left[\varepsilon(N+1)^2 \mid \mathbf{x}\right]\right\} = E\left\{\varepsilon(N+1)^2\right\} = \sigma_\varepsilon^2. \quad (20)$$

In (19), we know that $\varepsilon(N+1)$ is independent of the past samples of $x(1), x(2), \dots, x(N)$, and $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)$ for the AR process. In addition, the expression $\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}$ is a combination of $x(1), x(2), \dots, x(N)$, $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)$. So, $\varepsilon(N+1)$ is independent of $\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}$, and the third term of the right-hand side of (19) is equal to zero:

$$E\left\{\left[\varepsilon(N+1)\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} \mid \mathbf{x}\right]\right\} = 0. \quad (21)$$

It follows from (12) that the value of the second term of the right-hand side of (19) is equal to

$$\begin{aligned} \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} \\ = \mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T(\mathbf{x}_{q+1} - \mathbf{X}_A\boldsymbol{\beta}_A); \quad q \geq p. \end{aligned} \quad (22)$$

The expression $\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T(\mathbf{x}_{q+1} - \mathbf{X}_A\boldsymbol{\beta}_A)$ is a combination of $x(1), x(2), \dots, x(N)$. So, it follows from (22) that

the expectation of $\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}$ given \mathbf{x} is equal to $\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}$, that is,

$$\begin{aligned} E\left\{\left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2 \mid \mathbf{x}\right\} \\ = \left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2; \quad q \geq p. \end{aligned} \quad (23)$$

Therefore, using (20), (21), and (23), we can rewrite (19) as

$$PE(q) = \sigma_\varepsilon^2 + \left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2. \quad (24)$$

Taking the expectation of $PE(q)$ over the vector \mathbf{x} , it follows from (24) that

$$E_{\mathbf{x}}[PE(q)] = \sigma_\varepsilon^2 + E_{\mathbf{x}}\left\{\left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1}\right]^2\right\}. \quad (25)$$

It can be seen from (7)–(10) that

$$\begin{aligned} \mathbf{X}_A^T\boldsymbol{\varepsilon}_{q+1} &= \begin{bmatrix} \mathbf{x}_q & \mathbf{x}_{q-1} & \cdots & \mathbf{x}_1 \end{bmatrix}^T \boldsymbol{\varepsilon}_{q+1} \\ &= \begin{bmatrix} \mathbf{x}(q) & \mathbf{x}(q+1) & \cdots & \mathbf{x}(N-1) \end{bmatrix} \boldsymbol{\varepsilon}_{q+1} \\ &= \sum_{i=q}^{N-1} \mathbf{x}(i)\varepsilon(i+1). \end{aligned} \quad (26)$$

Substituting (26) into (25), we obtain

$$E_{\mathbf{x}}[PE(q)] = \sigma_\varepsilon^2 + E_{\mathbf{x}}\left\{\left[\mathbf{x}^T(N)\left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\sum_{i=q}^{N-1} \mathbf{x}(i)\varepsilon(i+1)\right]^2\right\}. \quad (27)$$

It is assumed that $x(\cdot)$ is a zero mean and covariance ergodic process. So, we define the covariance matrix \mathbf{R} and its estimate $\hat{\mathbf{R}}$ in the following way:

$$\mathbf{R} = E\left[\mathbf{x}(i)\mathbf{x}^T(i)\right], \quad (28)$$

$$\hat{\mathbf{R}} = \frac{\left(\mathbf{X}_A^T\mathbf{X}_A\right)}{(N-q)}. \quad (29)$$

Using (29) and (27), we have

$$\begin{aligned} E_{\mathbf{x}}[PE(q)] \\ = \sigma_\varepsilon^2 + \frac{1}{(N-q)^2} E_{\mathbf{x}}\left\{\left[\mathbf{x}^T(N)\hat{\mathbf{R}}^{-1}\sum_{i=q}^{N-1} \mathbf{x}(i)\varepsilon(i+1)\right]^2\right\}. \end{aligned} \quad (30)$$

To obtain a value for (30), we make the assumption that for sufficiently large $(N-q)$: $|\hat{\mathbf{R}}\mathbf{R}^{-1} - \mathbf{I}_q| < 1$, where $|\mathbf{A}|$ is the Euclidean norm of matrix \mathbf{A} . Under this assumption, it is shown in [19] that

$$\begin{aligned} \frac{1}{(N-q)^2} E_{\mathbf{x}}\left\{\left[\mathbf{x}^T(N)\hat{\mathbf{R}}^{-1}\sum_{i=q}^{N-1} \mathbf{x}(i)\varepsilon(i+1)\right]^2\right\} \\ = \frac{q}{(N-q)}\sigma_\varepsilon^2 + O\left[\frac{1}{(N-q)^{3/2}}\right]; \quad q \geq p, \end{aligned} \quad (31)$$

where $\beta_n = O(\alpha_n)$ means that for some positive number M , $|\beta_n| \leq M|\alpha_n|$ for every n .

Substituting (31) into (30), we get

$$E_x[\text{PE}(q)] = \sigma_\varepsilon^2 + \frac{q}{(N-q)}\sigma_\varepsilon^2 + O\left[\frac{1}{(N-q)^{3/2}}\right]; \quad q \geq p. \quad (32)$$

When $N-q$ tends to infinity, the third term in the right-hand side of (32) tends towards zero faster than the other terms. So, for large enough values of $N-q$

$$E_x[\text{PE}(q)] \approx \left[1 + \frac{q}{N-q}\right]\sigma_\varepsilon^2; \quad q \geq p. \quad (33)$$

It is shown in [16, 17] that

$$E_x[S^2(q)] \approx \left[1 - \frac{q}{N-q}\right]\sigma_\varepsilon^2; \quad q \geq p. \quad (34)$$

Combining (33) and (34), we obtain

$$E[\text{PE}(q)] \approx \left[\frac{1+q/(N-q)}{1-q/(N-q)}\right]E[S^2(q)]; \quad q \geq p. \quad (35)$$

The above relation can be used for estimating $\text{PE}(q)$ in the following way:

$$\widehat{\text{PE}}(q) = \left[\frac{1+q/(N-q)}{1-q/(N-q)}\right]S^2(q); \quad q \geq p. \quad (36)$$

It is reasonable to choose the integer q that minimizes this estimate of prediction error as the appropriate order for the AR process. So, we propose the finite sample FPE (FPEF) criterion as

$$\text{FPEF}(q) = \left[\frac{1+q/(N-q)}{1-q/(N-q)}\right]S^2(q). \quad (37)$$

The FPE criterion, which is an asymptotic estimate of the independent-realization prediction error, is defined in the following way [5]:

$$\text{FPE}(q) = \frac{N+q}{N-q}S^2(q) = \left[\frac{1+q/N}{1-q/N}\right]S^2(q). \quad (38)$$

As we mentioned earlier, the criterion defined by (38) is also used in the independent-realization case. When $N \gg q$, the FPE criterion given by (38) is an approximation of FPEF. In the asymptotic case $N \rightarrow \infty$ the criteria FPE and FPEF become identical.

3. AICF Criterion

We now give a mathematical derivation for finite sample AIC (AICF) criterion in the same-realization case starting from Kullback-Leibler (K-L) information. The K-L information is a measure of the distance between the true and the approximating pdfs for the data generated by the true pdf.

The K-L information for the approximating model $g(\mathbf{x}' | \boldsymbol{\theta})$ is given by [20]

$$I[f, g(\cdot | \boldsymbol{\theta})] = \int f(\mathbf{x}') \ln \left[\frac{f(\mathbf{x}')}{g(\mathbf{x}' | \boldsymbol{\theta})} \right] d\mathbf{x}', \quad (39)$$

where \mathbf{x}' is a vector of data generated by the true pdf f . We look for a unique value of parameter vector $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0$, that minimizes the K-L information. So, $\boldsymbol{\theta}_0$ is the solution to the following optimization problem [20]:

$$\min_{\boldsymbol{\theta} \in \Theta} \{I[f, g(\cdot | \boldsymbol{\theta})]\} = \int f(\mathbf{x}') \ln \left[\frac{f(\mathbf{x}')}{g(\mathbf{x}' | \boldsymbol{\theta}_0)} \right] d\mathbf{x}', \quad (40)$$

where Θ is the parameter space. In fact, we have to find the model order and the parameter values that minimize (39).

The expression in the right-hand side of (39) is not dependent on the observed data and f is unknown. So, we cannot compute (39) for different values of $\boldsymbol{\theta}$ and determine the value $\boldsymbol{\theta}_0$ that minimizes the K-L information. Therefore, we rewrite the K-L information in the following way:

$$\begin{aligned} I[f, g(\cdot | \boldsymbol{\theta})] &= \int f(\mathbf{x}') \ln[f(\mathbf{x}')] d\mathbf{x}' - \int f(\mathbf{x}') \ln[g(\mathbf{x}' | \boldsymbol{\theta})] d\mathbf{x}' \\ &= \text{constant} - E_{\mathbf{x}'}[\ln[g(\mathbf{x}' | \boldsymbol{\theta})]]. \end{aligned} \quad (41)$$

The first term in the right-hand side of (41) is not dependent on $\boldsymbol{\theta}$. So, instead of minimizing (41), we can minimize the following term:

$$-E_{\mathbf{x}'}[\ln[g(\mathbf{x}' | \boldsymbol{\theta})]]. \quad (42)$$

Note that the expression (42) is not dependent on \mathbf{x}' , but it depends on $\boldsymbol{\theta}$. The parameter vector $\boldsymbol{\theta}$ is unknown. So, in order to obtain an estimate of (42) we can replace $\boldsymbol{\theta}$ by its estimate. Suppose that N observations $x(1), x(2), \dots, x(N)$ are available. Thus, for each model order, it is reasonable to replace $\boldsymbol{\theta}$ by its maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ which is a function of the observed data. As $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, this replacement minimizes (42) for each value of the model order. Now, in a slightly simplified notation, minimizing the K-L information changes to minimizing the Kullback-Leibler index given by

$$\Delta = E_{\mathbf{x}'} \left\{ -2 \ln g[\mathbf{x}' | \hat{\boldsymbol{\theta}}(\mathbf{x}''), \mathbf{x}''] \right\}, \quad (43)$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ given the observed data vector \mathbf{x}'' , and $g[\mathbf{x}' | \boldsymbol{\theta}]$ is the likelihood function. Usually, instead of using the exact maximum likelihood estimator of $\boldsymbol{\theta}$, an approximation of it is used as $\hat{\boldsymbol{\theta}}$ in (43).

Now, we solve our AR model order selection problem in the same-realization case by minimizing the Kullback-Leibler index Δ given by (43). We assume that the input process $\varepsilon(n)$ to the AR model is white Gaussian noise (WGN). In our problem, $\hat{\boldsymbol{\theta}}$ (the approximation of the maximum likelihood

estimator of the AR parameter vector $\boldsymbol{\theta}$, \mathbf{x}' (the future data of the AR model that depends on the observed data), and \mathbf{x}'' (vector of the observed data of the AR model) are defined as follows:

$$\hat{\boldsymbol{\theta}} = (\mathbf{b}_A^T, \hat{\sigma}_\varepsilon^2)^T, \quad (44)$$

$$\mathbf{x}' = x(N+1), \quad (45)$$

$$\mathbf{x}'' = [x(1), x(2), \dots, x(N)] = \mathbf{x}. \quad (46)$$

Substituting (44)–(46) into likelihood function $g(\mathbf{x}' | \hat{\boldsymbol{\theta}}, \mathbf{x}'')$, we obtain

$$g(\mathbf{x}' | \hat{\boldsymbol{\theta}}, \mathbf{x}'') = g[x(N+1) | \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x}]. \quad (47)$$

When the input process $\varepsilon(n)$ to the AR model is white Gaussian noise (WGN), $g[x(N+1) | \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x}]$ can be written as

$$g[x(N+1) | \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x}] = \frac{\exp\left\{-[x(N+1) - \mathbf{x}^T(N)\mathbf{b}_A]^2 / 2\hat{\sigma}_\varepsilon^2\right\}}{\sqrt{2\pi\hat{\sigma}_\varepsilon^2}}. \quad (48)$$

Taking logarithm from both sides of (48), we get

$$\begin{aligned} \ln g(\mathbf{x}' | \hat{\boldsymbol{\theta}}, \mathbf{x}'') &= \ln g[x(N+1) | \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x}] \\ &= -\ln(\sqrt{2\pi}) - \ln(\hat{\sigma}_\varepsilon) \\ &\quad - \frac{[x(N+1) - \mathbf{x}^T(N)\mathbf{b}_A]^2}{2\hat{\sigma}_\varepsilon^2}. \end{aligned} \quad (49)$$

Substituting (45) and (49) into (43), we have

$$\begin{aligned} \Delta &= -E_{x(N+1)} \{2 \ln g[x(N+1) | \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x}]\} \\ &= 2E_{x(N+1)} \left\{ \ln(\sqrt{2\pi}) + \ln(\hat{\sigma}_\varepsilon) \right. \\ &\quad \left. + \frac{[x(N+1) - \mathbf{x}^T(N)\mathbf{b}_A]^2}{2\hat{\sigma}_\varepsilon^2} \mid \mathbf{b}_A, \hat{\sigma}_\varepsilon^2, \mathbf{x} \right\} \quad (50) \\ &= 2 \ln(\sqrt{2\pi}) + 2 \ln(\hat{\sigma}_\varepsilon) \\ &\quad + \frac{E_{x(N+1)} \{[x(N+1) - \mathbf{x}^T(N)\mathbf{b}_A]^2 \mid \mathbf{b}_A, \mathbf{x}\}}{\hat{\sigma}_\varepsilon^2}. \end{aligned}$$

The parameter $\hat{\sigma}_\varepsilon^2$ in (50) can be replaced by the residual variance $S^2(q)$ given by (3). Thus, combining (14), (15), and (50), we obtain

$$\Delta = \ln(2\pi) + \ln[S^2(q)] + \frac{\text{PE}(q)}{S^2(q)}. \quad (51)$$

The prediction error $\text{PE}(q)$ is unknown; so we replace it by its estimate given by (36) to obtain

$$\Delta \approx \ln(2\pi) + \ln[S^2(q)] + \left[\frac{1 + q/(N-q)}{1 - q/(N-q)} \right]; \quad q \geq p. \quad (52)$$

The order q that minimizes (52) can be selected as the best AR model order. So, omitting the constant $\ln(2\pi)$ from (52), we obtain the finite sample AIC (AICF) criterion for AR model order selection in the same-realization case as

$$\text{AICF}(q) = \ln[S^2(q)] + \left[\frac{1 + q/(N-q)}{1 - q/(N-q)} \right]. \quad (53)$$

The AIC criterion, which was derived for the independent-realization case, is defined in the following way [4]:

$$\text{AIC}(q) = \ln[S^2(q)] + \frac{2q}{N}. \quad (54)$$

4. Simulation Results

To investigate the effectiveness of the proposed criteria FPEF and AICF, we consider the problem of autoregressive model order selection for simulated data. The simulated data have been produced by AR(0), AR(1), AR(2), AR(3), and AR(7) processes defined as follows:

$$\begin{aligned} x[n] &= \varepsilon(n), \\ x[n] &= .95x[n-1] + \varepsilon(n), \\ x[n] &= -1.4x[n-1] - .5x[n-2] + \varepsilon(n), \\ x[n] &= -1.08x[n-1] - 0.37x[n-2] - 0.042x[n-3] + \varepsilon(n), \\ x[n] &= 2.8x[n-1] - 3.22x[n-2] + 1.96x[n-3] \\ &\quad - .68x[n-4] + .13x[n-5] - .013x[n-6] \\ &\quad + .0005x[n-7] + \varepsilon(n), \end{aligned} \quad (55)$$

where $\varepsilon(n)$ is WGN with zero mean and variance $\sigma_\varepsilon^2 = 1$. We can define the SNR parameter for each AR model as the ratio of the average output power of the AR model to σ_ε^2 in dB. The SNRs of AR(0), AR(1), AR(2), AR(3), and AR(7) models are 0 dB, 10 dB, 10.13 dB, 5.11 dB, and 23 dB, respectively.

For each of the four models given by (55), 5000 independent simulation runs of 20 consecutive samples of the AR process were generated. In each simulation run, the first nineteen samples were used for estimating the coefficients of the AR model, and the last sample was used for computing the prediction error. Candidate orders were considered from 0 to 8. For each candidate order, the prediction error has been approximated by averaging over 5000 simulation runs. This true prediction error and the estimated prediction errors given by FPE and FPEF are shown for AR(0), AR(1), AR(2), AR(3), and AR(7) models in Figures 1, 2, 3, 4, and 5, respectively. These figures illustrate that when q/N is not small, the FPEF gives an estimate of the prediction error that is much better than FPE.

Table 1 gives the mean of the prediction error for the AR order selection criteria FPE, FPEF, AIC, and AICF. The mean of the prediction error is computed for each AR order selection criterion by averaging the prediction error over 5000 intervals of simulated data for the orders that the AR

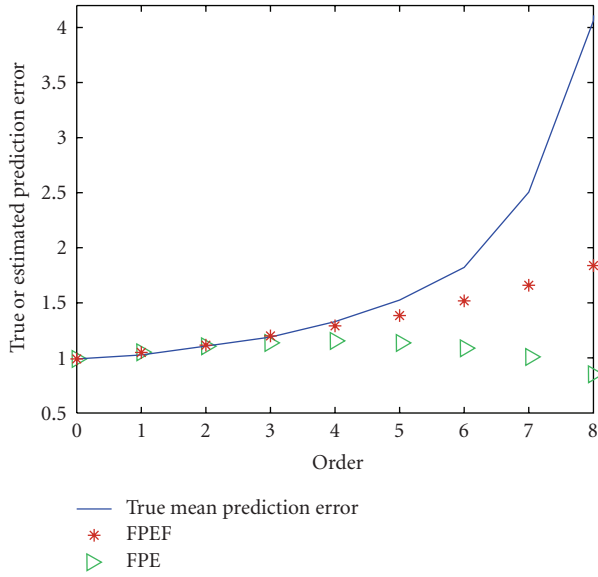


FIGURE 1: Estimation of prediction error for simulated AR(0) sequence with $N = 19$.

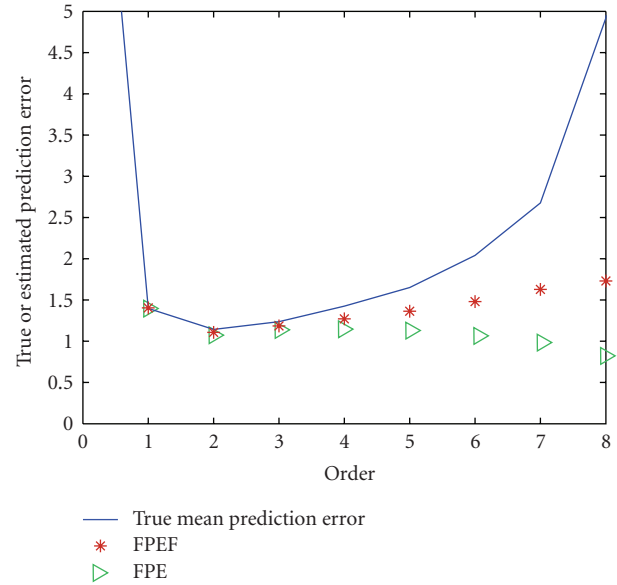


FIGURE 3: Estimation of prediction error for simulated AR(2) sequence with $N = 19$.

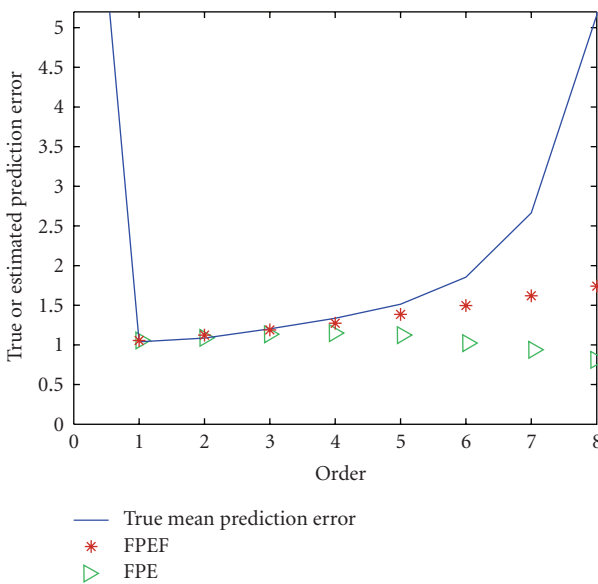


FIGURE 2: Estimation of prediction error for simulated AR(1) sequence with $N = 19$.

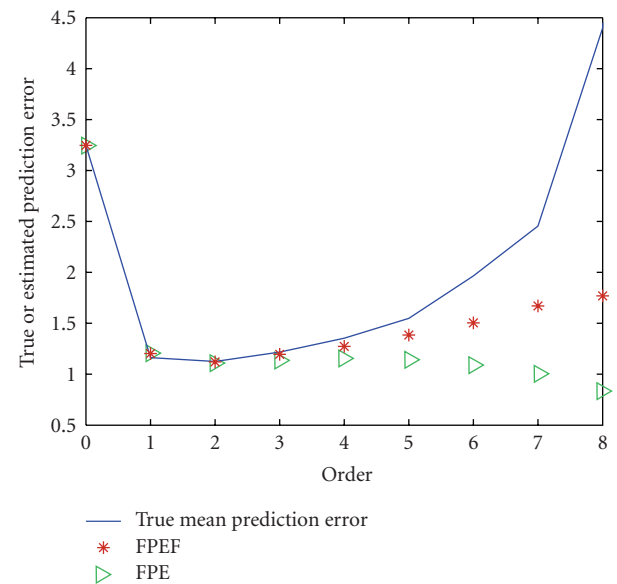


FIGURE 4: Estimation of prediction error for simulated AR(3) sequence with $N = 19$.

TABLE 1: Mean of prediction error for several order selection criteria.

	Criterion				
	MPE	FPE	FPEF	AIC	AICF
AR(0)	.31	3.35	2.02	3.17	1.02
AR(1)	.48	3.92	2.78	3.99	1.32
AR(2)	.51	4.42	2.87	4.55	1.39
AR(3)	.43	3.89	2.39	3.97	1.69
AR(7)	.48	6.18	4.50	6.24	1.88

order selection criterion has selected. The maximum and minimum candidate orders are $q_{\max} = 8$ and $q_{\min} = 0$, respectively, and the number of generated samples in each simulation run is 20 (the first nineteen samples are used for estimating the coefficients of the AR model and the last sample is used for computing the prediction error). The mean of the minimum prediction errors (MPE) that are possible in each run is also computed. It can be seen from Table 1 that the performance of AICF is better than the other criteria.

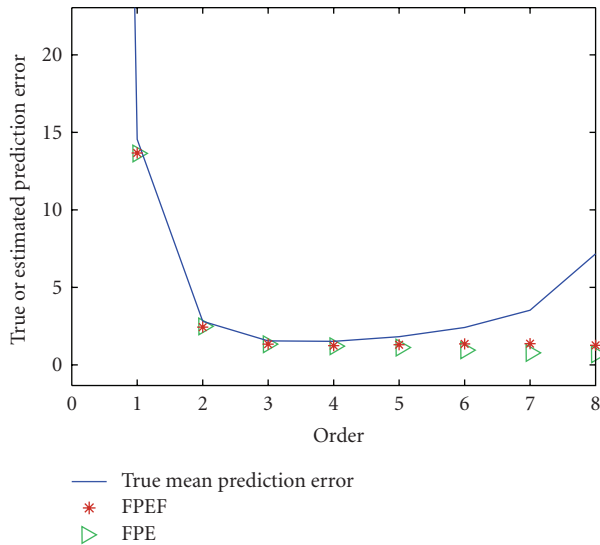


FIGURE 5: Estimation of prediction error for simulated AR(7) sequence with $N = 19$.

5. Conclusion

So far, few time series model selection theories have been established for same-realization prediction. In this paper, a new theoretical approximation was derived for the expectation of same-realization prediction error by using the LSF method for estimation of the AR parameters. Using this approximation and the approximation given in [16, 17] for the expectation of residual variance, the FPE and AIC criteria for AR order selection in the same-realization case were modified. The modified FPE and AIC criteria were called FPEF and AICF, respectively. Simulation results show that the bias in the estimates that FPEF gives for the prediction error is less than that of FPE. The performance of FPEF and AICF in AR model order selection was compared with FPE and AIC in the finite sample case. The results of this performance comparison showed that the performance of FPEF is better than FPE, and the performance of AICF is better than all other criteria. In the large sample case, the performance of FPEF and AICF is approximately identical to those of FPE and AIC, respectively.

References

- [1] G. E. P. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Cambridge University Press, Cambridge, UK, 1976.
- [2] S. M. Kay and S. L. Marple Jr., "Spectrum analysis—a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.
- [3] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.
- [4] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [5] H. Akaike, "Statistical predictor identification," *Annals of the Institute of Statistical Mathematics*, vol. 22, no. 1, pp. 203–217, 1970.
- [6] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [7] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [8] J. E. Cavanaugh, "A large-sample model selection criterion based on Kullback's symmetric divergence," *Statistics and Probability Letters*, vol. 42, no. 4, pp. 333–343, 1999.
- [9] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, pp. 297–307, 1989.
- [10] A.-K. Seghouane and M. Bekara, "A small sample model selection criterion based on Kullback's symmetric divergence," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3314–3323, 2004.
- [11] C.-K. Ing and C.-Z. Wei, "Order selection for same-realization predictions in autoregressive processes," *Annals of Statistics*, vol. 33, no. 5, pp. 2423–2474, 2005.
- [12] C.-K. Ing, "Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series," *Annals of Statistics*, vol. 35, no. 3, pp. 1238–1277, 2007.
- [13] M. Karimi, "On the residual variance and the prediction error for the LSF estimation method and new modified finite sample criteria for autoregressive model order selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2432–2441, 2005.
- [14] P. T. Broersen and H. E. Wensink, "On finite sample theory for autoregressive model order selection," *IEEE Transactions on Signal Processing*, vol. 41, no. 1, pp. 196–204, 1993.
- [15] P. M. T. Broersen and H. E. Wensink, "Autoregressive model order selection by a finite sample estimator for the Kullback-Leibler discrepancy," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 2058–2061, 1998.
- [16] M. Karimi, "Finite sample criteria for autoregressive model order selection," *Iranian Journal of Science and Technology, Transaction B*, vol. 31, no. 3, pp. 329–344, 2007.
- [17] M. Karimi, "A corrected FPE criterion for autoregressive processes," in *Proceedings of the 15th European Signal Processing Conference (EUSIPCO '07)*, pp. 803–806, Poznan, Poland, September 2007.
- [18] M. Karimi, "Finite sample AIC for autoregressive model order selection," in *Proceedings of the IEEE International Conference on Signal Processing and Communications (ICSPC '07)*, pp. 1219–1222, Dubai, United Arab Emirates, November 2007.
- [19] M. Wax, "Order selection for AR models by predictive least-squares," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 581–588, 1988.
- [20] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inferences*, Springer, New York, NY, USA, 2002.