ROBUST'2000, 393 – 400

SOME NOTES CONCERNING PREDICTION IN AR PROCESSES

PETR ZVÁRA

Abstract. The prediction of the $(n+s)$-th observation of the p-th order autoregressive process is studied. The mean squared error of the predictor (MSEP) when the autoregressive parameters are estimated by least squares is obtained to terms of order n^{-1} for some low order processes. It is shown that model overfitting increases MSEP. The naive prediction interval for X_{n+s} obtained by employing the estimated autoregressive model for prediction is considered. The overall coverage probability is evaluated to order n^{-1} in a special case. It is lower than the nominal one, because such prediction procedure ignores the uncertainty in the model parameters.

В работе изучается прогнозирование значения X_{n+s} в процессе авторегрессии порядка p с неизвестными коеффициентами. Приведенно явное выражение для члена порядка n^{-1} среднеквадратической ошибки прогноза в некоторых моделях. Далее рассмотрен наивный интервальный прогноз для X_{n+s}, построенный с помощю оценок неизвестных параметров. В частном случае вычислена его надежность.

1. Introduction

Let the autoregressive time series $\{X_t\}$ satisfy

$$X_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + e_t,\ \ t = 1, 2, \ldots, n, \tag{1}$$

where $\{e_t\}$ is a sequence of independent $N(0, \sigma^2)$ random variables and $X_0, X_{-1}, \ldots, X_{1-p}$ are given random variables. The characteristic equation associated with model (1) is

$$z^p - \sum_{j=1}^{p} a_j z^{p-j} = 0. \tag{2}$$

We assume that the process is a strictly stationary normal process, hence the roots of (2) are less than one in absolute value and $X_0, \ldots, X_{1-p}$ are normal random variables with the same covariance structure as $X_{t+p-1}, \ldots, X_t$ for all $t > 1 - p$.

We adopt a standard multivariate representation for the process (1). Let $\boldsymbol{X}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1}, 1)'$ and $\boldsymbol{e}_t = (e_t, 0, \ldots, 0)'$. Then we have

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{e}_t, \tag{3}$$

2000 *Mathematics Subject Classification.* Primary 60G25; Secondary 62M10.

Key words and phrases. Stationary process, prediction, coverage probability.

Acknowledgement. The research was supported by grants GAČR 201/00/0770 and MSM 113200008.

where

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p & a_0 \\ 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

The least squares predictor of X_{n+s} given a past history $\{X_{1-p}, \dots, X_n\}$ is $\bar{X}_{n+s} = a_0 + \sum_{j=1}^p a_j \bar{X}_{n+s-j}$, where $\bar{X}_t = X_t$ if $t \le n$. If the parameters $\boldsymbol{a} = (a_1, \dots, a_p, a_0)'$ and σ^2 must be estimated, the predictor

$$\hat{X}_{n+s} = \hat{a}_0 + \sum_{j=1}^{p} \hat{a}_j \hat{X}_{n+s-j},\ \hat{X}_t = X_t \text{ if } t \le n \tag{4}$$

is obtained by replacing the unknown $\boldsymbol{a}$ by an estimator $\hat{\boldsymbol{a}} = (\hat{a}_1, \dots, \hat{a}_p, \hat{a}_0)'$.

There are a number of commonly used estimation procedures for stationary X_t. In this text we consider the maximum likelihood estimators conditioned on $X_0, \dots, X_{1-p}$ (least squares estimators)

$$\hat{\boldsymbol{a}} = \Big(\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}\Big)^{-1}\Big(\sum_{t=1}^{n} \boldsymbol{X}_{t-1}X_t\Big),\ \hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}(X_t - \boldsymbol{X}'_{t-1}\hat{\boldsymbol{a}})^2. \tag{5}$$

We employ slightly different notation when the expectation $\mu = \mathsf{E}\, X_t$ is assumed to be known. Then a_0 is not to be estimated and the model (1) can be written as

$$Y_t = \sum_{j=1}^{p} a_j Y_{t-j} + e_t, t = 1, 2, \dots, n, \tag{6}$$

where $Y_t = X_t - \mu$. The multivariate representation for (6) is $\boldsymbol{Y}_t = \boldsymbol{B}\boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$, where $\boldsymbol{Y}_t = (Y_t, \dots, Y_{t-p+1})'$ and

$$\boldsymbol{B} = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

We define the least squares estimators $\boldsymbol{a}^*$ and σ^{2*} as

$$\boldsymbol{a}^* = \Big(\sum_{t=1}^{n} \boldsymbol{Y}_{t-1}\boldsymbol{Y}'_{t-1}\Big)^{-1}\Big(\sum_{t=1}^{n} \boldsymbol{Y}_{t-1}Y_t\Big),\ \sigma^{2*} = n^{-1}\sum_{t=1}^{n}(Y_t - \boldsymbol{Y}'_{t-1}\boldsymbol{a}^*)^2. \tag{7}$$

The predictor of X_{n+s} associated with $\boldsymbol{a}^*$ is

$$X^*_{n+s} = Y^*_{n+s} + \mu,\ Y^*_{n+s} = \sum_{j=1}^{p} a^*_j Y^*_{n+s-1},\ Y^*_t = X_t - \mu \text{ if } t \le n. \tag{8}$$

Fuller and Hasza (1980, Th. 1) investigated an AR(1) model and showed that the predictor (4) is unbiased for symmetric error distributions. Cryer, Nankervis and Savin (1990, Th. 6) extended their results to predictors based on fitted ARMA(p, q) models with exogenous nonrandom regressors.

Fuller and Hasza (1981, Cor. 2.1.) obtained an approximation for the variance of the predictor error $X_{n+s} - \hat{X}_{n+s}$ through terms of $O(n^{-1})$. They have shown that $\mathsf{E}\{(X_{n+s} - \hat{X}_{n+s})^2\}$ is the upper left element of the matrix

$$\sigma^2 \sum_{j=0}^{s-1} \boldsymbol{A}^j \boldsymbol{M} \boldsymbol{A}'^j + n^{-1}\sigma^2 \sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \boldsymbol{A}^j \boldsymbol{M} \boldsymbol{A}'^k \times \operatorname{Tr}\big\{(\boldsymbol{A}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{A}^{s-k-1})\big\} + O(n^{-3/2}), \tag{9}$$

where $\boldsymbol{\Gamma} = \mathsf{E}\{\boldsymbol{X}_t \boldsymbol{X}_t'\}$ and $\boldsymbol{M}$ is a matrix with one as the upper left element and zeros elsewhere. For $s = 1$ we have $\mathsf{E}\{(X_{n+1} - \hat{X}_{n+1})^2\} = \sigma^2[1 + n^{-1}(p+1)] + O(n^{-3/2})$. In section 2 we evaluate (9) for general prediction period s in some low order autoregressive models.

Since X_{n+s} is a random variable, a predictive region is relevant. Let $V(s) = \sigma^2 \sum_{j=0}^{s-1} w_j^2$, where the $\{w_j\}$ satisfy difference equations

$$w_j - \sum_{i=1}^{p} a_i w_{j-i} = 0, \quad j = 1, 2, \dots$$

subject to the initial conditions $w_0 = 1$ and $w_j = 0$ for $j < 0$. Then for a given $\alpha \in (0,1)$, a natural one-sided $100\alpha\%$ predictive interval for X_{n+s} is

$$\mathrm{PI}_s(\alpha) = [-\infty, \bar{X}_{n+s} + z_\alpha \sqrt{V(s)}], \tag{10}$$

since the conditional distribution of X_{n+s} given $\{X_{1-p}, \dots, X_n\}$ is normal with mean $\bar{X}_{n+s}$ and variance $V(s)$ (Montgomery et all, 1990).

A naive prediction region for X_{n+s} commonly used in textbooks on applied time series analysis as for example Montgomery et. all (1990) is a random set

$$\widehat{\mathrm{PI}}_s(\alpha) = [-\infty, \hat{X}_{n+s} + z_\alpha \sqrt{\hat{V}(s)}] \tag{11}$$

obtained by substituting the estimated parameters into (10). More precisely, $\hat{X}_{n+s}$ is defined in (4) and $\hat{V}(s) = \hat{\sigma}^2 \sum_0^{s-1} \hat{w}_j^2$, where $\hat{w}_j$ satisfy

$$\hat{w}_j - \sum_{i=1}^{p} \hat{a}_i \hat{w}_{j-i} = 0, \quad j = 1, 2, \dots$$

subject to $\hat{w}_0 = 1$ and $\hat{w}_j = 0$ for $j < 0$.When μ is known, we define similarly $\mathrm{PI}_s^*(\alpha)$ as a prediction region for X_{n+s} based on $\boldsymbol{a}^*$.

The overall coverage probability of $\widehat{\mathrm{PI}}_s(\alpha)$, $P[X_{n+s} \in \widehat{\mathrm{PI}}_s(\alpha)]$ is less than α due to the ignored increase in the mean squared error of prediction when employing the estimated autoregressive model for prediction. In section 3 we evaluate the overall coverage probability of the naive one step ahead prediction interval $\mathrm{PI}_1^*(\alpha)$ through terms of $O(n^{-1})$ assuming the variance of e_t is known.

2. Mean squared error of prediction in some low order autoregressive processes

Using similar arguments as in Fuller and Hasza (1981), one can show that when the expectation μ is known, the variance of $X_{n+s} - X_{n+s}^*$ is the upper left element of the matrix

(12)
$$\sigma^2 \sum_{j=0}^{s-1} \boldsymbol{B}^j \boldsymbol{M} \boldsymbol{B}'^j + n^{-1}\sigma^2 \sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \boldsymbol{B}^j \boldsymbol{M} \boldsymbol{B}'^k \times \operatorname{Tr}\bigl\{(\boldsymbol{B}^{s-j-1}\boldsymbol{\Gamma}_Y)'(\boldsymbol{\Gamma}_Y^{-1}\boldsymbol{B}^{s-k-1})\bigr\} + O(n^{-3/2}),$$

where $\boldsymbol{\Gamma}_Y = \mathsf{E}\{\boldsymbol{Y}_t\boldsymbol{Y}_t'\}$ and $\boldsymbol{M}$ is a matrix with one as the upper left element and zeros elsewhere.

2.1. **AR(1) process.** Consider first the AR(1) model with unknown expectation

$$X_t = a_0 + a_1 X_{t-1} + e_t, \quad t = 1, \dots, n, \tag{13}$$

where $e_t \sim N(0, \sigma^2)$, $|a_1| < 1$ and $X_0 \sim N(a_0(1-a_1)^{-1}, \sigma^2(1-a_1^2)^{-1})$. The predictor is $\hat{X}_{n+s} = \hat{a}_0 + \hat{a}_1\hat{X}_{n+s-1}$, where $\hat{X}_{n+s} = X_{n+s}$ for $s \le 0$ and

$$\begin{pmatrix}\hat{a}_1\\ \hat{a}_0\end{pmatrix} = \begin{pmatrix}\sum_{t=1}^n X_{t-1}^2 & \sum_{t=1}^n X_{t-1}\\ \sum_{t=1}^n X_{t-1} & n\end{pmatrix}^{-1} \begin{pmatrix}\sum_{t=1}^n X_{t-1}X_t\\ \sum_{t=1}^n X_t\end{pmatrix}.$$

In this case we have

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_0\\ 0 & 1\end{pmatrix}.$$

Matrix multiplication yields

$$\boldsymbol{A}^j = \begin{pmatrix} a_1^j & a_0\sum_0^{j-1} a_1^i\\ 0 & 1\end{pmatrix}$$

and

$$\boldsymbol{A}^j\boldsymbol{M}\boldsymbol{A}'^k = \begin{pmatrix} a_1^{j+k} & 0\\ 0 & 0\end{pmatrix}. \tag{14}$$

Evaluating moments of X_t up to second order, we find

$$\boldsymbol{\Gamma} = \mathsf{E}\{\boldsymbol{X}_t\boldsymbol{X}_t'\} = \begin{pmatrix} \frac{\sigma^2}{1-a_1^2} + \frac{a_0^2}{(1-a_1)^2} & \frac{a_0}{1-a_1}\\ \frac{a_0}{1-a_1} & 1\end{pmatrix}.$$

Calculating the trace of $(\boldsymbol{A}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{A}^{s-k-1})$, many terms vanish and we have

$$\operatorname{Tr}\{(\boldsymbol{A}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{A}^{s-k-1})\} = 1 + a_1^{2s-j-k-2}. \tag{15}$$

Combining (14) and (15) yields

$$\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \boldsymbol{A}^j\boldsymbol{M}\boldsymbol{A}'^k \operatorname{Tr}\{(\boldsymbol{A}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{A}^{s-k-1})\} = \begin{pmatrix} s^2a_1^{2s-2} + \left(\frac{1-a_1^s}{1-a_1}\right)^2 & 0\\ 0 & 0\end{pmatrix}.$$

Inserting into (9) we have

$$\mathsf{E}\{(X_{n+s} - \hat{X}_{n+s})^2\} = \sigma^2\sum_{j=0}^{s-1} a_1^{2j} + n^{-1}\sigma^2 s^2 a_1^{2s-2} + n^{-1}\sigma^2\Bigl(\frac{1-a_1^s}{1-a_1}\Bigr)^2 + O(n^{-3/2}),$$

which is Theorem 2 of Fuller and Hasza (1980).

When the expectation of X_t is known, $\operatorname{Tr}\{(\boldsymbol{B}^{s-j-1}\boldsymbol{\Gamma}_Y)'(\boldsymbol{\Gamma}_Y^{-1}\boldsymbol{B}^{s-k-1})\}$ simplifies to $a_1^{2s-j-k-2}$ and we find

$$\mathsf{E}\{(X_{n+s} - X_{n+s}^*)^2\} = \sigma^2\sum_{j=0}^{s-1} a_1^{2j} + n^{-1}\sigma^2 s^2 a_1^{2s-2} + O(n^{-3/2}). \tag{16}$$

2.2. **AR(2) process.** Consider now the strictly stationary AR(2) model with zero expectation

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + e_t, \quad t = 1, \dots, n, \tag{17}$$

where $e_t \sim N(0, \sigma^2)$. The predictor is $X^*_{n+s} = a^*_1 X^*_{n+s-1} + a^*_2 X^*_{n+s-2}$, where $X^*_{n+s} = X_{n+s}$ if $s \le 0$ and

$$\begin{pmatrix} a^*_1 \\ a^*_2 \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^n X^2_{t-1} & \sum_{t=1}^n X_{t-1}X_{t-2} \\ \sum_{t=1}^n X_{t-1}X_{t-2} & \sum_{t=1}^n X^2_{t-2} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^n X_{t-1}X_t \\ \sum_{t=1}^n X_{t-2}X_t \end{pmatrix}.$$

The parametr matrix of multivariate representation for (17) is

$$\boldsymbol{B} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix}.$$

Denote the roots of characteristic equation $z^2 - a_1 z - a_2 = 0$ by z_1 and z_2, then $a_1 = z_1 + z_2$ and $a_2 = -z_1 z_2$. Stationarity condition implies that $|z_i| < 1$, $i = 1, 2$. The zero mean AR(1) process is a special case of (17) when $z_2 = 0$.

Denote the rows of $\boldsymbol{B}$ by $\boldsymbol{r}_0 = (1, 0)'$ and $\boldsymbol{r}_1 = (a_1.a_2)'$. Then the rows of $\boldsymbol{B}^j$ are $\boldsymbol{r}_j$ and $\boldsymbol{r}_{j-1}$. They satisfy the relation $\boldsymbol{r}_j = a_1\boldsymbol{r}_{j-1} + a_2\boldsymbol{r}_{j-2}$. Solving these difference equations subject to initial conditions above, we obtain

$$\boldsymbol{B}^j = \frac{1}{z_1 - z_2} \begin{pmatrix} z_1^{j+1} - z_2^{j+1} & z_1 z_2^{j+1} - z_1^{j+1} z_2 \\ z_1^j - z_2^j & z_1 z_2^j - z_1^j z_2 \end{pmatrix}.$$

Matrix multiplication yields

$$\boldsymbol{B}^j \boldsymbol{M} \boldsymbol{B}'^k = \frac{1}{(z_1 - z_2)^2} \begin{pmatrix} (z_1^{j+1} - z_2^{j+1})(z_1^{k+1} - z_2^{k+1}) & (z_1^{j+1} - z_2^{j+1})(z_1^k - z_2^k) \\ (z_1^j - z_2^j)(z_1^{k+1} - z_2^{k+1}) & (z_1^j - z_2^j)(z_1^k - z_2^k) \end{pmatrix}.$$

Using the relation $\sum_{j=0}^{s-1} z^j = (1 - z^s)(1 - z)^{-1}$ if $|z| < 1$, we find that the upper left element of $\sum_0^{s-1} \boldsymbol{B}^j \boldsymbol{M} \boldsymbol{B}'^j$ is

$$\frac{1}{(z_1 - z_2)^2} \{ z_1^2 \frac{1 - z_1^{2s}}{1 - z_1^2} - 2 z_1 z_2 \frac{1 - (z_1 z_2)^s}{1 - z_1 z_2} + z_2^2 \frac{1 - z_2^{2s}}{1 - z_2^2} \}. \tag{18}$$

The matrix of second moments of $\boldsymbol{X}_t$ is

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_Y = \mathsf{E}\{(X_t, X_{t-1})'(X_t, X_{t-1})\} = \gamma_0 \begin{pmatrix} 1 & \frac{z_1+z_2}{1+z_1z_2} \\ \frac{z_1+z_2}{1+z_1z_2} & 1 \end{pmatrix},$$

where $\gamma_0 = \mathsf{E}\, X_t^2$ is the variance of the process. The inverse is

$$\boldsymbol{\Gamma}^{-1} = \gamma_0^{-1} \frac{1 + z_1 z_2}{(1 - z_1^2)(1 - z_2^2)} \begin{pmatrix} 1 + z_1 z_2 & -z_1 - z_2 \\ -z_1 - z_2 & 1 + z_1 z_2 \end{pmatrix}.$$

Multiplicating

$$\boldsymbol{\Gamma}^{-1} \boldsymbol{B}^{s-k-1} = \gamma_0^{-1} \frac{1 + z_1 z_2}{(1 - z_1^2)(1 - z_2^2)(z_1 - z_2)}$$
$$\times \begin{pmatrix} z_1 z_2^{s-k-1}(1 - z_2^2) - z_1^{s-k-1} z_2 (1 - z_1^2) & z_1^{s-k-1} z_2^2 (1 - z_1^2) - z_1^2 z_2^{s-k-1}(1 - z_2^2) \\ z_1^{s-k-1}(1 - z_1^2) - z_2^{s-k-1}(1 - z_2^2) & z_1 z_2^{s-k-1}(1 - z_2^2) - z_1^{s-k-1} z_2 (1 - z_1^2) \end{pmatrix}$$

and

$$\boldsymbol{\Gamma}' \boldsymbol{B}^{s-j-1\prime} = \frac{\gamma_0}{(1 + z_1 z_2)(z_1 - z_2)}$$
$$\times \begin{pmatrix} z_1^{s-j}(1 - z_2^2) - z_2^{s-j}(1 - z_1^2) & z_1^{s-j-1}(1 - z_2^2) - z_2^{s-j-1}(1 - z_1^2) \\ z_1^{s-j+1}(1 - z_2^2) - z_2^{s-j+1}(1 - z_1^2) & z_1^{s-j}(1 - z_2^2) - z_2^{s-j}(1 - z_1^2) \end{pmatrix},$$

we obtain

$$\operatorname{Tr}\{(\boldsymbol{B}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^{s-k-1})\} = \frac{1}{(z_1-z_2)^2}\big\{(1-z_1z_2)^2(z_1^{2s-j-k-2}+z_2^{2s-j-k-2}) - (1-z_1^2)(1-z_2^2)(z_1^{s-j-1}z_2^{s-k-1}+z_1^{s-k-1}z_2^{s-j-1})\big\}.$$

Using relation $\sum_{k=0}^{s-1} z_1^k z_2^{s-1-k} = (z_1^s - z_2^s)(z_1-z_2)^{-1} = K$ say, we find that the upper left element of the matrix

$$\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \boldsymbol{B}^j\boldsymbol{M}\boldsymbol{B}'^k \operatorname{Tr}\{(\boldsymbol{B}^{s-j-1}\boldsymbol{\Gamma})'(\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^{s-k-1})\}$$

is

$$\begin{aligned}&\frac{(1-z_1z_2)^2}{(z_1-z_2)^4}\big\{K^2(z_1^2+z_2^2) - 2Ks(z_1^sz_2+z_1z_2^s) + s^2(z_1^{2s}+z_2^{2s})\big\}\\ &+\frac{(1-z_1^2)(1-z_2^2)}{(z_1-z_2)^4}\big\{2K^2z_1z_2 - 2Ks(z_1^{s+1}+z_2^{s+1}) + 2s^2z_1^sz_2^s\big\}.\end{aligned} \tag{19}$$

Inserting (18) and (19) into (9) we obtain

(20)

$$\begin{aligned}\mathsf{E}\{(X_{n+s}-X_{n+s}^*)^2\} &= \frac{\sigma^2}{(z_1-z_2)^2}\Big\{z_1^2\frac{1-z_1^{2s}}{1-z_1^2} - 2z_1z_2\frac{1-(z_1z_2)^s}{1-z_1z_2} + z_2^2\frac{1-z_2^{2s}}{1-z_2^2}\Big\}\\ &+ n^{-1}\sigma^2\frac{(1-z_1z_2)^2}{(z_1-z_2)^4}\big\{K^2(z_1^2+z_2^2) - 2Ks(z_1^sz_2+z_1z_2^s) + s^2(z_1^{2s}+z_2^{2s})\big\}\\ &+ n^{-1}\sigma^2\frac{(1-z_1^2)(1-z_2^2)}{(z_1-z_2)^4}\big\{2K^2z_1z_2 - 2Ks(z_1^{s+1}+z_2^{s+1}) + 2s^2z_1^sz_2^s\big\}.\end{aligned}$$

When $s=1$ we have $K=1$ and find $\mathsf{E}\{(X_{n+1}-X_{n+1}^*)^2\} = \sigma^2(1+2n^{-1})$, which is the same expression as that obtained by other authors.

2.3. **Application.** As an application of the result (20) we can evaluate the effect of overfitting on the mean squared prediction error. Let $z_2 = a_2 = 0$ in (17) which implies $a_1 = z_1$. Thus we fit the AR(2) model when the true model is AR(1). Then (20) reduces to

(21)

$$\mathsf{E}\{(X_{n+s}-\hat{X}_{n+s})^2\} = \sigma^2\frac{1-a_1^{2s}}{1-a_1^2} + n^{-1}\sigma^2(s-1)^2a_1^{2s-4} + n^{-1}\sigma^2 2sa_1^{2s-2} + O(n^{-3/2}).$$

Since $(s-1)^2+2sa_1^2-s^2a_1^2 = (1-a_1^2)(s-1)^2+a_1^2$ is always positive, we infer from (16) and (21) that overfitting the zero mean AR(1) model by one additional autoregressive parameter results in increase of the mean squared prediction error. The amount of increase can be analytically expressed as $n^{-1}\sigma^2a_1^{2s-4}\{(1-a_1^2)(s-1)^2+a_1^2\}+O(n^{-3/2})$ and tends to zero as the sample size approaches infinity. The one-step ahead mean squared prediction error is $\sigma^2+2n^{-1}\sigma^2$ when fitting overfitted AR(2) model while only $\sigma^2+n^{-1}\sigma^2$ when fitting correct AR(1) model.

3. Coverage probability of naive prediction intervals

Consider the naive s-step ahead prediction interval $\widehat{\mathrm{PI}}_s(\alpha)$ for X_{n+s} defined in (11). There are two kinds of coverage probabilities:

1. For fixed sample information $\boldsymbol{X} = (X_{1-p}, \dots, X_n)'$ (and thus fixed $\boldsymbol{X}_n$, $\hat{\boldsymbol{a}}$, $\hat{\sigma}^2$ and $\widehat{\mathrm{PI}}_s(\alpha)$) is the conditional probability

$$\begin{aligned}\mathrm{CP}[\widehat{\mathrm{PI}}_s(\alpha)|\hat{\boldsymbol{a}}, \hat{\sigma}^2, \boldsymbol{X}_n] &= P[X_{n+s} \in \widehat{\mathrm{PI}}_s(\alpha)|\hat{\boldsymbol{a}}, \hat{\sigma}^2, \boldsymbol{X}_n] \\ &= \Phi\Big(z_\alpha \sqrt{V(s)^{-1}\hat{V}(s)} + V(s)^{-1}(\hat{X}_{n+s} - \bar{X}_{n+s})\Big),\end{aligned}$$

because the conditional distribution of X_{n+s} is normal with mean $\bar{X}_{n+s}$ and variance $V(s)$. Here $\Phi(t)$ denotes the distribution function of standard normal distribution and $z_\alpha = \Phi(\alpha)^{-1}$ its α-quantile.
2. From sample to sample, the conditional coverage probability is random because $\widehat{\mathrm{PI}}_s(\alpha)$ depends on $\hat{\boldsymbol{a}}$ and $\boldsymbol{X}_n$. The unconditional (overall) coverage probability for the prediction interval procedure is

$$\mathrm{UCP}[\widehat{\mathrm{PI}}_s(\alpha)] = P[X_{n+s} \le \hat{X}_{n+s} + z_\alpha\sqrt{\hat{V}(s)}] = \mathsf{E}\,\{\mathrm{CP}[\widehat{\mathrm{PI}}_s(\alpha)|\hat{\boldsymbol{a}}, \hat{\sigma}^2, \boldsymbol{X}_n]\},$$

where the expectation is w.r.t. the random $\hat{\boldsymbol{a}}, \hat{\sigma}^2$ and $\boldsymbol{X}_n$.

Beran (1990) has shown in his Example 1 that $\mathrm{UCP}[\mathrm{PI}_1^*(\alpha)] = \alpha-(2n)^{-1}z_\alpha\phi(z_\alpha)+o(n^{-1})$ for the AR(1) process with known mean and $\sigma^2 = 1$ also known. We extend this result to general order and give the order of error. We have

Theorem 1. *Assume that $\{X_t\}_{t=1-p}^n$ is strictly stationary AR(p) process defined in* (1)*, where $\mathsf{var}\, e_t = \sigma^2$, the order p and expectation μ are known. Let the parameters $\boldsymbol{a} = (a_1, \dots, a_p)'$ be estimated by $\boldsymbol{a}^*$ defined in* (7)*. Let X_{n+1}^* be defined as $\mu+\sum_{j=1}^p a_j^*(X_{n+j-1}-\mu)$. Then the overall coverage probability of the naive one-step ahead prediction interval $\mathrm{PI}_1^*(\alpha) = [-\infty, X_{n+1}^* + z_\alpha\sigma]$ is*

$$\mathrm{UCP}[\mathrm{PI}_1^*(\alpha)] = \alpha - (2n)^{-1}pz_\alpha\phi(z_\alpha) + O(n^{-3/2}),$$

where $\phi(t) = (2\pi)^{-1/2}\exp(-t^2/2)$ is the density of $N(0,1)$.

Proof. Without loss of generality assume $\mu = 0$. Since σ^2 is known, $V(1)^* = V(1) = \sigma^2$ and $\mathrm{PI}_1^*(\alpha) = (-\infty, X_{n+1}^* + z_\alpha\sigma)$.
The conditional coverage probability is $\mathrm{CP}(\mathrm{PI}_1^*(\alpha)|\boldsymbol{a}^*, \boldsymbol{X}_n) = \Phi(z_\alpha + \delta_n)$, where $\delta_n = \sigma^{-1}(X_{n+1}^* - \bar{X}_{n+1})$.
The distribution function of Gaussian distribution has continuous derivatives of all orders, thus the Taylor expansion yields

$$\mathrm{CP}(\mathrm{PI}_1^*(\alpha)|\boldsymbol{a}^*, \boldsymbol{X}_n) = \alpha + \delta_n\phi(z_\alpha) + \frac{\delta_n^2}{2}\phi'(z_\alpha) + \frac{\delta_n^3}{6}\phi''(z_n), \tag{22}$$

where z_n is random variable between z_α and $z_\alpha + \delta_n$.

Since both X_{n+1}^* and $\bar{X}_{n+1}$ are unbiased predictors for X_{n+1} (Fuller and Hasza, 1980, Cryer et all, 1990), we have

$$\mathsf{E}\,\delta_n = 0. \tag{23}$$

Rewrite

$$\sigma^2\delta_n^2 = (X_{n+1}^* - \bar{X}_{n+1})^2 = \Big[\sum_{j=1}^p (a_j^* - a_j)X_{n-j+1}\Big]^2 = \boldsymbol{X}_n'(\boldsymbol{a}^* - \boldsymbol{a})(\boldsymbol{a}^* - \boldsymbol{a})'\boldsymbol{X}_n.$$

Following Fuller and Hasza (1981), the conditional expectation is

$$\sigma^2\mathsf{E}\,\{\delta_n^2|\boldsymbol{X}_n\} = n^{-1}\sigma^2\boldsymbol{X}_n'\Gamma^{-1}\boldsymbol{X}_n + O(n^{-3/2}),$$

where $\Gamma = \mathsf{E}\,\{\boldsymbol{X}_t\boldsymbol{X}_t'\}$ as in section 2. Using formula for the expectation of quadratic form we find

$$\begin{aligned} \mathsf{E}\,\delta_n^2 &= n^{-1}\mathsf{E}\left\{\boldsymbol{X}_n'\Gamma^{-1}\boldsymbol{X}_n\right\} + O(n^{-3/2}) \\ &= n^{-1}\left[\mathsf{E}\left\{\mathrm{Tr}(\Gamma^{-1}\mathsf{var}\,\boldsymbol{X}_n)\right\} + \mathsf{E}\,\boldsymbol{X}_n'\Gamma^{-1}\mathsf{E}\,\boldsymbol{X}_n\right] + O(n^{-3/2}) \\ &= n^{-1}p + O(n^{-3/2}). \end{aligned} \tag{24}$$

Since $\phi(t)''$ is product of a polynomial and $\exp(-t^2/2)$, it is bounded, thus $|\phi(z_n)''| \le M_1$ for some $M_1 > 0$. Rewrite

$$\sigma^3\delta_n^3 = \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(a_j^* - a_j)(a_k^* - a_k)(a_l^* - a_l)X_{n-j+1}X_{n-k+1}X_{n-l+1}.$$

We have

$$\sigma^3\mathsf{E}\,|\delta_n^3| \le \sum_{j,k,l}\mathsf{E}\left\{|a_j^* - a_j|\dots|X_{n-j+1}|\dots\right\} \le \sum_{j,k,l}\sqrt[6]{\mathsf{E}\left\{|a_j^* - a_j|^6\right\}\dots\mathsf{E}\left\{|X_{n-j+1}|^6\right\}\dots}$$

from Holder inequality. Now $M_2 = E\{|X_t|^6\}$ is finite because X_t is Gaussian, thus

$$E|\delta_n^3| \le \sigma^{-3}\sqrt{M_2}\sum_{j,k,l}\sqrt[6]{\mathsf{E}\left\{|a_j^* - a_j|^6\right\}\mathsf{E}\left\{|a_k^* - a_k|^6\right\}\mathsf{E}\left\{|a_l^* - a_l|^6\right\}}.$$

Following Bhansali and Papangelou (1991), $\mathsf{E}\left\{|a_j^* - a_j|^6\right\} = O(n^{-3})$ and we find $\mathsf{E}\,|\delta_n^3| = O(n^{-3/2})$. Finally

$$\mathsf{E}\left\{|\delta_n^3\phi(z_n)''|\right\} \le M_1\mathsf{E}\,|\delta_n^3| = O(n^{-3/2}). \tag{25}$$

Combining (23), (24), (25) and $\phi(t)' = -t\phi(t)$ and inserting into (22) we obtain

$$\mathsf{E}\left\{\mathrm{CP}(\mathrm{PI}_1^*(\alpha)|\boldsymbol{a}^*, \boldsymbol{X}_n)\right\} = \alpha - (2n)^{-1}pz_\alpha\phi(z_\alpha) + O(n^{-3/2}),$$

which was to be shown. Q.E.D.

We could extent our theorem to the model with unknown mean provided the result of Bhansali and Papangelou (1991),

$$\mathsf{E}\left\{|\hat{a}_j - a_j|^k\right\} = O(n^{-k/2}),\ 0,\dots,p$$

is valid for such model. Unfortunately as far as we know, there is no such generalization in the literature. The formula for the more general case, not proven yet would be

$$\mathrm{UCP}[\widehat{\mathrm{PI}}_1(\alpha)] = \alpha - (2n)^{-1}(p+1)z_\alpha\phi(z_\alpha).$$

References

[1] Beran, R. (1990). Calibrating prediction regions. *Journal of the American Statistical Association* **85**, 715–723.

[2] Bhansali, R.J. and Papangelou, F. (1991). Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *The Annals of Statistics*, **19**, 1155–1162.

[3] Cryer, J.D., Nankervis, J.C. and Savin, N.E. (1990). Forecast error symmetry in ARIMA models. *Journal of the American Statistical Association*, **85**, 724–728.

[4] Fuller, W.A. and Hasza, D.P. (1980). Predictors for the first-order autoregressive process. *Journal of Econometrics*, **13**, 139–157.

[5] Fuller, W.A. and Hasza, D.P. (1981). Properties of predictors for autoregressive time series. *Journal of the American Statistical Association*, **76**, 155–161.

[6] Montgomery, D.C., Johnson, L.A. and Gardiner, J.S. (1990). *Forecasting & Time Series Analysis.* New York: McGraw-Hill.

UK MFF, KPMS, Sokolovsk 83, 186 75 Praha
E-mail address: zvarap@karlin.mff.cuni.cz