Communications in Statistics—Simulation and Computation®, 44: 1187–1196, 2015

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2013.809189

Resampling for Order Estimation of Autoregressive Models with Missing Data

ABDELAZIZ EL MATOUAT,[1] FREEDATH DJIBRIL MOUSSA,[2] AND HASSANIA HAMZAOUI[2]

[1]University of Le Havre, Le Havre, France
[2]University of Fez, Fez, Morocco

In this article, we consider the order estimation of autoregressive models with incomplete data using the expectation–maximization (EM) algorithm-based information criteria. The criteria take the form of a penalization of the conditional expectation of the log-likelihood. The evaluation of the penalization term generally involves numerical differentiation and matrix inversion. We introduce a simplification of the penalization term for autoregressive model selection and we propose a penalty factor based on a resampling procedure in the criteria formula. The simulation results show the improvements yielded by the proposed method when compared with the classical information criteria for model selection with incomplete data.

Keywords Autoregressive model; EM algorithm; Information criteria; Missing data; Resampling.

Mathematics Subject Classification 62M10; 62F07.

1. Introduction

Let $\{x_t\}$ be a pth-order causal autoregressive process AR(p) satisfying

$$x_t - \sum_{j=1}^{p} \phi_j x_{t-j} = \varepsilon_t, \forall t \in \mathbb{Z},$$

where $\{\varepsilon_t\}$ is a white noise process with mean 0 and variance σ_p^2 and $\phi_1, \ldots, \phi_p$ are real coefficients such that $1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| < 1$. The order p, the coefficients $\phi_1, \ldots, \phi_p$, and the variance σ_p^2 are supposed unknown. Given a sample $\underline{x}$ of size n, p can be estimated by minimizing information criteria like Akaike information criterion (AIC; Akaike, 1973) or Bayesian information criterion (BIC; Schwarz, 1978). These criteria have the form

$$\text{IC}(k) = \log \hat{\sigma}_k^2 + k \frac{C_n}{n},$$

Received August 29, 2012; Accepted May 21, 2013
Address correspondence to Abdelaziz El Matouat, University of Le Havre, 25, rue Philippe Lebon, BP420, Le Havre, 76057 France; E-mail: abdelaziz.elmatouat@univ-lehavre.fr

where k is the order of the autoregressive candidate process, $\hat{\sigma}_k^2$ is a $\sqrt{n}$-consistent estimator of the white noise variance, and C_n is a specified term that may be constant ($C_n = 2$ for AIC) or depend on the sample size ($C_n = \log n$ for BIC). In the criterion's formulation, the log-likelihood term is a decreasing function of the order and it measures the adequacy of the approximating model to the data. The penalization term $k\frac{C_n}{n}$ that increases along with the model complexity is added in order to pick a parsimonious model.

In practice, the application of these criteria may not lead to satisfactory order selection especially with small samples. The choice of the penalty factor $\frac{C_n}{n}$ affects the performance of the criteria. AIC can overfit the model while BIC is prone to underfitting. To overcome these defects, Chen et al. (1993) suggested a resampling procedure to obtain a data-adaptive penalty factor for complete data.

In this article, we shall be concerned with the order determination of an autoregressive process with missing data. For model selection in incomplete data settings, the information criteria have been modified using the expectation–maximization (EM) algorithm (Dempster et al., 1977). Cavanaugh and Shumway (1998) derived a variant of AIC called AIC_{cd} and considering the appropriate penalization, we define a modified BIC denoted as BIC_{cd}. Those two criteria that take into account the uncertainty due to the missing data are written as the sum of two components where the first is the conditional expectation of the complete data log-likelihood given the observed data. The second component is a penalization term with a penalty factor that depends on the sample size and the impact of the missing data on the fitted model. We modify AIC_{cd} and BIC_{cd} by introducing a penalty factor defined from resampled data. Our procedure extends Chen et al.'s (1993) technique to the incomplete data case.

The article is organized as follows: we present in Section 2 the autoregressive parameters estimation methods considered in our procedure: the Yule–Walker and maximum likelihood estimation and the EM algorithm. In Section 3, we give the missing data information criteria for autoregressive order estimation and we introduce the proposed simplification of their penalization. Section 4 is devoted to the choice of the resampling penalty factor, we present the derivation of the resampling procedure and we describe its practical implementation. Finally, the improvements yielded by our method are validated on simulated data in Section 5.

2. Autoregressive Parameters Estimation Methods

2.1. *Autoregressive Parameters Estimation with a Complete Sample*

Let $\underline{x} = (x_1, \ldots, x_n)$ be a sample of size n of a causal autoregressive model AR(p). For $k = 1, 2, \ldots$, the coefficient vector $\phi_k = (\phi_{k1}, \ldots, \phi_{kk})'$ and the variance σ_k^2, based on fitting an AR(k) process to the data, satisfy the Yule–Walker equations:

$$\phi_k = \Gamma_k^{-1}\gamma_k, \tag{1}$$

$$\sigma_k^2 = \gamma(0) - \phi_k\gamma_k = \gamma(0)\prod_{i=1}^{k}(1 - \phi_{ii}^2), \tag{2}$$

where $\gamma(.)$ is the autocovariance function of the process, $\Gamma_k = [\gamma(i-j)]_{i,j=1}^{k}$ is the covariance matrix, and $\gamma_k = (\gamma(1), \ldots, \gamma(k))'$.

The Yule–Walker estimates $\tilde{\phi}_k$ and $\tilde{\sigma}_k^2$ of ϕ_k and σ_k^2 are obtained by equating the sample autocovariance function $\tilde{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h} x_t x_{t+h}$ to $\gamma(h)$ in Eqs. (1) and (2) and solving the resulting equations.

An alternative method for the parameters estimation is the maximum likelihood. The Gaussian joint density function of the sample is

$$f\left(\underline{x}|\phi_k, \sigma_k^2\right) = \left(2\pi\sigma_k^2\right)^{-\frac{n}{2}} |M_k^{(k)}|^{\frac{1}{2}} \exp\left[-\frac{S(\phi_k)}{2\sigma_k^2}\right],$$

where $M_k^{(k)} = \Gamma_k^{-1}\sigma_k^2$ and

$$S(\phi_k) = \underline{x}_k' M_k^{(k)} \underline{x}_k + \sum_{t=k+1}^{n} \left(x_t - \phi_{k1}x_{t-1} - \cdots - \phi_{kk}x_{t-k}\right)^2,$$

with $\underline{x}_k = (x_1, \ldots, x_k)$. The maximum likelihood estimates (MLEs) of ϕ_k and σ_k^2 maximize the log-likelihood

$$\begin{aligned} l\left(\phi_k, \sigma_k^2, \underline{x}\right) &= -\frac{n}{2}\log\sigma_k^2 + \frac{1}{2}\log|M_k^{(k)}| - \frac{S(\phi_k)}{2\sigma_k^2}, \\ &\approx -\frac{n}{2}\log\sigma_k^2 - \frac{S(\phi_k)}{2\sigma_k^2}. \end{aligned} \tag{3}$$

The variance estimate is given by

$$\hat{\sigma}_k^2 = \frac{S(\hat{\phi}_k)}{n}, \tag{4}$$

and the coefficient vector estimate $\hat{\phi}_k$ is obtained by minimizing $S(\phi_k)$. The details of computation are described in Box et al. (1994).

2.2. *Autoregressive Parameters Estimation with Missing Data*

When the sample $\underline{x}$ contains missing values, we can write $\underline{x} = (x_{\text{obs}}, x_{\text{mis}})$, where x_{obs} denotes the observed data and x_{mis}, the missing data. The sample $\underline{x}$ is referred to as the complete data. The MLEs of parameters are found with the EM algorithm (Dempster et al., 1977). Let $\theta = (\phi_k, \sigma_k^2)$ denote the parameter vector and $Q(\theta_1, \theta_2) = E[l(\theta_1, \underline{x})|x_{\text{obs}}, \theta_2]$. Each iteration of the algorithm consists in two steps:

E step (Expectation): To estimate the conditional expectation $Q(\theta; \theta^{(t)})$ of the complete data log-likelihood given x_{obs} and using the current estimate $\theta^{(t)}$ of θ.

M step (Maximization): To choose $\theta^{(t+1)}$ that maximizes $Q(\theta; \theta^{(t)})$ with respect to θ.

These steps are repeated until a stopping criterion is met. The EM algorithm implicitly defines a map $F : \theta^{(t)} \longmapsto \theta^{(t+1)}$. By construction, the map F is derivable and a first-order Taylor expansion of $\theta^{(t+1)} = F(\theta^{(t)})$ about the MLE $\hat{\theta}$ leads to the approximation

$$(\theta^{(t+1)} - \hat{\theta})' \approx (\theta^{(t)} - \hat{\theta})' DF,$$

where DF is the Jacobian matrix for F at the point $\hat{\theta}$. The computation of DF is done via numerical differentiation and is discussed in Meng and Rubin (1991). This matrix is useful to obtain both the observed data and complete data information matrices of the parameter estimates.

The EM algorithm derived quantities $\hat{\theta}$, $Q(\hat{\theta};\hat{\theta})$, and the information matrices are used to define the missing data information criteria.

3. Missing Data Information Criteria for Autoregressive Model Selection

The classical model selection criteria like AIC or BIC depend on the observed data log-likelihood. Shimodaira (1994) argued that it is more natural or desirable to consider a criterion based on the complete data and proposed the PDIO criterion where the penalization is EM algorithm related. On the motivation provided by Shimodaira (1994), Cavanaugh and Shumway (1998) argued that in incomplete data framework, AIC can be interpreted as a measure of separation between the fitted model for the observed (incomplete) data and the true generating one that presumely gave rise to the incomplete data. Therefore, they formulated an EM algorithm-based version of AIC denoted as $\mathrm{AIC_{cd}}$ criterion as an estimation of the expected complete data Kullback–Leibler divergence between the true unknown model and the fitted one. Cavanaugh and Neath (1999) generalized the derivation of BIC to any model. Similarly, we define the $\mathrm{BIC_{cd}}$ criterion for model selection from incomplete data. In $\mathrm{AIC_{cd}}$ and $\mathrm{BIC_{cd}}$ criteria, both the log-likelihood and the penalization terms depend on the complete data. The general form of these criteria is

$$\mathrm{IC_{cd}}(k) = -2Q(\hat{\theta};\hat{\theta}) + \mathrm{Tr}(I_{oc}.I_o^{-1})C_n,$$

where I_o is the observed information matrix evaluated at the MLE $\hat{\theta}$ and I_{oc} is the conditional expectation of the complete data information matrix given the observed data evaluated at $\hat{\theta}$. Typical values of C_n are 2 for $\mathrm{AIC_{cd}}$ and $\log n$ for $\mathrm{BIC_{cd}}$.

For an autoregressive process, the formula of the criteria is given by

$$\mathrm{IC_{cd}}(k) = \log \hat{\sigma}_k^2 + \mathrm{Tr}\big(I_{oc}.I_o^{-1}\big)\frac{C_n}{n}, \tag{5}$$

where k is the order of the AR(k) candidate model and $\hat{\sigma}_k^2$ is the estimate of the variance obtained with the EM algorithm.

For a given model, I_o^{-1} and I_{oc} matrices are not directly evaluated. Cavanaugh and Shumway (1998) showed that $\mathrm{Tr}(I_{oc}.I_o^{-1})$ can be written as $k+\mathrm{Tr}[DF(I-DF)^{-1})]$, where DF is the Jacobian matrix of the map F and I is the identity matrix of size k. Thus,

$$\mathrm{IC_{cd}}(k) = \log \hat{\sigma}_k^2 + (k + \mathrm{Tr}[DF(I-DF)^{-1})])\frac{C_n}{n}. \tag{6}$$

Set n_{obs} be the number of observed data and n_{mis} be the number of missing data such that $n = n_{\mathrm{obs}} + n_{\mathrm{mis}}$. From Meng and Rubin (1991), $I_o = V^{-1}$, where V is the sample variance covariance matrix of the parameters estimates. For the AR(k) model, the variance-covariance matrix of the parameters estimates based on n_{obs} data is $V = n_{\mathrm{obs}}^{-1}\hat{\sigma}_k^2\hat{\Gamma}_k^{-1}$, where $\hat{\Gamma}_k$ is the sample autocovariance matrix, hence $I_o = n_{\mathrm{obs}}\hat{\sigma}_k^{-2}\hat{\Gamma}_k$.

Since the complete data density function belongs to the regular exponential family, the matrix I_{oc} is the complete data Fisher information matrix (see, e.g., McLachlan and Krishnan, 1997). For an autoregressive model, the Fisher information matrix supplies the inverse of the variance-covariance of the parameter estimates (Box et al., 1994), thus we have $I_{oc} = n\hat{\sigma}_k^{-2}\hat{\Gamma}_k$.

Therefore, $\mathrm{Tr}(I_{oc}.I_o^{-1}) = \frac{n}{n_{\mathrm{obs}}}k$ and we deduce the following criteria for autoregressive order estimation with incomplete data

$$\mathrm{IC}_{\mathrm{cd}}(k) = \log \hat{\sigma}_k^2 + \frac{C_n}{n_{\mathrm{obs}}}k. \tag{7}$$

When the sample $\underline{x}$ is complete, the $\mathrm{IC}_{\mathrm{cd}}$ criteria are exactly IC since $n_{\mathrm{obs}} = n$.

Remark. Contrasting the representations Eq. (6) and Eq. (7) of $\mathrm{IC}_{\mathrm{cd}}$ and using the decomposition $\frac{C_n}{n_{\mathrm{obs}}} = \frac{C_n}{n} + \frac{C_n}{n}\frac{n_{\mathrm{mis}}}{n_{\mathrm{obs}}}$, we deduce that

$$\frac{n_{\mathrm{mis}}}{n_{\mathrm{obs}}}k = \mathrm{Tr}[DF(I - DF)^{-1})].$$

This agrees with the interpretation of $\mathrm{Tr}[DF(I - DF)^{-1})]$ as the extent to which the amount of missing data affects the fitted model given by Cavanaugh and Shumway (1998). They also stated that this term will be substantial in settings where the amount of missing data is large relative to the complexity of the fitted model as can be seen in the equation above.

4. Resampling Penalty Factor for Incomplete Data

4.1. *Derivation of the Resampling Procedure*

Computing the order estimate with missing data model selection criteria requires the EM algorithm. At the final iteration of the algorithm, the missing data x_{mis} are estimated so that we complete the former incomplete data. Let $\tilde{x}$ be the completed data. Although the number of missing data n_{mis} may be large, in practice n_{mis} is generally small as compared to the size n of the sample. So, $\tilde{x}$ can be viewed as data from the autoregressive model. Then we apply the resampling procedure of Chen et al. (1993) to find a penalty factor. This procedure is based on the properties of autoregressive parameters and the Yule–Walker equations. The data-generating process is supposed to be an AR(p), hence $\phi_p \neq 0$. For a candidate model AR(k), if $k \geq p$, the coefficient vector $\phi_k = (\phi_{k1}, \ldots, \phi_{kk})'$ is $\phi_k = (\phi_1, \ldots, \phi_p, 0, \ldots, 0)'$. Thus,

$$\phi_{ki} = \begin{cases} \phi_i & \text{if } i = 1, \ldots, p, \\ 0 & \text{if } i = p+1, \ldots, k. \end{cases}$$

From the Yule–Walker Eq. (2), we have

$$\sigma_k^2 = \gamma(0)\prod_{i=1}^{k}\left(1 - \phi_{ii}^2\right) = \gamma(0)\prod_{i=1}^{p}\left(1 - \phi_{ii}^2\right)\prod_{i=p+1}^{k}\left(1 - \phi_{ii}^2\right).$$

It follows that

$$\sigma_k^2 = \gamma(0)\prod_{i=1}^{p}\left(1 - \phi_{ii}^2\right) = \sigma_p^2 \text{ if } k \geq p. \tag{8}$$

If $k < p$, we have $\dfrac{\sigma_p^2}{\sigma_k^2} = \dfrac{\gamma(0)\prod_{i=1}^{p}\left(1 - \phi_{ii}^2\right)}{\gamma(0)\prod_{i=1}^{k}\left(1 - \phi_{ii}^2\right)} = \prod_{i=k+1}^{p}(1 - \phi_{ii}^2).$

Since ϕ_{ii} are autocorrelations, $\phi_{ii}^2 \leq 1$ and we get

$$\sigma_p^2 \leq \sigma_k^2 \qquad \text{if } k < p. \tag{9}$$

For the order selection, the penalty factor should be chosen such that

$$\log \sigma_p^2 + p\frac{C_n}{n_{\text{obs}}} \leq \log \sigma_k^2 + k\frac{C_n}{n_{\text{obs}}} \text{ for any } k \neq p,$$

which implies that

$$\text{for } k > p \qquad \frac{C_n}{n_{\text{obs}}} \geq \frac{\log \sigma_p^2 - \log \sigma_k^2}{k - p} = 0 \text{ by Eq. (8)},$$

$$\text{and for } k < p \qquad \frac{C_n}{n_{\text{obs}}} \leq \frac{\log \sigma_k^2 - \log \sigma_p^2}{p - k} = b \geq 0 \text{ by Eq. (9)}.$$

Hence, we have $0 \leq \frac{C_n}{n_{\text{obs}}} \leq b$. Since b is unknown, an interval $[a_n, b_n]$ that contains $\frac{C_n}{n_{\text{obs}}}$ and converges to $[0, b]$ is built using the following proposition.

Proposition 4.1. *Let $(x_1, \ldots, x_n)$ be a sample of size* n *from an* AR(p) *process and $K \geq p$ a fixed integer. If we define*

$$a_n = \begin{cases} 0 & \text{if } p = K, \\ \max_{p<j\leq K} \frac{\log \tilde{\sigma}_p^2 - \log \tilde{\sigma}_j^2}{j-p} & \text{if } p < K, \end{cases}$$

and

$$b_n = \begin{cases} \infty & \text{if } p = 0, \\ \min_{0<j<p} \frac{\log \tilde{\sigma}_j^2 - \log \tilde{\sigma}_p^2}{p-j} & \text{if } p > 0, \end{cases}$$

where $\tilde{\sigma}_0^2 = \tilde{\gamma}(0)$. Then, as $n \to \infty$, we have

1. *$a_n 1 \xrightarrow{\text{a.s.}} 0$ and $b_n \xrightarrow{\text{a.s.}} 0$. If $\min_{1\leq j\leq p} |\phi_{jj}| > 0$, then $b > 0$.*
2. *If $a_n \leq b_n$, then for any scriptstyle$\frac{C_n}{n_{\text{obs}}} \in [a_n, b_n]$,*

$$\log \tilde{\sigma}_p^2 + p\frac{C_n}{n_{\text{obs}}} = \min_{0\leq j\leq K} \left\{ \log \tilde{\sigma}_j^2 + j\frac{C_n}{n_{\text{obs}}} \right\}.$$

This proposition is the same as Proposition 2.1. in Chen et al. (1993) with $\frac{C_n}{n_{\text{obs}}}$ rather than $\frac{C_n}{n}$.

Proposition 4.1 indicates how to choose a penalty factor leading to the selection of the correct order when p is known. Since it is not the case, a resampling procedure is used to find a sequence of nonempty admissible sets for $\frac{C_n}{n_{\text{obs}}}$. Set $K_1 < K$, where K is the maximum order of the competing models. For $k = 0, 1, \ldots, K_1$, samples $y^{(k)}$ are generated from an AR(k) process defined with parameter estimates of the original data. For each $k \leq K_1$, an interval $I_n^{(k)} = [a_n^{(k)}, b_n^{(k)}]$ for the penalty factor is computed by applying Proposition 4.1. If the intersection of these intervals $I_n = \bigcap_{k=0}^{K_1} I_n^{(k)}$ is nonempty, it provides a range of values for a suitable penalty factor. The construction of I_n and its asymptotic behavior are given by the following proposition.

Proposition 4.2. (Chen et al., 1993) *Suppose that the conditions in Proposition 4.1 are satisfied and* $E[\varepsilon_t^4] < \infty$. *Let* $K_1 \leq p$ *be a nonnegative integer. We define*

$$\tilde{\varepsilon}_t = x_t - \tilde{\phi}_{K1}x_{t-1} - \tilde{\phi}_{K2}x_{t-2} - \cdots - \tilde{\phi}_{KK}x_{t-K}, \; t = K+1, \ldots, n.$$

For $k = 0, 1, \ldots, K_1$, *let* $\{y_1^{(k)}, \ldots, y_n^{(k)}\}$ *be a sequence of observations from the* AR(k) *model:*

$$y_t^{(k)} = \tilde{\phi}_{k1}y_{t-1}^{(k)} - \cdots - \tilde{\phi}_{kk}y_{t-k}^{(k)} + \varepsilon_t^*,$$

where $\{\varepsilon_t^*\}$ *is an* iid *sequence whose distribution is the empirical distribution corrected to have the mean 0 of* $\{\tilde{\varepsilon}_t\}$. *For* $k = 0$, $y_t^0 = \varepsilon_t^*$.

For each $k = 0, \ldots, K_1$, *let* $I_n^{(k)} = [a_n^{(k)}, b_n^{(k)}]$ *denote the interval obtained by applying Proposition 4.1 to the* AR(k) *series* $y_t^{(k)}$. *Then for almost all sample of sequences of* $\{x_t\}$, *we have*

1. $a_n = \max_{0\leq k\leq K_1} a_n^{(k)} \xrightarrow{P_n} 0$,
2. $b_n = \min_{0\leq k\leq K_1} b_n^{(k)} \xrightarrow{P_n} b \geq 0$, *where* $\xrightarrow{P_n}$ *denotes convergence in probability conditional on* $x_1, \ldots, x_n$.
3. *If* $a_n \leq b_n$, *then* $I_n = \bigcap_{k=0}^{K_1} I_n^{(k)} = [a_n, b_n]$ *converges in conditional probability to a nonempty set.*

The resampling penalty factor is defined by

$$C_{nr} = \begin{cases} a_n + cb_n \frac{C_n}{n_{obs}} & \text{if } a_n < b_n, \\ b_n \frac{C_n}{n_{obs}} & \text{else,} \end{cases}$$

where c is such that $C_{nr} \leq b_n$. We denote IC_{cdr} the criteria with the resampling penalty C_{nr}

$$\text{IC}_{cdr}(k) = \log \tilde{\sigma}_k^2 + kC_{nr}.$$

4.2. *Implementation*

Let $\underline{x} = (x_{obs}, x_{mis})$ be an incomplete sample of size n from an autoregressive model. Set K a large integer and $K_1 < K$. The resampling procedure to obtain the penalty factor C_{nr} for the missing data criteria consists of four steps.

Step 1 Choose initial values for the parameters and run the EM algorithm to obtain the completed sample $\tilde{x} = (x_1, \ldots, x_n)$ in fitting an AR(K) model to $\underline{x}$.

Step 2 Apply Proposition 4.2 to generate n_r data $y_1^{(k)}, \ldots, y_{n_r}^{(k)}$ and compute $a_n^{(k)}$ and $b_n^{(k)}$ for $k = 0, \ldots, K_1$.

Step 3 Compute $a_n = \max_{0\leq k\leq K_1} a_n^{(k)}$ and $b_n = \min_{0\leq k\leq K_1} b_n^{(k)}$.

Step 4 If $a_n < b_n$, choose a value C_{nr} in $[a_n, b_n]$. Otherwise, reduce the value of K_1 by 1 and return to Step 3.

Table 1
Frequencies of estimated orders over 100 replications with sample size $n = 40$ and different sizes n_r for the test series y in the resampling procedure

Criteria	Orders											
	$n_r = n = 40$						$n_r = 100$					
	1	2	3	4	5	6	1	2	3	4	5	6
AIC	19	73	5	2	1	0	19	73	5	2	1	0
BIC	35	62	2	0	1	0	35	62	2	0	1	0
AIC_r	8	78	6	1	5	2	3	77	9	3	5	3
BIC_r	10	79	4	1	4	2	6	80	6	2	5	1

To evaluate the performance of the proposed procedure, we compare the criteria AIC_{cd}, BIC_{cd}, AIC_{cdr}, and BIC_{cdr} on simulated data from an autoregressive model.

5. Numerical Study

In our simulations, we considered the following autoregressive model

$$\text{AR}(2): \quad x_t - 1.4x_{t-1} + 0.49x_{t-2} = \varepsilon_t,$$

where $\{\varepsilon_t\} \sim \mathcal{N}(0, 1)$. We generated 100 samples of size $n = 40$ and $n = 100$ and we estimated the order with AIC, BIC, AIC_r, and BIC_r. Next, percentages $P_{\text{mis}} = 10\%$ and $P_{\text{mis}} = 20\%$ of data were discarded at random. We computed the order estimates with the missing data criteria AIC_{cd}, BIC_{cd}, AIC_{cdr}, and BIC_{cdr}. In the EM algorithm, the estimates from the largest block of observed data were used as starting values. For the resampling scheme, we considered $K = 6$ and $K_1 = 2$. To reduce sampling errors, we simulated 50 replications of the test series y and took the average of the variances over these replications to obtain a_n and b_n. The size n_r of the generated data y can be chosen larger than n. For samples of size 40 in absence of missing data, we apply the resampling procedure with samples y of size $n_r = n = 40$ and $n_r = 100$. The results reported in Table 1 indicate that there is less underfitting with $n_r = 100$ as compared to $n_r = 40$. Hence, we used $n_r = 100$ in all the simulations.

The penalty factor C_{nr} satisfies

$$a_n \leq C_{\text{nr}} = a_n + cb_n \frac{C_n}{n_{\text{obs}}} \leq b_n.$$

Thus, for the constant c, we have

$$0 \leq c \leq \left(1 - \frac{a_n}{b_n}\right) \frac{n_{\text{obs}}}{C_n}.$$

To keep the same value of c for AIC_{cdr} and BIC_{cdr} penalty factors, c must be chosen such that

$$c \leq \left(1 - \frac{a_n}{b_n}\right) \frac{n_{\text{obs}}}{\log n} = \inf\left(\left(1 - \frac{a_n}{b_n}\right) \frac{n_{\text{obs}}}{2}, \left(1 - \frac{a_n}{b_n}\right) \frac{n_{\text{obs}}}{\log n}\right).$$

Table 2
Frequencies of estimated orders over 100 replications from the AR(2) model with sample size $n = 40$

P_{mis}	Criteria	Orders					
		1	2	3	4	5	6
0	AIC_{cd}	19	73	5	2	1	0
	BIC_{cd}	35	62	2	0	1	0
	AIC_{cdr}	3	77	9	3	5	3
	BIC_{cdr}	6	80	6	2	5	1
0.1	AIC_{cd}	28	57	7	4	3	1
	BIC_{cd}	47	48	4	1	0	0
	AIC_{cdr}	12	59	14	5	7	3
	BIC_{cdr}	16	64	10	6	3	1
0.2	AIC_{cd}	35	52	5	5	3	0
	BIC_{cd}	49	49	0	2	0	0
	AIC_{cdr}	15	59	6	8	6	6
	BIC_{cdr}	19	63	4	6	3	5

In our simulations, we used $c=(1-\frac{a_n}{b_n})\frac{n}{2\log n}$. All the computations were conducted using the software R. The results obtained for frequencies of estimated orders are reported in Tables 2 and 3.

From the simulation results, we observe that the criteria AIC_{cdr} and BIC_{cdr} with the resampling penalty factor perform better than AIC_{cd} and BIC_{cd} in identifying the true model. As the frequency of missing data grows, the correct order selection is deteriorated for all the criteria but our procedure still yields better performance. AIC_{cdr} and BIC_{cdr} underfit to a much lesser extent than AIC_{cd} and BIC_{cd}, respectively. The correct identification is

Table 3
Frequencies of estimated orders over 100 replications from the AR(2) model with sample size $n = 100$

P_{mis}	Criteria	Orders					
		1	2	3	4	5	6
0	AIC_{cd}	2	79	16	2	0	1
	BIC_{cd}	4	93	3	0	0	0
	AIC_{cdr}	1	93	4	1	0	1
	BIC_{cdr}	1	98	0	0	0	1
0.1	AIC_{cd}	3	74	16	3	3	1
	BIC_{cd}	9	88	2	1	0	0
	AIC_{cdr}	2	88	7	1	2	0
	BIC_{cdr}	7	91	1	0	1	0
0.2	AIC_{cd}	10	63	11	12	2	2
	BIC_{cd}	15	81	4	0	0	0
	AIC_{cdr}	9	76	7	6	2	0
	BIC_{cdr}	12	83	1	3	1	0

improved with samples of size 100 as compared with samples of size 40. We also observe that the BIC_{cdr} criterion outperforms the other criteria.

Conclusion

In this article, we presented a refinement of the information criteria for autoregressive model selection with incomplete data. We considered the EM algorithm-based information criteria AIC_{cd} and BIC_{cd} and we proposed a procedure to find resampling penalty factors for these criteria. We obtained AIC_{cdr} and BIC_{cdr} and compared them with AIC_{cd} and BIC_{cd} criteria on simulated samples. Our numerical study indicates that the order selection performance was improved with the resampling penalty factor and underfitting was reduced. Thus, the resampling method remains interesting for autoregressive model selection from incomplete data.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Csaki, F., eds. *Second International Symposium of Information Entropy Theory.* Budapest: Akademia Kiado, pp. 267–281.

Box, G. E. P., Jenkins, G. M., Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control.* 3rd ed. Englewood Cliffs: Prentice Hall.

Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods. Springer Series in Statistics.* 2nd ed.

Cavanaugh, J. E., Neath, A. A. (1999). Generalizing the derivation of the Schwarz information criterion. *Communications in Statistics Theory and Methods* 28: 49–66.

Cavanaugh, J. E., Shumway, R. H. (1998). An Akaike information criterion for model selection in the presence of incomplete data. *Journal of Statistical Planning and Inference* 67: 45–65.

Chen, C., Davis, A. R., Brockwell, P. J., Bai, Z. (1993). Order determination for autoregressive processes using resampling methods. *Statistica Sinica* 3: 481–500.

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

Ibrahim, J. G., Zhu, H., Tang, N. (2008). Model selection criteria for missing-data problems using the EM algorithm. *Journal of the American Statistical Association* 103(484): 1648–1658.

Little, R. J. A., Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

McLachlan, G., Peel, D. (2000). *Finite Mixture Models.* New York: Wiley.

McLachlan, G. J., Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: Wiley-Interscience.

Meng, X. L., Rubin, D. B. (1991). Using EM algorithm to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association* 86: 899–909.

Rao, C. R., Wu, Y. (2001). *On Model Selection. IMS Lecture Notes Monograph Series.* Vol. 38.

R Development Core Team, (2011). *R: A Language and Environment for Statistical Computing.* Vienna Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6: 461–464.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike information criterion. *Biometrika* 63: 117–126.

Shimodaira, H. (1994). A new criterion for selecting models from partially observed data. In: *Selecting Models from Data: Artificial Intelligence and Statistics IV (Lecture Notes in Statistics).* Vol. 89. New York: Springer, pp. 21–29.