Prediction Intervals for Some ARIMA Time Series

Mulubrhan G. Haile and David J. Olive *
Southern Illinois University

September 20, 2022

Abstract

Prediction intervals and regions for ARIMA time series models are given, including prediction intervals after model selection. When consistent estimators are used, the forecast residuals are consistent estimators of the forecast errors. Find a prediction interval for a future forecast error, then shift the interval to be centered at the point estimator of the h-step ahead forecast.

KEY WORDS: model selection.

1 Introduction

This paper suggests prediction intervals for a wide variety of time series models. This section reviews some time series models and model selection. A *time series* $\{Y_t\} = Y_1, ..., Y_n$ consists of observations Y_t collected sequentially at times $1, ..., n$. Many time series models have the form

$$Y_t = \tau + \sum_i \psi_i Y_{t-ik_i} + \sum_j \nu_j e_{t-jk_j} + e_t \tag{1}$$

where the errors $\{e_t\}$ are independent and identically distributed (iid) unobserved random variables. Unless stated otherwise, assume the mean $E(e_t) = 0$ and the variance $V(e_t) = \sigma^2 = \sigma_e^2$. For example, the Box, Jenkins, and Reinsel (1994) multiplicative seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ time series models have this form.

To describe several important time series models, we will use the R software notation and write a moving average parameter θ with a positive sign. Many references and software will write the model with a negative sign for the moving average parameters. A *moving average* MA(q) times series is

$$Y_t = \tau + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t$$

*Mulubrhan G. Haile is Visiting Assistant Professor, Westminster College, Fulton, MO, and David J. Olive is Professor, School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, IL 62901-4408 (E-mail: *dolive@siu.edu*).

where $\theta_q \neq 0$. An *autoregressive* AR(p) times series is

$$Y_t = \tau + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

where $\phi_p \neq 0$. An *autoregressive moving average* ARMA(p, q) times series is

$$Y_t = \tau + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{2}$$

where $\theta_q \neq 0$ and $\phi_p \neq 0$.

We usually want the ARMA(p, q) model to be weakly stationary, causal, and invertible. Let $Z_t = Y_t - \mu$ where $\mu = E(Y_t)$ if $\{Y_t\}$ is weakly stationary. Then the causal property implies that $Z_t = \sum_{j=1}^{\infty} \psi_j e_{t-j} + e_t$, which is an MA($\infty$) representation, where the $\psi_j \to 0$ rapidly as $j \to \infty$. Invertibility implies that $Z_t = \sum_{j=1}^{\infty} \chi_j Z_{t-j} + e_t$, which is an AR($\infty$) representation, where the $\chi_j \to 0$ rapidly as $j \to \infty$. Thus if the ARMA(p, q) model is weakly stationary, causal, and invertible, then Y_t depends almost entirely on nearby lags of Y_t and e_t, not on the distant past. Also, the time series model $\approx$ AR(p_y) $\approx$ MA(q_y) for some positive integers p_y and q_y that do not depend on the sample size n.

To describe ARIMA models, let the difference operator $\bigtriangledown = (1 - B)$ where the backshift operator or lag operator B satisfies $BW_t = W_{t-1}$ and $B^j W_t = W_{t-j}$. Let $X_t = \bigtriangledown^d Y_t = (1 - B)^d Y_t$ be the differenced time series. The first difference is $X_t = \bigtriangledown Y_t = (1 - B) Y_t = Y_t - Y_{t-1}$. The second difference is $X_t = \bigtriangledown^2 Y_t = \bigtriangledown(\bigtriangledown Y_t) = Y_t - 2Y_{t-1} + Y_{t-2}$. If Y_t follows an ARIMA(p, d, q) model, want X_t to follow a weakly stationary and invertible ARMA(p, q) = ARIMA($p, 0, q$) model. Typically $d = 0$ or 1, but occasionally $d = 2$. Usually $\tau = 0$ if $d > 1$. The ARIMA($p, d = 1, q$) model is $Y_t = \tau + (1+\phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + \cdots + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t$.

This paper derives prediction intervals for the above time series models, and another topic in this paper is prediction intervals after time series model selection. For ARMA model selection, let the full model be an ARMA(p_{max}, q_{max}) model. For AR model selection $q_{max} = 0$, while for MA model selection $p_{max} = 0$. For nonseasonal time series, Granger and Newbold (1977, p. 178) suggested using $p_{max} = 13$ for AR model selection, and we may use $p_{max} = q_{max} = 5$ for ARMA model selection, and $q_{max} = 13$ for MA model selection. For ARMA model selection, there are $J = (p_{max} + 1)(q_{max} + 1)$ ARMA(p, q) submodels where p ranges from 0 to p_{max} and q ranges from 0 to q_{max}. For AR and MA model selection there are $J = p_{max} + 1$ and $J = q_{max} + 1$ submodels, respectively. Assume the true (optimal) model is an ARMA(p_S, q_S) model with $p_S \leq p_{max}$ and $q_S \leq q_{max}$. Let the selected model I be an ARMA(p_I, q_I) model. Then the model underfits unless $p_I \geq p_S$ and $q_I \geq q_S$. For AR model selection, the probability of underfitting goes to 0 if the Akaike (1973) AIC, Schwartz (1978) BIC, or Hurvich and Tsai (1989) AIC_C criterion are used, at least if the e_t are iid $N(0, \sigma^2)$. Also see Claeskens and Hjort (2008, pp. 39, 40, 45, 46), Hannan and Quinn (1979), and Shibata (1976).

More notation is needed for model selection. Let the full model be the AR(p_{max}), MA(q_{max}), or ARMA(p_{max}, q_{max}) model. Let $\boldsymbol{\beta}$ be a $b \times 1$ vector. For ARMA model selection, let $\boldsymbol{\beta} = (\boldsymbol{\phi}^T, \boldsymbol{\theta}^T)^T = (\phi_1, ..., \phi_{p_{max}}, \theta_1, ..., \theta_{q_{max}})^T$ with $b = p_{max} + q_{max}$. For AR model selection, let $\boldsymbol{\beta} = (\phi_1, ..., \phi_{p_{max}})^T$ with $b = p_{max}$, and for MA model selection, let $\boldsymbol{\beta} = (\theta_1, ..., \theta_{q_{max}})^T$ with $b = q_{max}$. Hence $\boldsymbol{\beta} = (\beta_1, ..., \beta_{p_{max}}, \beta_{p_{max}+1}, ..., \beta_{p_{max}+q_{max}})^T$. Let $S = \{1, ..., p_S, p_{max} + 1, ..., p_{max} + q_S\}$ index the true ARMA(p_S, q_S) model. If

$S = \emptyset$ is the empty set, then the time series random variables $Y_1, ..., Y_n$ are iid. Let $I = \{1, ..., p_I, p_{max}+1, ..., p_{max}+q_I\}$ index the ARMA(p_I, q_I) model. Let $\hat{\boldsymbol{\beta}}_{I,0}$ be a $b \times 1$ estimator of $\boldsymbol{\beta}$ which is a obtained by padding $\hat{\boldsymbol{\beta}}_I$ with zeroes. If $\boldsymbol{\beta}_I = (\phi_1, ..., \phi_{p_I}, \theta_1, ..., \theta_{q_I})^T$, then $\hat{\boldsymbol{\beta}}_{I,0} = (\hat{\phi}_1, ..., \hat{\phi}_{p_I}, 0, .., 0, \hat{\theta}_1, ..., \hat{\theta}_{q_I}, 0, ..., 0)^T$. If $q_I = 0$, then $\hat{\boldsymbol{\beta}}_{I,0} = (\hat{\phi}_1, ..., \hat{\phi}_{p_I}, 0, .., 0)^T$. If $p_I = 0$ then $\hat{\boldsymbol{\beta}}_{I,0} = (0, ..., .., 0, \hat{\theta}_1, ..., \hat{\theta}_{q_I}, 0, ..., 0)^T$. If $I = \emptyset$ with $p_I = q_I = 0$, then define $\hat{\boldsymbol{\beta}}_{I,0} = \mathbf{0}$, the $b \times 1$ vector of zeroes. The submodel I underfits unless $S \subseteq I$.

For example, if $p_{max} = q_{max} = 5$, then $S = \{1, 6, 7\}$ corresponds to the ARMA(1,2) model, and $I = \{1, 6, 7, 8\}$ corresponds to the ARMA(1,3) model. Then $\hat{\boldsymbol{\beta}}_S = (\hat{\phi}_1, \hat{\theta}_1, \hat{\theta}_2)^T$, $\hat{\boldsymbol{\beta}}_{S,0} = (\hat{\phi}_1, 0, 0, 0, 0, \hat{\theta}_1, \hat{\theta}_2, 0, 0, 0)^T$, and $\hat{\boldsymbol{\beta}}_{I,0} = (\hat{\phi}_1, 0, 0, 0, 0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, 0, 0)^T$.

The model I_{min} corresponds to the model that minimizes the AIC, AIC_C, or BIC criterion. Then the model selection estimator $\hat{\boldsymbol{\beta}}_{MS} = \hat{\boldsymbol{\beta}}_{I_{min},0}$. Assume $\hat{\boldsymbol{\beta}}_{MS} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities $\pi_{kn} = P(I_{min} = I_k)$ for $k = 1, ..., J$. Haile and Olive (2022) gave the large sample theory for $\hat{\boldsymbol{\beta}}_{MS}$, and used bootstrap confidence regions for hypothesis testing.

Section 2 illustrates h-step ahead prediction intervals with ARIMA models. Section 3 illustrates prediction regions with random walk models. Section 4 gives some examples and simulations.

2 Prediction Intervals

For forecasting, predict the test data $Y_{n+1}, ..., Y_{n+L}$ given the past training data $Y_1, ..., Y_n$. A large sample $100(1-\delta)\%$ prediction interval (PI) for Y_{n+h} is $[L_n, U_n]$ where the coverage $P(L_n \le Y_{n+h} \le U_n) = 1 - \delta_n$ is eventually bounded below by $1 - \delta$ as $n \to \infty$. We often want $1 - \delta_n \to 1 - \delta$ as $n \to \infty$. By construction, some of the prediction intervals will have training data coverage $\approx 1 - \alpha_n$ where $1 - \alpha_n \ge 1 - \delta$, and $1 - \alpha_n \to 1 - \delta$ as $n \to \infty$. A large sample $100(1-\delta)\%$ PI is asymptotically optimal if it has the shortest asymptotic length: the length of $[L_n, U_n]$ converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the population shorth: the shortest interval covering at least $100(1-\delta)\%$ of the mass.

The shorth estimator of the population shorth will be defined below and used to create large sample PIs that do not require knowing the distribution of the errors e_t. If the data are $Z_1, ..., Z_n$, let $Z_{(1)} \le \cdots \le Z_{(n)}$ be the order statistics. Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x (e.g., $\lceil 7.7 \rceil = 8$). Consider intervals that contain c cases $[Z_{(1)}, Z_{(c)}], [Z_{(2)}, Z_{(c+1)}], ..., [Z_{(n-c+1)}, Z_{(n)}]$. Compute $Z_{(c)} - Z_{(1)}, Z_{(c+1)} - Z_{(2)}, ..., Z_{(n)} - Z_{(n-c+1)}$. Then the estimator shorth(c) $= [Z_{(s)}, Z_{(s+c-1)}]$ is the interval with the shortest length.

Suppose the data $Z_1, ..., Z_n$ are iid and a large sample $100(1-\delta)\%$ PI is desired for a future value Z_f such that $P(Z_f \in [L_n, U_n]) \to 1 - \delta$ as $n \to \infty$. The shorth(c) interval is a large sample $100(1-\delta)\%$ PI if $c/n \to 1 - \delta$ as $n \to \infty$, that often has the asymptotically shortest length. Frey (2013) showed that for large $n\delta$ and iid data, the shorth($k_n = \lceil n(1-\delta) \rceil$) prediction interval has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$, and used the large sample $100(1-\delta)\%$ PI shorth(c) =

$$[Z_{(s)}, Z_{(s+c-1)}] \quad \text{with} \quad \mathrm{c} = \min(\mathrm{n}, \lceil \mathrm{n}[1 - \delta + 1.12\sqrt{\delta/\mathrm{n}}\ ]\ \rceil). \tag{3}$$

Some more notation is needed before deriving PIs for time series. Suppose the training data set is $Y_1, ..., Y_t$. The h-step ahead forecast for a future value Y_{t+h} is $\hat{Y}_t(h)$ and the h-step ahead forecast residual is $\hat{e}_t(h) = Y_{t+h} - \hat{Y}_t(h)$. For example, a common choice for model (1) is

$$\hat{Y}_t(h) = \hat{\tau} + \sum_i \hat{\psi}_i Y^*_{t+h-ik_i} + \sum_j \hat{\nu}_j \hat{e}^*_{t+h-jk_j}$$

where $\hat{e}_t$ is the tth residual, $Y^*_{t+h-ik_i} = Y_{t+h-ik_i}$ if $h - ik_i \leq 0$, $Y^*_{t+h-ik_i} = \hat{Y}_t(h - ik_i)$ if $h - ik_i > 0$, $\hat{e}^*_{t+h-jk_j} = \hat{e}_{t+h-jk_j}$ if $h - jk_j \leq 0$, and $\hat{e}^*_{t+h-jk_j} = 0$ if $h - jk_j > 0$, and the forecasts $\hat{Y}_t(1), \hat{Y}_t(2), ..., \hat{Y}_t(L)$ are found recursively if there is data $Y_1, ..., Y_t$. Typically the residuals $\hat{e}_t = \hat{e}_{t-1}(1)$ are the 1-step ahead forecast residuals and the fitted or predicted values $\hat{Y}_t = \hat{Y}_{t-1}(1)$ are the 1-step ahead forecasts.

Example 1 is useful to illustrate the forecasts. The R software produces $\hat{e}_t$ and $\hat{Y}_t = Y_t - \hat{e}_t$ for $t = m+1, ..., m+n_1$ where there are n_1 1-step ahead forecast residuals $\hat{e}_t = \hat{e}_{t-1}(1)$ available, often with $m = 0$ and $n_1 = n$. In the examples, we get the formulas $\hat{Y}_n(h)$, and then replace n by t so that the test data formula is applied to the training data. Then the general formula for an ARMA(p, q) model is $\hat{Y}_t(h) = \hat{\tau} + \hat{\phi}_1 \hat{Y}_t(h-1) + \hat{\phi}_2 \hat{Y}_t(h-2) + \cdots + \hat{\phi}_{h-1} \hat{Y}_t(1) + \hat{\phi}_h Y_t + \cdots + \hat{\phi}_p Y_{t+h-p} + \hat{\theta}_h \hat{e}_t + \cdots + \hat{\theta}_q \hat{e}_{t+h-q}$ for $1 < h \leq min(p, q)$. Assume there are n_h forecast residuals $\hat{e}_t(h)$ available from the training data.

Example 1. Suppose the training data is $Y_1, ..., Y_n$. a) Consider an MA(2) model: $Y_t = \tau + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t$. The R software produces $\hat{e}_t$ and $\hat{Y}_t = Y_t - \hat{e}_t$ for $t = 1, ..., n$ where $\hat{Y}_t = \hat{Y}_{t-1}(1) = \hat{\tau} + \hat{\theta}_1 \hat{e}_{t-1} + \hat{\theta}_2 \hat{e}_{t-2}$ and $\hat{e}_t(1) = Y_{t+1} - \hat{Y}_t(1)$ for $t = 3, ..., n$. Also, $\hat{Y}_n(1) = \hat{\tau} + \hat{\theta}_1 \hat{e}_n + \hat{\theta}_2 \hat{e}_{n-1}$. Hence there are $n_1 = n$ 1-step ahead forecast residuals $\hat{e}_t = \hat{e}_{t-1}(1)$ available. Similarly, $\hat{Y}_t(2) = \hat{\tau} + \hat{\theta}_2 \hat{e}_t$ for $t = 1, ..., n$. Hence the 2-step ahead forecast residuals are available for $t = 3, ..., n-2$. Now $\hat{Y}_t(h) = \hat{\tau} \approx \overline{Y}$ for $h > 2$. Hence there are n h-step ahead forecast residuals $Y_t - \overline{Y}$ for $h > 2$ and $t = 1, ..., n$.
b) Consider an ARMA(1,1) model: $Y_t = \tau + \phi_1 Y_{t-1} + \theta_1 e_{t-1} + e_t$. For $h = 1$, $\hat{Y}_t(1) = \hat{\tau} + \hat{\phi}_1 Y_t + \hat{\theta}_1 \hat{e}_t$. For $h > 1$, $\hat{Y}_t(h) = \hat{\tau} + \hat{\phi}_1 \hat{Y}_t(h-1)$.
c) Consider an AR(1) model: $Y_t = \tau + \phi_1 Y_{t-1} + e_t$. For $h = 1$, $\hat{Y}_t(1) = \hat{\tau} + \hat{\phi}_1 Y_t$. If $\hat{Y}_t(0) = Y_t$, then $\hat{Y}_t(h) = \hat{\tau} + \hat{\phi}_1 \hat{Y}_t(h-1) = \hat{\tau}(1 + \hat{\phi}_1 + \cdots + \hat{\phi}_1^{h-1}) + \hat{\phi}_1^h Y_t = \dfrac{1 - \hat{\phi}_1^h}{1 - \hat{\phi}_1} \hat{\tau} + \hat{\phi}_1^h Y_t$.
For a weakly stationary AR(1) time series, a good estimation method will have $|\hat{\phi}_1| < 1$.
d) Consider an ARIMA(1,1,1) model with $\tau = 0$: $Y_t = (1 + \phi_1) Y_{t-1} - \phi_1 Y_{t-2} + \theta_1 e_{t-1} + e_t$. Then $\hat{Y}_t(1) = (1 + \hat{\phi}_1) Y_t - \hat{\phi}_1 Y_{t-1} + \hat{\theta}_1 \hat{e}_t$, $\hat{Y}_t(2) = (1 + \hat{\phi}_1) \hat{Y}_t(1) - \hat{\phi}_1 Y_t$, and $\hat{Y}_t(h) = (1 + \hat{\phi}_1) \hat{Y}_t(h-1) - \hat{\phi}_1 \hat{Y}_t(h-2)$ for $h > 2$.
e) Consider an ARIMA(0,1,1) model with $\tau = 0$: $Y_t = Y_{t-1} + \theta_1 e_{t-1} + e_t$. Then $\hat{Y}_t(1) = Y_t + \hat{\theta}_1 \hat{e}_t$, and $\hat{Y}_t(h) = \hat{Y}_t(h-1) = \hat{Y}_t(1)$ for $h \geq 2$.
f) Consider an ARIMA(0,2,2) model with $\tau = 0$: $Y_t = 2Y_{t-1} - Y_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t$. Then $\hat{Y}_t(1) = 2Y_t - Y_{t-1} + \hat{\theta}_1 \hat{e}_t + \hat{\theta}_2 \hat{e}_{t-1}$, $\hat{Y}_t(2) = 2\hat{Y}_t(1) - Y_t + \hat{\theta}_2 \hat{e}_t$, and $\hat{Y}_t(h) = 2\hat{Y}_t(h-1) - \hat{Y}_t(h-2)$ for $h \geq 3$.

The basic idea for getting prediction intervals for the test data is now given. Find the forecast formulas for the test data $Y_{n+1}, ..., Y_{n+L}$, and apply the formulas to the training data $Y_1, ..., Y_n$ to get forecast residuals. Assume consistent estimators are used so that the forecast residuals are consistent estimators of the forecast errors. Apply the shorth to the n_h forecast residuals $\hat{e}_t(h)$ to get $[L_n(h), U_n(h)]$, a PI for a future forecast error. Then the PI for Y_{n+h} is $[\hat{Y}_n(h) + L_n(h), \hat{Y}_n(h) + U_n(h)]$. Since the forecast residuals tend to underestimate the forecast errors, small correction factors are needed for small n. This idea is illustrated for ARIMA models, but also works for many other time series methods, including seasonal ARIMA models. Similar PIs and prediction regions were derived for multiple linear regression, nonlinear models of the form $Y = m(\boldsymbol{x}) + e$, and multivariate linear regression by Olive (2007, 2013, 2017ab, 2018).

Often time series PIs assume normality, and do not work well unless the errors e_t are iid $N(0, \sigma_e^2)$. For many time series models, a large sample normal $100(1-\delta)\%$ PI for Y_{t+h} is

$$[L_n, U_n] = \hat{Y}_t(h) \pm t_{1-\delta/2, n-p-q} SE(\hat{Y}_t(h)). \tag{4}$$

Suppose that as $n \to \infty$, $\hat{Y}_t(h) \xrightarrow{P} E(Y_{t+h}) = \mu_{t+h}$ and $SE(\hat{Y}_t(h)) \xrightarrow{P} SD(Y_{t+h}) = \sigma_{t+h}$. Thus $\hat{Y}_t(h)$ and $SE(\hat{Y}_t(h))$ are consistent estimators of μ_{t+h} and σ_{t+h}, respectively. These quantities are conditional on the past, but the conditioning is suppressed. Then
$P(Y_{t+h} \in [L_n, U_n]) \approx P(Y_{t+h} \in [\mu_{t+h} - z_{1-\delta/2}\sigma_{t+h}, \mu_{t+h} + z_{1-\delta/2}\sigma_{t+h}]) =$
$P[|Y_{t+h} - \mu_{t+h}| < z_{1-\delta/2}\sigma_{t+h}]$ "$\geq$" $1 - \frac{1}{z^2_{1-\delta/2}}$ assuming Chebyshev's inequality holds to a good approximation. Hence a 95% PI could have coverage as low as 74% and a 99.7% PI could have coverage as low as 89%. If n is large, a nominal 95% PI uses $t_{1-\delta/2, n-p-q} \approx 1.96$ while using $z_{1-\delta/2} = 5$ has coverage that is eventually bounded below by 96% as $n \to \infty$. The t cutoff 1.96 tends to be too low while the Chebyshev cutoff 5 tends to be too high in that the PI length will be too long and the coverage too high.

The following new PI ignores the time series structure of the data. Let $\overline{e}_t = Y_t - \overline{Y}$, and let shorth$(c_1 = \lceil n(1-\delta) \rceil) = [L_n(h), U_n(h)]$ be computed from the $\overline{e}_t$. Then the large sample $100(1-\delta)\%$ shorth(c_1) PI for Y_{t+h} is

$$[L_n, U_n] = [\overline{Y} + b_n L_n(h), \overline{Y} + b_n U_n(h)] \tag{5}$$

where $b_n = \left(1 + \frac{15}{n}\right)\sqrt{\frac{n+1}{n-1}}$. Note that this PI is the same for all h. For weakly stationary, causal, and invertible ARMA(p, q) models, this PI is too long for h near 1, but should have short length for large h and if $h > q$ for an MA(q) model. This PI is the Olive (2013) PI suggested for Y_f when $Y_1, ..., Y_n$ and Y_f are iid.

PI (5) works for two reasons. First, a weakly stationary, causal ARMA(p, q) time series follows an MA(∞) model which is approximately an MA(q_y) time series where q_y depends on the time series but not on n. Such time series tend to be ergodic: see White (1984, p. 46). For ergodic data from a unimodal distribution, Chen and Shao (1999) proved the sample shorth converges to the unique population shorth. Second, we show that PI (5) works for MA(q) models. Thus PI (5) will also work for MA(∞) time series models. For the MA(q) model, $e_t(h) = \theta_1 e_{t+h-1} + \theta_2 e_{t+h-2} + \cdots + \theta_{h-1} e_{t+1} + e_{t+h}$ for $h \leq q$, $e_t(h) = Y_{t+h} - \mu$ for $h > q$, the $e_t(h)$ are identically distributed for fixed h, and the random variables $e_j(h), e_{j+h}(h), e_{j+2(h)}(h), ...$ are iid for fixed $h \leq q$. For $h \leq q$, there are h iid sequences starting at $j = 1, 2, ..., h$, respectively. For $h > q$ there are $q + 1$ iid sequences starting at $j = 1, ..., (q + 1)$. Since the sample percentiles of the iid sequences converge in probability to the population percentiles for fixed h, so do the sample percentiles of all of the data. Hence $P(e_t(h) \in [L_n(h), U_n(h)]) \approx 1 - \delta$ as $n \to \infty$ for the MA(q) model if consistent estimators are used.

One step ahead PIs for differenced time series are simple using Equation (5). If $Y_{t+1} = X_{t+1} - X_t$, then a nonparametric one step ahead PI for X_{t+1} is

$$[X_t + L_n, X_t + U_n] = [X_t + \overline{Y} + b_n L_n(h), X_t + \overline{Y} + b_n U_n(h)].$$

Here $X_{t+1} = X_t + Y_{t+1}$ is a random walk where $e_t = Y_t$ has an $MA(\infty)$ representation. See Section 3 for a random walk where the e_t are iid. If $Y_{t+1} = X_{t+1} - 2X_{t-1} + X_{t-1}$, then a nonparametric one ahead PI for X_{t+1} is

$$[2X_t - X_{t-1} + L_n, 2X_t - X_{t-1} + U_n].$$

The following PI is new and takes into account the time series structure of the data. A similar idea in Masters (1995, p. 305) is to find the n_h h-step ahead forecast residuals and use percentiles to make PIs for Y_{t+h} for $h = 1, ..., L$. For ARIMA(p, d, q) models, let $c_2 = \lceil n_h(1 - \delta_n) \rceil$ and compute shorth(c_2) $= [L_n(h), U_n(h)]$ of the h-step ahead forecast residuals $\hat{e}_t(h)$. Let $a_h = \left(1 + \frac{15}{n_h}\right) \sqrt{\frac{n_h}{n_h - p - q}}$. Then a large sample $100(1 - \delta)\%$ PI for Y_{t+h} is

$$[L_n, U_n] = [\hat{Y}_n(h) + a_h L_n(h), \hat{Y}_n(h) + a_h U_n(h)] \tag{6}$$

where $1 - \delta_n = \min(1 - \delta + 0.05, 1 - \delta + (p + q)/n_h)$ for $\delta > 0.1$ and $1 - \delta_n = \min(1 - \delta/2, 1 - \delta + 10(p + q)\delta/n_h)$ for $\delta \le 0.1$. The correction factor helps compensate for undercoverage when $n_h \ge 20(p + q)$, and similar correction factors are used in Olive (2007, 2017, 2018) and Pelawa Watagoda and Olive (2021) to create prediction intervals for regression models and prediction regions for multivariate regression models. Note that for $h = 1$, an estimator for $\sigma^2 = V(e)$ is

$$\hat{\sigma}^2 = \frac{1}{n_1 - p - q} \sum_{i=1}^{n_1} \hat{e}_i^2 \approx \frac{1}{n_1} \sum_{i=1}^{n_1} e_i^2,$$

suggesting that

$$\sqrt{\frac{n_1}{n_1 - p - q}} \;\; \hat{e}_i \approx e_i.$$

Next we consider time series PIs after model selection. Let the full model be the ARMA(p_{max}, q_{max}) model. Let I_{min} be the ARMA(p_m, q_m) model that minimized a criterion such as AIC, AIC_C, or BIC. Find $\hat{Y}_n(h)$ and the forecast residuals $\hat{e}_t(h)$ for the selected model I_{min}. For $h = 1$ we will use the residuals $\hat{e}_t$. Let $k = p_m + q_m + a$ where a is usually 0 or 1, and $\tilde{e}_t(h) = \sqrt{\frac{n}{n - k}} \hat{e}_t(h)$. Let $q_n = \min(1 - \delta + 0.05, 1 - \delta + k/n_h)$ for $\delta > 0.1$ and

$$q_n = \min(1 - \delta/2, 1 - \delta + 10\delta k/n_h), \quad \text{otherwise.}$$

If $1 - \delta < 0.999$ and $q_n < 1 - \delta + 0.001$, set $q_n = 1 - \delta$. Then compute the shorth(c_{mod}) PI $[\hat{L}_n(h), \hat{U}_n(h)]$ from the n_h scaled forecast residuals $\tilde{e}_t(h)$ with

$$c_{mod} = \min(n_h, \lceil n_h[q_n + 1.12\sqrt{\delta/n_h}\ ]\ \rceil). \tag{7}$$

Then the new large sample $100(1-\delta)\%$ PI for Y_{n+h} is

$$[L_n, U_n] = [\hat{Y}_t(h) + \hat{L}_n(h), \hat{Y}_t(h) + \hat{U}_n(h)]. \tag{8}$$

Similar correction factors were used by Olive, Rathnayake, and Haile (2021) for prediction intervals for regression models, such as generalized linear models, after variable selection.

Why might PIs (6) and (8) have good coverage? For both the test data and the training data, $Y_{t+h} = \hat{Y}_t(h) + \hat{e}_t(h) = \mu_{t+h} + e_t(h)$. First, consider the training data where n_h forecast residuals $\hat{e}_t(h)$ exist. Then the proportion of $Y_{t+h} \in [\hat{Y}_t(h) + L_n(h), \hat{Y}_t(h) + U_n(h)]$ = the proportion of the n_h forecast residuals $\hat{e}_t(h) \in [L_n(h), U_n(h)] \approx 1-\delta_n \geq 1-\delta$ by construction. Hence the training data coverage is good. If the selected fitted model is good, and the test data behaves like the training data, then we expect the test data coverage to be good. Hence we need consistent estimators and large n.

Second, assume the time series follow a weakly stationary ARMA model, and suppose $\hat{Y}_t(h)$ is a consistent estimator of μ_{t+h} and $\hat{e}_t(h)$ estimates $e_t(h)$ in that $\hat{e}_t(h) - e_t(h) \xrightarrow{D} 0$ as $n \to \infty$. Also assume that the percentiles of $\hat{e}_t(h)$ estimate the percentiles of $e_t(h)$ such that $P(e_t(h) \in [L_n(h), U_n(h)]) \to 1-\delta$ as $n \to \infty$. Then $P(Y_{n+h} \in [\hat{Y}_n(h) + L_n(h), \hat{Y}_n(h) + U_n(h)]) \approx P(e_t(h) \in [L_n(h), U_n(h)]) \approx 1-\delta$. These assumptions are roughly the assumptions made when normality is assumed, which makes the time series strictly stationary. For $h = 1$, the $\{\hat{e}_{t+1}\} = \{\hat{e}_t(1)\}$ estimate the iid $\{e_t\}$, and these assumptions may be reasonable if consistent estimators are used and n is large. For weakly stationary ARMA models, $\mu_{t+h} \to \mu$, $\hat{Y}_t(h) \to \mu$, and $\hat{e}_t(h)$ estimates $Y_{t+h} - \mu$ as $h \to \infty$. Lee and Scholtes (2014) discuss when the percentiles of forecast errors are consistent for ARMA models.

If the model selection estimator is based on a consistent estimator and the probability that the model selection estimator underfits goes to 0 as $n \to \infty$, then the model selection estimators tend to be consistent by Haile and Olive (2022). For example, use the Yule Walker estimator and AIC for AR(p) variable selection. We recommend using PI (5) when $n_h < 50$ if the fitted full ARMA(p_{max}, q_{max}) model is weakly stationary, causal, and invertible. PI (5) is also useful for MA(q) models if $h > q$, and for MA(∞) models if h is large.

3 Examples and Simulations

Example 2. The presidents data set, available from *R*, consists of (approximately) quarterly approval rating for the President of the United States from the first quarter of 1945 to the last quarter of 1974 with six missing values. Figure 2 shows the time series where the last value was used as test data. The two horizontal lines correspond to PI (5). The two circles correspond to the one step ahead PI (6) for the fitted AR(1) model, and did contain the test value. Note that the one step ahead PI is much shorter than PI (5).

The remainder of this section gives simulations for the methods described in this paper. More simulations and tables are in Haile (2022). With 5000 runs, coverages between 0.94 and 0.96 suggest that there is no reason to believe that the nominal coverage

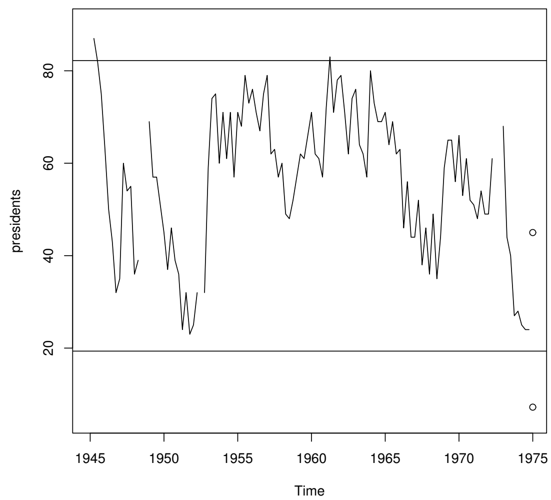

Figure 1: PI Plot of the Presidents Data Set

is not 0.95, while coverages between 0.48 and 0.52 suggest that there is no reason to believe that the nominal coverage is not 0.5.

The first simulation used the MA(2) model where the distribution of the iid e_t was N(0,1), t_5, $U(-1,1)$ or (EXP(1) - 1), a shifted exponential distribution. All these distributions have mean 0, but the fourth distribution is not symmetric. The Gaussian maximum likelihood estimator (GMLE) $(\hat{\theta}_1, \hat{\theta}_2)$ was used since this estimator is consistent for a large class of error distributions. See Haile and Olive (2022) for references. The simulation generated 5000 time series of length $n + L$ and PIs were found for the test data $Y_{n+1}, ..., Y_{n+L}$ using the training data $Y_1, ..., Y_n$. The simulations used $L = 7$ and 95% and 50% nominal PIs. The PIs used were the normal PI (4), PI (5) for Y_{t+h} where $h > 2$, and PI (6) for $h = 1, 2$. These PIs are denoted by N and A respectively in the tables. The simulated coverages and average lengths of the PI are shown. For $h = 1$, the asymptotic optimal lengths of the 95% PIs are 3.92, 5.141, 1.9, and 2.996, while the asymptotic lengths of the normal nominal 95% PIs are $3.92\sigma = 3.92$, 5.061, 2.263, and 3.92 for the N(0,1), t_5, $U(-1,1)$ and (EXP(1) - 1) distributions, respectively.

From Table 1 for normal errors, note that for $n = 400$, the coverages and lengths of PIs (5) and (6) were similar to the those of PI (4). PIs (5) and (6) were longer than the normal PI (4) for $n = 100$ and normal errors. For t_5 errors, the 95% normal PI (4) worked well since $5.141 \approx 5.061$, but the nominal 50% normal PI (4) had coverage that was too high and the average lengths were too large. The alternative PIs had coverage near 50% with shorter average lengths. For uniform errors, the normal PIs (4) were too long and the coverage was too high for 95% PIs. The alternative PIs (5) and (6) had good coverage. From Table 2 with EXP(1) - 1 errors, for 95% PIs the normal PIs (4) were usually longer than the alternative PIs (5) and (6). For the 50% PIs, the normal PIs (4) were too long with coverage that was too high. The alternative PIs (5) and (6) were shorter with good coverage.

Table 1: Normal Errors

δ	n	PI	h=1	h=2	h=3	h=4	h=5	h=6	h=7
0.05	100	N	0.9354	0.9428	0.9526	0.9456	0.9496	0.9410	0.9442
0.05	100		3.900	4.087	4.214	4.214	4.214	4.214	4.214
0.05	100	A	0.9520	0.9652	0.9586	0.9518	0.9576	0.9510	0.9530
0.05	100		4.329	4.746	4.480	4.480	4.480	4.480	4.480
0.05	400	N	0.9444	0.9444	0.9506	0.9466	0.9536	0.9522	0.9442
0.05	400		3.913	4.077	4.182	4.182	4.182	4.182	4.182
0.05	400	A	0.9444	0.9480	0.9468	0.9464	0.9512	0.9460	0.9478
0.05	400		3.980	4.192	4.209	4.209	4.209	4.209	4.209
0.5	100	N	0.4888	0.4968	0.5004	0.4856	0.4966	0.4914	0.4948
0.5	100		1.326	1.388	1.431	1.431	1.431	1.431	1.431
0.5	100	A	0.5100	0.5162	0.5004	0.4898	0.4998	0.4892	0.4926
0.5	100		1.459	1.533	1.496	1.496	1.496	1.496	1.496
0.5	400	N	0.4940	0.49304	0.5028	0.5100	0.4884	0.4858	0.4924
0.5	400		1.344	1.399	1.435	1.435	1.435	1.435	1.435
0.5	400	A	0.4906	0.4902	0.4894	0.5020	0.4816	0.4808	0.4800
0.5	400		1.356	1.413	1.432	1.432	1.432	1.432	1.432

Table 2: EXP(1) - 1 Errors

δ	n	PI	h=1	h=2	h=3	h=4	h=5	h=6	h=7
0.05	100	N	0.9424	0.9516	0.9478	0.9458	0.9432	0.9422	0.9446
0.05	100		3.872	4.052	4.177	4.177	4.177	4.177	4.177
0.05	100	A	0.9590	0.9712	0.9562	0.9512	0.9492	0.9492	0.9550
0.05	100		3.726	4.389	4.047	4.047	4.047	4.047	4.047
0.05	400	N	0.9556	0.9442	0.9496	0.9446	0.9458	0.9414	0.9486
0.05	400		3.908	4.072	4.177	4.177	4.177	4.177	4.177
0.05	400	A	0.9598	0.9540	0.9496	0.9472	0.9462	0.9434	0.9504
0.05	400		3.224	3.689	3.8093	3.809	3.809	3.809	3.809
0.5	100	N	0.5250	0.5418	0.5528	0.5516	0.5620	0.5494	0.5546
0.5	100		1.323	1.382	1.425	1.425	1.425	1.425	1.425
0.5	100	A	0.5070	0.5068	0.5018	0.4920	0.4956	0.5012	0.5012
0.5	100		0.901	1.023	1.029	1.029	1.029	1.029	1.029
0.5	400	N	0.5358	0.5618	0.5620	0.5550	0.5604	0.5454	0.5568
0.5	400		1.342	1.397	1.432	1.432	1.432	1.432	1.432
0.5	400	A	0.5004	0.5042	0.4984	0.5028	0.4934	0.4842	0.4974
0.5	400		0.760	0.905	0.970	0.970	0.970	0.970	0.970

Table 3: One Step Ahead PI after Model Selection

n	dist	PI 6	PI 8	PI F	PI 4
100	N	0.9582	0.9592	0.9442	0.9476
100		4.4553	4.3214	3.8857	3.9341
100	t5	0.9504	0.9550	0.9412	0.9434
100		5.7340	5.6747	5.0015	5.06377
100	U	0.9728	0.9776	0.9842	0.9860
100		2.3876	2.1992	2.2538	2.2819
100	sExp	0.9536	0.9540	0.9406	0.9424
100		4.0179	3.7989	3.8504	3.8983
400	N	0.9458	0.9500	0.9470	0.9476
400		4.2054	3.9990	3.9119	3.9239
400	t5	0.9432	0.9444	0.9404	0.9412
400		5.4640	5.2364	5.0455	5.0609
400	U	0.9518	0.9576	0.9988	0.9992
400		2.2084	1.9644	2.2593	2.2662
400	sExp	0.9558	0.9578	0.9508	0.9518
400		3.8057	3.2935	3.9047	3.9166
800	N	0.9516	0.9526	0.9514	0.9520
800		4.1704	3.9445	3.9147	3.9206
800	t5	0.9458	0.9480	0.9452	0.9456
800		5.4334	5.1604	5.0491	5.0568
800	U	0.9500	0.9524	0.9994	0.9994
800		2.1838	1.9255	2.2605	2.2640
800	sExp	0.9438	0.9438	0.9410	0.9410
800		3.7821	3.1842	3.9147	3.9207

The second simulation was for PI (8) after model selection with the GMLE and AIC_C. The full model was the ARMA(5,5) model. The true model was an MA(2) model. The same error types as in the first simulation were used. The simulations used the 1-step ahead PI with $h = 1$ for ease of programming. PIs (4), (6), and (8) were used, as well as the PI given by the Hyndman and Khandakar (2008) *forecast package*, denoted by (F). Table 3 gives some results. PIs (4) and (F) were very similar. PI (8) tended to have the shortest average length unless the e_t were iid $N(0, \sigma_e^2)$ or t_5.

4 Discussion

The one step ahead prediction intervals (6) and (8) are easy to program since the software provides the residuals. See the last paragraph of this section. The Chebyshev "95%" prediction intervals such as (4) are useful even after model selection, provided that consistent estimators of μ_{t+h} and σ_{t+h} are used, but the asymptotic coverage could be between 0.74 and 1.0, depending on the error distribution. Using a holdout sample of the last few h cases of the training data set can be useful for checking various prediction intervals for coverage and length for competing times series models.

The correction factors for the prediction intervals of this paper help compensate for estimation of the model parameters and model selection for moderate n. Hyndman and Athanasopoulos (2018, last paragraph of $\oint$8.8) note that ARIMA-based prediction intervals tend to be too narrow, so actual coverage is less than the nominal coverage. See Bhansali (1981) for the effects of estimating the order of the time series model. Data sets where the future data does not behave like the past data are common, and then the prediction intervals tend to perform poorly.

There is a large literature on time series PIs, especially for AR(p) models, and the bootstrap is often used. It may be difficult to bootstrap a model selection estimator. Then data splitting may be useful. See Alonso, Peńa, and Romo (2002, 2003), Brockwell and Davis (2016), Clements and Kim (2007), deLuna (2000), Hyndman and Athanasopoulos (2018), Kabaila and He (2007), Lu and Wang (2020), Pan and Politis (2016a), Pascual, Romo, and Ruiz (2001), Thombs and Schucany (1990), Vidoni (2009), and Wolf and Wunderli (2015) for references. Some papers on the shorth include Chen and Shao (1999), Grübel (1988), and Einmahl and Mason (1992). See Hong, Kuffner, and Martin (2018) for why classical PIs after AIC variable selection do not work.

Plots and simulations were done in *R*. See R Core Team (2020). Programs are in the collection of functions *tspack.txt*. See (http://parker.ad.siu.edu/Olive/tspack.txt). The function `locpi` gets PI (5). The function `locpi2` needs the forecast residuals, and finds $[L_n, U_n]$ used in PIs (6) and (8). The function `onesteppi` gets the one step ahead PI for seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ models with period s where the 6 estimated parameters need to be given. The function can handle missing values entered as NA. The function `pimasim2` was used for Tables 1 and 2. For Table 3, the function `pitsvssim` simulates PI (8) after model selection using the GMLE with AIC_C using the *R* function `auto.arima` from the Hyndman and Khandakar (2008) *forecast package*. Also see Hyndman and Athanasopoulos (2018).

REFERENCES

Akaike, H. (1973), "Information Theory as an Extension of the Maximum Likelihood Principle," in *Proceedings, 2nd International Symposium on Information Theory*, eds. Petrov, B.N., and Csakim, F., Akademiai Kiado, Budapest, 267-281.

Alonso, A.M., Peńa, D., and Romo, J. (2002), "Forecasting Time Series with Sieve Bootstrap," *Journal of Statistical Planning and Inference*, 100, 1-11.

Alonso, A.M., Peńa, D., and Romo, J. (2003), "On Sieve Bootstrap Prediction Intervals," *Statistical & Probability Letters*, 65, 13-20.

Bhansali, R.J. (1981), "Effects of Not Knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction-I," *Journal of the American Statistical Association*, 76, 588-597.

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

Brockwell, P.J., and Davis, R.A. (2016), *Introduction to Time Series and Forecasting*, 3rd ed., Springer, New York, NY.

Chen, M.H., and Shao, Q.M. (1999), "Monte Carlo Estimation of Bayesian Credible and HPD Intervals," *Journal of Computational and Graphical Statistics,* 8, 69-92.

Claeskens, G., and Hjort, N.L. (2008), *Model Selection and Model Averaging*, Cambridge University Press, New York, NY.

Clements, M.P., and Kim, N. (2007), "Bootstrapping Prediction Intervals for Autoregressive Time Series," *Computational Statistics & Data Analysis*, 51, 3580-3594.

deLuna, X. (2000), "Prediction Intervals Based on Autoregressive Forecasts," *Journal of the Royal Statistical Society, D,* 49, 87-93.

Einmahl, J.H.J., and Mason, D.M. (1992), "Generalized Quantile Processes," *The Annals of Statistics*, 20, 1062-1078.

Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference*, 143, 1039-1048.

Granger, C.W.J., and Newbold, P. (1977), *Forecasting Economic Time Series*, Academic Press, New York, NY.

Grübel, R. (1988), "The Length of the Shorth," *The Annals of Statistics,* 16, 619-628.

Haile, M.G. (2022), "Inference for Time Series after Variable Selection," Ph.D. Thesis, Southern Illinois University. See (http://parker.ad.siu.edu/Olive/shaile.pdf).

Haile, M.G., and Olive, D.J. (2022), "Bootstrapping ARMA Time Series Models after Model Selection," unpublished manuscript. See (http://parker.ad.siu.edu/Olive/pptsboot.pdf).

Hannan, E.J., and Quinn, B.G. (1979), "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, B*, 41, 190-195.

Hong, L., Kuffner, T.A., and Martin, R. (2018), "On Overfitting and Post-Selection Uncertainty Assessments," *Biometrika*, 105, 221-224.

Hurvich, C., and Tsai, C.L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297-307.

Hyndman, R.J., and Athanasopoulos, G. (2018), *Forecasting: Principles and Practice*, 2nd edition, OTexts: Melbourne, Australia. https://OTexts.org/fpp2/

Hyndman, R.J., and Khandakar, Y. (2008), "Automatic Time Series Forecasting: the Forecast Package for R," *Journal of Statistical Software*, 27, 1-22.

Kabaila, P., and He, Z. (2007), "Improved Prediction Limits for AR(p) and ARC(p) Processes," *Journal of Time Series Analysis*, 29, 213-223.

Lee, Y.S., and Scholtes, S. (2014), "Empirical Prediction Intervals Revisited," *International Journal of Forecasting,* 30, 217-234.

Lu, X., and Wang, L. (2020), "Bootstrap Prediction Interval for ARMA Models with Unknown Orders," *Revstat-Statistical Journal*, 18, 375-396.

Masters, T. (1995), *Neural, Novel, & Hybrid Algorithms for Time Series Prediction*, Wiley, New York, NY.

Mykland, P.A. (2003), "Financial Options and Statistical Prediction Intervals," *The Annals of Statistics*, 31, 1413-1438.

Olive, D.J. (2007), "Prediction Intervals for Regression Models," *Computational Statistics & Data Analysis,* 51, 3115-3122.

Olive, D.J. (2013), "Asymptotically Optimal Regression Prediction Intervals and Prediction Regions for Multivariate Data," *International Journal of Statistics and Probability*, 2, 90-100.

Olive, D.J. (2017a), *Robust Multivariate Analysis*, Springer, New York, NY.

Olive, D.J. (2017b), *Linear Regression*, Springer, New York, NY.

Olive, D.J. (2018), "Applications of Hyperellipsoidal Prediction Regions," *Statistical Papers*, 59, 913-931.

Olive, D.J., Rathnayake, R.C., and Haile, M.G. (2022), "Prediction Intervals for GLMs, GAMs, and Some Survival Regression Models," *Communication in Statistics: Theory and Methods*, 52, 8012-8026.

Pan, L., and Politis, D.N. (2016a), "Bootstrap Prediction Intervals for Linear, Nonlinear, and Nonparametric Autoregressions," *Journal of Statistical Planning and Inference*, 177, 1-27.

Pan, L., and Politis, D.N. (2016b), "Bootstrap Prediction Intervals for Markov Processes," *Computational Statistics & Data Analysis*, 100, 467-494.

Pankratz, A. (1983), *Forecasting with Univariate Box-Jenkins Models*, Wiley, New York, NY.

Pascual, L., Romo, J., and Ruiz, E., (2004),"Bootstrap Predictive Inference for ARIMA Processes," *Journal of Time Series Analysis*, 25, 449-465.

Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), "Comparing Six Shrinkage Estimators with Large Sample Theory and Asymptotically Optimal Prediction Intervals," *Statistical Papers,* 62, 2407-2431.

R Core Team (2020), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

Ross, S.M. (2014), *Introduction to Probability Models*, 11th ed., Academic Press, San Diego, CA.

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's Information Criterion," *Biometrika*, 63, 117-126.

Thombs, L.A. and Schucany, W.R. (1990), "Bootstrap Prediction Intervals for Autoregression," *Journal of the American Statistical Association*, 85, 486-492.

Vidoni, P. (2009), “A Simple Procedure for Computing Improved Prediction Intervals for Autoregressive Models,” *Journal of Time Series Analysis*, 30, 577-590.

White, H. (1984), *Asymptotic Theory for Econometricians,* Academic Press, San Diego, CA.

Wolf, M., and Wunderli, D. (2015), “Bootstrap Joint Prediction Regions,” *Journal of Time Series Analysis*, 36, 352-376.