International Journal of Statistics and Applied Mathematics 2024; 9(3): 65-70

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(3): 65-70

https://www.mathsjournal.com
Received: 21-03-2024
Accepted: 02-05-2024

Mosisa Aga
Auburn University Montgomery, Alabama, USA

Corresponding Author:
Mosisa Aga
Auburn University Montgomery, Alabama, USA

Bootstrapping the coefficients of a polynomial regression with ARFIMA errors

Mosisa Aga

DOI: https://dx.doi.org/10.22271/maths.2024.v9.i3a.1732

Abstract
The purpose of this paper is to establish the validity of a bootstrap least square estimate of a polynomial regression model exhibiting an autoregressive fractionally integrated moving average ARFIMA (p,d,q) errors. Under standard conditions on the regression parameters and the error components, the bootstrap is shown to be valid. In other words, for a $p\times1$ vector β of unknown parameters, $\hat{\beta}_m$ a 'modified' least square estimate of β, $\hat{\beta}^*$a bootstrap estimate of β, and $C \in R^k$ such that $C'(\hat{\beta}_m - \beta)$ has finite variance, it is shown that the distribution of $C'(\hat{\beta}^* - \hat{\beta}_m)$ converges to that of $C'(\hat{\beta}_m - \beta)$, uniformly in C. This work is an extension of that of Freedman (1981) and Eck (2018) to the case where the error term is a strongly dependent time series.

Keywords: Bootstrap, least square estimate, linear regression, long memory, Mallows Metric.
2000 Mathematics Subject Classification: 62M10

1. Introduction

The linear regression model is an important tool in statistical analyses in which we study the effects of explanatory variables or covariates on a response variable. Regression analysis is primarily used for predicting values of the response variable at interesting values of the predictor variables, discovering the predictors that are associated with the response variable, and estimating how changes in the predictor variables affects the response variable (Eck, 2018) [6]. Following the introduction of the bootstrap approximation technique by Efron (1979) [7], the first systematic treatments of bootstrapping of regression models was probably the work of Freedman (1981) [9] in which the validity of bootstrapping of the coefficient of "regression models" (where the design matrix is not random) and that of "correlation models" (where the design matrix is allowed to be random) are established. During subsequent years Shorack (1982) [16], Freedman and Peter (1984) [10], Weber (1984) [19], Wu (1986), Shao (1988) [15], Efron (1991) [8], and others have rigorously expanded the applicability of the bootstrap to various aspects of regression models.

Shorack (1982) [16] established the validity of bootstrap for robust M-estimates of a linear regression in which the error terms are independent and identically distributed (iid) random variables. Freedman and Peter (1984) [10] studied the application of bootstrap in estimating standard errors for regression coefficients obtained by constrained generalized least squares with an estimated covariance matrix where the errors are assumed to be iid. Weber (1984) [19] examined bootstrapping of functions of the parameters of regression models with iid errors. Wu (1986) considered three bootstrap methods in estimating the variance of least square estimators of a regression model with uncorrelated errors. Shao (1988) [15] investigated the application of the jackknife and the bootstrap in estimating the bias and variance of the parameter of a linear model when the error terms are independent and heteroscedastic. Edron (1991) discussed the estimation of regression percentiles

More recently, Eck (2018) [6] extended the work of Freedman (1981) [9] to multivariate linear regression for the case where the error terms are independent. All of the above works on the application of bootstrap to regression models have one thing in common: they all require the error terms to be independent.

Among the few works in the literature on bootstrapping linear regression models with dependent error structures are Stute (1995) [18], McKnight *et al*. (2000) [14], Aga (2007) [1], and Aga (2023b) [2]. Stute (1995) [18] and McKnight *et al*. (2000) [14] both deal with bootstrapping regression models with short memory auto-regressive (AR) errors. Although Aga (2007) [1] and Aga (2023b) [2] consider bootstrapping of linear regression models with long memory errors, both of them deal with bootstrapping the variance parameters of the error components, not the regression parameter β. This paper builds on the methods presented by Freedman (1981) [9] and Eck (2018) [6] and investigates the validity of bootstrap procedure as it is applied to the regression coefficient β of a polynomial regression model whose error component is the well-known long memory ARFIMA (p,d,q) errors.

Consider a polynomial regression model with long memory

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots \beta_q t^q + \varepsilon_t = \sum_{i=0}^{r} \beta_i t^i + \varepsilon_t \tag{1.1}$$

where $\{\varepsilon_t, t \geq 1\}$ is a stationary, Gaussian, and long memory discrete time series with mean zero and and spectral density $f_\theta(\lambda)$ for $\lambda \in (-\pi, \pi)$, where $\theta = (\theta_1, \theta_2, \dots, \theta_m)' \in \mathrm{R}^m$ and

$$f_\theta(\lambda) = O(|\lambda|^{-2d-\delta}) \tag{1.2}$$

as $|\lambda| \to 0$, $\forall \delta > 0$, $d \in (0,1/2)$, and $\theta_1 = d$, referred to as the "long-memory parameter" of the process (see Andrews *et al*. (2006) [3].

Given a finite sample of the process, (1.1) may be written in matrix notation as

$$Y = Z\beta + \varepsilon \tag{1.3}$$

where $(Y_1, \dots, Y_n)'$ and $(\varepsilon_1, \dots, \varepsilon_n)'$ are $n \times 1$ vectors, $\beta = (\beta_0, \dots, \beta_r)$ is $(r + 1) \times 1$, and the design matrix Z given by $Z = [Z^{(1)}, \dots, Z^{(r+1)}]$ and $Z^{(j)} = (1, 2^{j-1}, \dots, n^{(j-1)})'$, $j = 1,\dots,r + 1$. We assume that Z has the full rank $(r + 1)$, where $(r + 1) \leq n$.

Let $\Sigma = \gamma_\varepsilon(s,t)$ be the covariance function of ε_t. If the error term ε_t is assumed to be a white noise, then the ordinary least squares estimate $\hat{\beta}$ of β is given by

$$\hat{\beta} = (Z'Z)^{-1}Z'Y. \tag{1.4}$$

in which case $\gamma_\varepsilon(s, t) = 0$ for $s \neq t$, and $\gamma_\varepsilon(t, t) = \sigma^2$, independent of t. In our current model the error term ε_t is a long memory time series and therefore, using (1.4) is inappropriate (Shumway R. H. *et al*. (2017) [17].

The most well-known model for long-memory processes $\{\varepsilon_t\}$ satisfying (1.2) is the autoregressive fractionally integrated moving average ARFIMA (p,d,q) process introduced by Hosking (1980) [12] and Grangr *et al*. (1981) and defined by

$$\phi(B)\varepsilon_t = \theta(B)(1 - B)^{-d}\epsilon_t, \tag{1.5}$$

where B is the back-shift operator, $\phi(B) = 1+\phi_1 B+ \dots +\phi_p B^p$ and $\theta(B) = 1+\theta_1 B+ \dots + \theta_q B^q$ are autoregressive and moving-average operators, $\phi(B)$ and $\theta(B)$ have no common roots, $d \in \left(0, \frac{1}{2}\right)$, and $(1 - B)^{-d}$ defined by the binomial formula $(1 - B)^{-d} = \sum_{j=0}^{\infty} \eta_j B^j$, where

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \tag{1.6}$$

and Γ is the gamma function, and ϵ_t is a white noise sequence with finite variance σ^2. Let $\pi(B) = (1 - B)^d \theta(B)^{-1}\phi(B)$. Then, multiplying both sides of (1.5) by $(1 - B)^d \theta(B)^{-1}$ we obtain

$$\pi(B)\varepsilon_t = \epsilon_t \tag{1.7}$$

and consequently, (1.3) can now be transformed to a classical linear model with a white noise errors having the form

$$\tilde{Y} = \tilde{Z}\beta + \epsilon, \tag{1.8}$$

where $\tilde{Y} = \pi(B)Y$, $\tilde{Z} = \pi(B)Z$, and $\epsilon = (\epsilon_1, \dots, \epsilon_n)$.

Therefore, the modified least squares estimate $\hat{\beta}_m$ of β based on (1.8) is given by

$$\hat{\beta}_m = \left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'\tilde{Y} \tag{1.9}$$

and the residual vector $\hat{\epsilon} = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_n)$ is given by

$$\hat{\epsilon} = \tilde{Y} - \tilde{Z}\hat{\beta}_m. \tag{1.10}$$

The remainder of the paper proceeds as follows. In section 2 we present the bootstrap procedure proposed for our model. In section 3 we define the Mallows metric that is vital in proving the main result and state and prove a number of lemmas associated with it. In section 4 we state and prove the main theorem about the validity of the bootstrap procedure described in section 2. Section 5 presents a brief conclusion.

The Bootstrap Procedure

The main problem in finding $\hat{\beta}_m$ as given in (1.9) is that we do not typically know the behavior of the error term ε_t and hence the coefficient $\pi(B)$ in (1.8) is not predetermined. An easy way to tackle this problem was first presented in Cochrane et. al (1949) and with the aid of the emergence of cheap computing, the technique is modified and presented in Algorithm 1 below.

Algorithm 1

1. First, run an ordinary regression of Y_t on Z_t and compute the ordinary least square estimate $\hat{\beta} = (Z'Z)^{-1}Z'Y$ as if the errors ε_t are uncorrelated. Retain the residuals $\hat{\varepsilon}_t = Y_t = \sum_{i=0}^{r} \beta_i t^i$.
2. Fit ARFIMA (p,d,q) to the errors $\hat{\varepsilon}_t$ and find the estimates $\Theta = (\hat{\sigma}^2, \hat{d}, \hat{\phi}, \hat{\theta})$ and let $\hat{\epsilon}_t$ be an estimate of ϵ_t given in (1.5).
3. Let $\hat{\pi}(B) = (1 - B)^d\, \hat{\theta}(B)^{-1}\hat{\phi}(B)$ be an estimate of $\pi(B) = (1 - B)^d \theta(B)^{-1}\phi(B)$ computed using the estimated parameters in step 2 above.
4. Compute a modified least squares estimate $\hat{\beta}_m$ of the regression model with ARFIMA (p,d,q) errors using (1.9) and replacing $\pi(B)$ by $\hat{\pi}(B)$ from step 3 above.
5. Perform residual analysis on $\hat{\epsilon}_t$ for whiteness, and adjust the model if necessary.

Now we turn our focus to the main purpose of this paper, that is, to describe a bootstrap procedure of obtaining an estimate $\hat{\beta}^*$ of β and establish that, under mild conditions, the bootstrap approximation is valid.

Bootstrapping is one of the different re-sampling techniques in which a series of random samples are drawn a large number of times with replacement from an original sample Y obtained from the population of interest. The statistic of interest is then calculated from each of the bootstrap samples and an approximate of the sampling distribution of the statistic is obtained from the calculated values.

In this section we describe a bootstrap procedure for estimating the vector of unknown parameters β of the regression model (1.3). Let $Y = (Y_1, ..., Y_n)$ be a sample from our linear regression model with strongly dependent errors as described in section 1. In a regression model with deterministic design it is appropriate to resample residuals (Freedman (1981) [9], Eck (2018) [6]. Let $\hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_n)$ be a column n-vector of observable residuals estimated as described in Algorithm 1 and let $\bar{\epsilon} = \sum_{i=1}^{n} \hat{\epsilon}_i$. Let $\hat{F}_n$ be the empirical distribution of the centered residuals, assigning mass $1/n$ to each $\hat{\epsilon}_i - \hat{\epsilon}$, for $i = 1, ..., n$. Given $Y = (Y_1, ..., Y_n)$, let $\epsilon^* = (\epsilon_1^*, ..., \epsilon_n^*)$ be conditionally independent, drawn from the common distribution $\hat{F}_n$ and let

$$\tilde{Y}^* = \hat{\pi}(B) Z \hat{\beta}_m + \epsilon^* \qquad (2.1)$$

where $\hat{\pi}$ and $\hat{\beta}_m$ are as computed in Algorithm 1. Then, a procedure to compute a bootstrap estimate $\hat{\beta}^*$ of β can be described by the following algorithm:

Algorithm 2

1. Set B and initialize $b = 1$.
2. Resample residuals $\hat{\epsilon}_i - \bar{\epsilon}$, $i = 1, ..., n$ from $\hat{F}_n$, with replacement, and compute let $\hat{\epsilon}_b^* = (\hat{\epsilon}_{1b}^*, ..., \hat{\epsilon}_{nb}^*)$; store $\hat{\epsilon}_b^*$.
3. Compute $\tilde{Y}_b^* = \hat{\pi}(B) Z \hat{\beta}_m + \hat{\epsilon}_b^*$.
4. Compute $\hat{\beta}_b^* = (\tilde{Z}'\tilde{Z})^{-1} \tilde{Z}' \tilde{Y}_b^*$; store $\hat{\beta}_b^*$, and set $b = b + 1$.
5. Repeat steps 2-4 $(B - 1)$ times.
6. Compute $\hat{\beta}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_b^*$ and $\hat{\epsilon}^* = (\hat{\epsilon}_1^*, ..., \hat{\epsilon}_n^*)$ where $\hat{\epsilon}_i^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\epsilon}_{ib}^*$ for $i = 1, ..., n$.

The Mallows Metrics

We present a Mallows metric (Mallows, 1972) as a central tool to prove the main results of this paper. The Mallows metrics for probabilities in R^k, $k \geq 1$, relative to Euclidean norm was the key ingredient needed to establish the validity of the residual bootstrap approximation for uni-variate linear regression (Bickel *et al*., (1981) [4], Freedman, (1981) [9] and that of a multivariate linear regression (Eck, D. J., (2018) [6], both with white noise errors.

Under the condition that $\hat{\pi}(B)$ can be estimated and be used in the bootstrap approximation, we can now state the following assumption about the model (1.8).

- **A1.** The $n\times(r+1)$ matrix $\tilde{Z}$ is not random and has the full rank $(r+1)$, where $(r+1) \leq n$..
- **A2.** The components $\epsilon_1, ..., \epsilon_n$ of the error term ϵ in (1.8) are independent with unknown common distribution F and unknown finite variance σ^2.

Let $P = \tilde{Z}'\tilde{Z}$. Then P is positive definite and therefore, it has a unique positive definite square root; that is $\sqrt{P}$. $\hat{\beta}_m$ has mean β and variance covariance matrix $\sigma^2 P^{-1}$. Suppose the transformed regression (1.8) satisfy assumptions A1 and A2. Let $\mathcal{F}_P$ and $\mathcal{G}_P$ be two alternate distributions of $\sqrt{P}(\hat{\beta}_m - \beta)$, when F and G, respectively, are alternate distribution of the $\epsilon's$. In practice F is the distribution of the residuals and G is the distribution of the centered residuals. Likewise, if C is a $(r+1)\times 1$ coefficient vector, let $\mathcal{F}_C$ and $\mathcal{G}_C$, respectively, denote the exact alternate distributions of $C'(\hat{\beta}_m - \beta)$, normalized so that $C'(\hat{\beta}_m - \beta)$ has variance σ^2. It follows that

$$C'(\tilde{Z}'\tilde{Z})^{-1}C = 1. \qquad (3.1)$$

We shall define the Mallows metric for probabilities in R^k and state some results needed in our current set up.

Definition 3.1. Let X and Y be random vectors in R^k and let F_X and F_Y represent their corresponding probability distributions. The Mallows metric between F_X and F_Y is given by

$$\rho_\alpha^k(F_X, F_Y) = \inf_{F_X, F_Y} E^{1/\alpha}(||X - Y||^\alpha \qquad (3.2)$$

where the infimum is taken over all random vectors $X, Y \in R^k$ and $||\cdot||$ is the Euclidean norm.

Lemma 3.2. Let F and G be two possible alternate distributions of the errors ϵ given in (1.8). Assume that both F and G have mean 0 and finite variance. Then,

(a) $\rho_2[\mathcal{F}_P, \mathcal{G}_P]^2 \leq (r+1)\rho_2(F,G)^2$.

(b) $\rho_2[\mathcal{F}_C, \mathcal{G}_C]^2 \leq (r+1)\rho_2(F,G)^2$.

Proof

Let ϵ_{iF} be independent having distribution F and let ϵ_F be the $n\times 1$ column vector with components ϵ_{iF}. That is, $\epsilon_F = (\epsilon_{1F}, ..., \epsilon_{nF})$. Similarly, let ϵ_{iG} be independent having distribution G and let ϵ_G be the $n\times 1$ column vector with components ϵ_{iG}. Let A be an arbitrary $(r+1)\times n$ matrix. Let $\mathcal{F}_A$ be the distribution of $A\epsilon_F$, and let $\mathcal{G}_A$ be the distribution of $A\epsilon_G$. Then by Lemma 8.9 of Bickel et. al (1981) [4] we have	
$\rho_2[\mathcal{F}_A, \mathcal{G}_A]^2 \leq trace(AA')\rho_2(F,G)^2$. Moreover,	(3.3)

$$\sqrt{P}(\hat{\beta}_m - \beta) = \sqrt{P}((\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{Y} - \beta)$$

$$= \sqrt{P}((\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'(\tilde{Z}\beta + \epsilon) - \beta)$$

$$= \sqrt{P}((\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{Z}\beta + \sqrt{P}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\epsilon) - \beta)$$

$$= (\sqrt{P})^{-1}\tilde{Z}'\epsilon, \qquad (3.4)$$

where the second equality holds because $\tilde{Y}$ is replaced by $\tilde{Z}\beta + \epsilon$ (from (1.8)) and the last equality holds because $\tilde{Z}'\tilde{Z}$ *is* replaced by P. Now, replacing the matrix A in (3.3) by $(\sqrt{P})^{-1}\tilde{Z}'$ we obtain

$$\begin{aligned} AA' &= (\sqrt{P})^{-1}\tilde{Z}'((\sqrt{P})^{-1}\tilde{Z}')' \\ &= (\sqrt{P})^{-1}\tilde{Z}'\tilde{Z}(\sqrt{P})^{-1} \\ &= (\sqrt{P})^{-1}P(\sqrt{P})^{-1} \\ &= I_{(r+1)} \end{aligned} \qquad (3.5)$$

Where $I_{(r+1)}$ is the $(r+1)\times(r+1)$ identity matrix. Since $I_{(r+1)}$ has trace $(r+1)$ the inequality in part (a) follows.

(a) Replacing A by $C'(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'$ we obtain

$AA' = C'(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\ (C'(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}')'$
$= C'(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\ \tilde{Z}'(\tilde{Z}'\tilde{Z})^{-1}$ C
$= C'I(r+1)(\tilde{Z}'\tilde{Z})^{-1}C$
$= I(r+1)$ (3.6)

where the last equality holds because $C'(\tilde{Z}'\tilde{Z})^{-1}C = 1$ by (3.1). This verifies the inequality in part (b) of the lemma. The next lemma is a slightly modified version of Lemma 2.2 of Freedman (1981) [9] and is used in the proof of Theorem 4.1.

Lemma 3.3. Let F_n be the empirical distribution of $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ and $\hat{F}_n$ be the empirical distribution of the centered residuals $(\hat{\epsilon}_n - \bar{\epsilon}, \ldots, \hat{\epsilon}_n - \bar{\epsilon})$, where $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i$. Then

$$E[\rho_2(\hat{F}_n, F_n)^2] \le \frac{\sigma^2(r+2)}{n} \quad (3.7)$$

In the proof of Theorem 4.1, section 4 we will make use of the Levy-Prokhorov metric defined as follows.

Definition 3.4: Let α and β be two probability distributions on R^k. Then, the Levy-Prokhorov metric $\rho_{lp}(\eta,\beta)$ is given by

$$\rho_{lp}(\eta,\beta) = \inf_{\varepsilon \ge 0, K}[\eta(K) \le \beta(K^{\varepsilon}) + \varepsilon \ \ and \ \ \beta(K) \le \eta(K^{\varepsilon}) + \varepsilon \quad (3.8)$$

for all compact $K \in R^k$, where $K^{\varepsilon} = \{x \in R^k / \|x\| \le \varepsilon\}$.
The next lemma establishes the relationship between the Mallows metric and the Levy-Prokhorov metric.

Lemma 3.5. For any two probability measures η and β on R^k we have

$$\rho_{lp}(\eta,\beta) \le \rho_{\alpha}(\eta,\beta)^{\frac{\alpha}{\alpha+1}} \quad (3.9)$$

Proof. For all compact $K \in R^k$ and for all $\varepsilon \ge 0$ we have $\eta(K) = P(X \in K)$	
$\le P(Y \in K_{\varepsilon}) + P(\|X - Y\| \ge \varepsilon)$	(3.10)

$\le P(Y \in K_{\varepsilon}) + \varepsilon^{-\alpha}E\{\|X - Y\|^{\alpha}\}$

where the last inequality holds by Chebychev's inequality. Setting

$$\varepsilon = \rho_{\alpha}(\eta,\beta)^{\frac{\alpha}{\alpha+1}}$$

and choosing X and Y to minimize the expected value, (3.10) becomes

$$\begin{aligned}\eta(K) &\le P(Y \in K_{\varepsilon}) + E\left[\|X-y\|^{\alpha}\right]^{-\frac{\alpha}{\alpha+1}} E|X-y\|^{\alpha} \\ &\le \beta(K_{\varepsilon}) + E\left[\|X-y\|^{\alpha}\right]^{\frac{\alpha^2}{\alpha+1}} \\ &\le \beta(K_{\varepsilon}) + \left(E\left[\|X-y\|^{\alpha}\right]^{\frac{1}{\alpha}}\right)^{\frac{\alpha}{\alpha+1}} \\ &= \beta(K_{\varepsilon}) + \rho_{\alpha}(\eta,\beta)^{\frac{\alpha}{\alpha+1}}.\end{aligned} \quad (3.11)$$

Similarly, $\beta(K) \le \eta(K_{\varepsilon}) + \rho_{\alpha}(\eta,\beta)^{\frac{\alpha}{\alpha+1}}$ for all compact $K \in R^k$ and $\varepsilon \ge 0$ which proves (3.9).
Consider a coefficient vector C satisfying (3.1). Let $\hat{\beta}_m$ be as given in (1.9), $\hat{\epsilon} = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)$ be as computed in Algorithm 1. Let $\hat{\beta}^*$ and $\hat{\beta}^* = (\hat{\epsilon}_1^*, \ldots, \hat{\epsilon}_n^*)$ be bootstrap estimates computed using Algorithm 2 and let

$$\hat{\sigma}^{*^2} = \frac{1}{n-(r+1)}\sum_{i=1}^{n}(\hat{\epsilon}_i^* - \bar{\hat{\epsilon}}^*)^2 \quad (3.12)$$

Where $\bar{\hat{\epsilon}}^* = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^*$ and let

$$\hat{\sigma}^2 = \frac{1}{n-(r+1)}\sum_{i=1}^{n}(\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2 \quad (3.13)$$

Where

$$\bar{\hat{\epsilon}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i.$$

The bootstrap principle is that the distribution of $C'(\hat{\beta}^* - \hat{\beta}_m)$ and $\frac{C'(\hat{\beta}^* - \hat{\beta}_m)}{\hat{\sigma}^*}$, which can be computed from the data, approximate the distributions of $C'(\hat{\beta}_m - \beta)$ and $\frac{C'(\hat{\beta}_m - \beta)}{\hat{\sigma}}$, respectively. The main theorem of this paper given below justifies the use of bootstrap in this context for our current polynomial regression model with long memory errors as given in (1.8).

Validity of the Bootstrap
Theorem 4.1. Suppose assumptions **A1** and **A2** hold on the regression model (1.8). Let $C \in R^k$ satisfy (3.1) and let $\mathcal{F}_C^*$ and $\mathcal{F}_C$ be the distributions of $C'(\hat{\beta}^* - \hat{\beta}_m)$ and $C'(\hat{\beta}_m - \beta)$, respectively. Likewise, let $\mathcal{G}_C^*$ and $\mathcal{G}_C$ be the distributions of $\frac{C'(\hat{\beta}^* - \hat{\beta}_m)}{\hat{\sigma}^*}$ and $\frac{C'(\hat{\beta}_m - \beta)}{\hat{\sigma}}$, respectively. Condition on $\tilde{Y}_1, \ldots, \tilde{Y}_n$, let $n \to \infty$ and assume that if $(r + 1) \to \infty$, then $(r + 1)/n \to 0$. Then
(a) $\rho_2(\mathcal{F}_C^*, \mathcal{F}_C) \to 0$ uniformly in C.
(b) The conditional distribution of $\hat{\sigma}^*$ converges weakly to σ.
(c) $\rho_2(\mathcal{G}_C^*, \mathcal{G}_C) \to 0$ uniformly in C.

Proof.
(a) Replacing G by $\hat{F}_n$ in

Lemma 3.2 (b) we have

$$\rho 2(\mathcal{F}_C^*, \mathcal{F}_C) \le \rho 2(F, \hat{F}_n)^2 \le \rho 2(F, Fn)^2 + \rho 2(Fn, \hat{F}_n)^2. \quad (4.1)$$

But $\rho 2(Fn, \hat{F}_n)^2 \to o$ by

Lemma 3.3 because $(r + 2)/n \to o$ by assumption. On the other hand, $\rho_2(F,F_n) \to 0$ by Lemma 8.4 of Bickel et. al (1981) [4]. This proves (a). We first show that $\hat{\sigma}_n \to \sigma$ almost everywhere. Let

$$\sigma_n^2 = \frac{1}{n-(r+1)}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2 \quad (4.2)$$

where $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i$. Note that

$$\sigma_n = \sqrt{\frac{1}{n-(r+1)}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2} = \frac{1}{\sqrt{n-(r+1)}}\|\epsilon - \bar{\epsilon}\| \quad (4.3)$$

where $\epsilon - \bar{\epsilon}$ is to mean $(\epsilon_1 - \epsilon, \ldots, \epsilon_n - \bar{\epsilon})$. Similarly, we have

$$\hat{\sigma}_n = \sqrt{\frac{1}{n-(r+1)}\sum_{i=1}^{n}(\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2} = \frac{1}{\sqrt{n-(r+1)}}\|\hat{\epsilon} - \bar{\hat{\epsilon}}\| \quad (4.4)$$

where $\hat{\epsilon} - \bar{\hat{\epsilon}}$ represents $(\hat{\epsilon}_i - \bar{\hat{\epsilon}}, ..., \hat{\epsilon}_n - \bar{\hat{\epsilon}})$ and $\bar{\hat{\epsilon}}$ is as given in (3.13). Using properties of norms we have

$$\begin{aligned}(\hat{\sigma}_n - \sigma_n)^2 &= (\frac{1}{\sqrt{n-(r+1)}}\|\hat{\epsilon} - \bar{\hat{\epsilon}}\| - \frac{1}{\sqrt{n-(r+1)}}\|\epsilon - \bar{\epsilon}\|)^2 \\ &\le \frac{1}{n-(r+1)}\|(\hat{\epsilon} - \bar{\hat{\epsilon}}) - (\epsilon - \bar{\epsilon})\|^2 \\ &= \frac{1}{n-(r+1)}\|(\hat{\epsilon} - \epsilon) - (\bar{\hat{\epsilon}} - \bar{\epsilon})\|^2 \\ &= \frac{1}{n-(r+1)}\left[\|(\hat{\epsilon} - \epsilon)\|^2 - n(\bar{\hat{\epsilon}} - \bar{\epsilon})^2\right] \\ &\le \frac{1}{n-(r+1)}\|(\hat{\epsilon} - \epsilon)\|^2 \qquad (4.5)\end{aligned}$$

The last term in (4.5) go to zero almost everywhere by Lemma 2.4 of Freedman (1981) [9] and therefore $\hat{\sigma}_n \to \sigma_n$ almost everywhere. Now, let

$$\sigma_n^{*2} = \frac{1}{n-(r+1)}\sum_{i=1}^{n}(\epsilon_i^* - \bar{\epsilon}^*)^2 \qquad (4.6)$$

Where

$\epsilon^* = (\epsilon_1^*, ..., \epsilon_n^*)$ is as given in (2.1), $\bar{\epsilon}^* = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^*$ and let σ_n^{*2} be as given in (3.12). Then condition on $\tilde{Y}_1, ..., \tilde{Y}_n$ and utilizing the procedure in (4.5) we obtain

$$\begin{aligned}E^2(|\hat{\sigma}_n^* - \sigma_n^*|) &\le E[(\hat{\sigma}_n^* - \sigma_n^*)^2] \\ &\le E\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i^* - \epsilon_i^*)^2\right] \quad as\ in\ (4.5) \\ &\le \frac{\hat{\sigma}_n^2(r+1)}{n} \quad applying\ (3.7)\ to\ (\hat{\epsilon}^*, \epsilon^*) \\ &\to 0 \quad almost\ everywhere\ by\ (4.5). \qquad (4.7)\end{aligned}$$

It remains to show that $\hat{\sigma}_n^* \to \sigma$. By Lemma 8.6 of Bickel *et al*. (1981) [4] we have

$$\rho_1\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^{*2}, \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\right) \le \rho_1(\epsilon_i^{*2}, \epsilon_i^2) \qquad (4.8)$$

Note that ϵ_i^* has distribution $\hat{F}_n$ and ϵ_i has distribution F. By Lemma 2.6 of Freedman (1981) [9] we have $\rho_2(\hat{F}_n, F) \to 0$ almost everywhere and therefore, by Lemma 8.5 of Bickel *et al*. (1981) [4] we have $\rho_1(\epsilon_i^{*2}, \epsilon_i^2) \to 0$. It follows that $\hat{\sigma}_n^* \to \sigma$. This completes the proof of (b).

(b) follows from (a) and (b) and (3.9).

Corollary 4.2. Suppose assumptions A1 and A2 hold on the regression model (1.8). For $P = \tilde{Z}'\tilde{Z}$, let $\mathcal{F}_P^*$ and $\mathcal{F}_P$ be the distributions of $\sqrt{P}(\hat{\beta}^* - \hat{\beta}_m)$ and $\sqrt{P}(\hat{\beta}_m - \beta)$, respectively. Likewise, let $\mathcal{G}_P^*$ and $\mathcal{G}_P$ be the distributions of $\frac{\sqrt{P}(\hat{\beta}^* - \hat{\beta}_m)}{\hat{\sigma}^*}$ and $\frac{\sqrt{P}(\hat{\beta}_m - \beta)}{\hat{\sigma}}$, respectively. Condition on $\tilde{Y}_1,...,\tilde{Y}_n$ let r be fixed and $n \to \infty$. Then

(a) $\rho_2(\mathcal{F}_P^*, \mathcal{F}_P) \to 0$.
(b) $\rho_2(\mathcal{G}_P^*, \mathcal{G}_P) \to 0$.

Proof. This follows from Theorem 4.1 replacing C by $\sqrt{P}$ since $\sqrt{P}$ satisfies (3.1).

Conclusion

In this paper we provided a mathematical justification for the validity of the bootstrap for estimating the coefficient vector of a polynomial regression whose error terms are the most well-known long memory time series, namely, the autoregressive fractionally integrated moving average ARFIMA (p,d,q). Under broad conditions on the polynomial regression and standard regularity conditions on long memory error component, it is shown that a bootstrap procedure is valid. This offers guide for practitioners on utilizing the suggested bootstrap procedure of a given regression model with the specific long memory series applying techniques whose validity is supported by theory. Here the concept of validity has the meaning that the bootstrap distribution function converges to the asymptotic distribution function based on the original observations. As stated earlier, this work is an extension of the results by Freedman (1981) [9] and Eck (2018) [6], in which bootstrapping of a linear regression model with white noise errors is studied, to the case where the error terms are the well-known long memory ARFIMA (p,d,q) process. The extension is obtained in the following way.

First, we transform the original model to the case having white noise error by multiplying both sides by the polynomial $\pi(B) = (1 - B)^d\theta(B)^{-1}\phi(B)$ as shown in (1.8). Second, we estimate the parameters of the ARFIMA (p,d,q), compute an estimate $\hat{\pi}(B)$ of $\pi(B)$, Compute a modified least squares estimate $\hat{\beta}_m$ of the regression model, and compute an estimate of the white noise $\hat{\epsilon}$ using Algorithm 1.

Third, utilizing the estimated parameters we transform equation (1.8) to the case where $\tilde{Y}$ and $\tilde{Z}$ can be determined from the original model.

Fourth, we present a procedure of obtaining a bootstrap estimate $\hat{\beta}^*$ of β using Algorithm 2.

Fifth, we defined the Mallows metric and presented three related lemmas that are relevant in proving the main result of the paper.

Sixth, we state and prove the validity of the bootstrap procedure in Theorem 4.1.

References

1. Aga M, Sun T. Parametric Bootstrap Confidence Intervals for Linear Regression Processes with long-memory Errors. Sankhya: The Indian Journal of Statistics. 2007;69(4):615-634. Available from: https://www.jstor.org/stable/25664582.
2. Aga M. Bootstrap Probability Errors of the Whittle MLE for Linear Regression Processes with Strongly Dependent Disturbances. Int J Stat Prob. 2023;12(4):26-39. Available from: https://doi.org/10.5539/ijsp.v12n4p26.
3. Andrews D, Lieberman O, Marmer V. Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes. J Econometrics. 2006;133(2):673-702. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0304407605001302.
4. Bickel P, Freedman D. Some Asymptotic Theory for the Bootstrap. Ann Stat. 1981;9:1196-1217. Available from: https://www.jstor.org/stable/pdf/2240410.pdf.
5. Cochrane D, Orcutt GH. Applications of Least Squares Regressions to Relationships Containing Autocorrelated Errors. J Am Stat Assoc. 1949;44:32-61. Available from: https://www.jstor.org/stable/2280349.

6. Eck DJ. Bootstrapping Multivariate Linear Regression Models. Stat Prob Lett. 2018;134:141-149. Available from: https://doi.org/10.1016/j.spl.2017.11.001.
7. Efron B. Bootstrap Methods: another look at the jackknife. Ann Stat. 1979;7(1):1-26. DOI: 10.1214/aos/1176344552.
8. Efron B. Regression Percentiles Using Asymmetric Squared Error Loss. Stat Sinica. 1991;1(1):93-125. Available from: https://www.jstor.org/stable/24303995.
9. Freedman DA. Bootstrapping Regression Models. Ann Stat. 1981;9(6):1218-1228. DOI: 10.1214/aos/1176345638.
10. Freedman DA, Peters SC. Bootstrapping Regression Equation: Some Empirical Results. J Am Stat Assoc. 1984;79(385):97-106. Available from: https://doi.org/10.2307/2288341.
11. Granger CWJ, Joyeux R. An Introduction to Long-range Time Series and Fractional Differencing. J Time Series Anal. 1981;1(1):15-29. Available from: https://doi.org/10.1111/j.1467-9892.1980.tb00297.x.
12. Hosking JRM. Fractional Differencing. Biometrika. 1981;68(1):165-176. Available from: https://doi.org/10.2307/2335817.
13. Mallows CL. A Note on Asymptotic Joint Normality. Ann Math Stat. 1972;43(2):508-515. DOI: 10.1214/aoms/1177692631.
14. McKnight SD, McKean J, Huitema B. A Double Bootstrap Method to Analyze Linear Models with Autoregressive Error Terms. Psychol Methods. 2000;5(1):87-101. DOI: 10.1037//1082-989X.5.1.87.
15. Shao J. On Resampling Methods for Variance and Bias Estimation in Linear Models. Ann Stat. 1988;16(3):986-1008. Available from: https://www.jstor.org/stable/2241616.
16. Shorack GR. Bootstrapping robust regression. Commun Stat - Theory Methods. 1982;11(9):961-972. DOI: 10.1080/03610928208828286.
17. Shumway RH, Stoffer DS. Time Series and Its Applications. 4th ed. Springer; c2017. p. 142-145.
18. Stute W. Bootstrap of a Linear Model with AR-Error Structure. Metrika. 1995;42:395-410. Available from: https://link.springer.com/article/10.1007/BF01894336.
19. Weber NC. On resampling techniques for regression models. Stat Prob Lett. 1984;2(5):275-278. Available from: https://doi.org/10.1016/0167-7152(84)90064-6.
20. Wu CJF. Jackknife, bootstrap and other resampling methods in regression analysis (with Discussion). Ann Stat. 1986;14(4):1261-1295. Available from: https://www.jstor.org/stable/2241454.