Autoregressive Tempered Fractionally Integrated Moving Average Model

Dinesh Reddy Poddaturi

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/creativecomponents

Part of the Applied Statistics Commons, and the Longitudinal Data Analysis and Time Series Commons

Recommended Citation

https://lib.dr.iastate.edu/creativecomponents/95

This Creative Component is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Creative Components by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Autoregressive Tempered Fractionally Integrated Moving Average Model

by

Dinesh Reddy Poddaturi

A Creative Component submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Statistics

Program of Study Committee:
Farzad Sabzikar, Major Professor
Cindy Yu
Lily Wang

Iowa State University
Ames, Iowa
2018

Copyright © Dinesh Reddy Poddaturi, 2018. All rights reserved.
TABLE OF CONTENTS

LIST OF TABLES ................................................................. iv
LIST OF FIGURES ............................................................... v
ACKNOWLEDGMENTS ............................................................. vi
ABSTRACT ........................................................................... vii
CHAPTER 1. Introduction ....................................................... 1
CHAPTER 2. ARMA and ARFIMA models ................................. 3
  2.1 ARMA model .............................................................. 3
      2.1.1 Definition ......................................................... 3
      2.1.2 Causality and Invertibility .................................... 4
      2.1.3 Autocovariance .................................................. 4
      2.1.4 Spectral Density .................................................. 6
  2.2 ARFIMA model .............................................................. 6
      2.2.1 Definition ......................................................... 6
      2.2.2 Causality and Invertibility .................................... 7
      2.2.3 Autocovariance .................................................. 8
      2.2.4 Spectral Density .................................................. 8
  2.3 Long-Memory and Short-Memory Processes ...................... 8
  2.4 Whittle and Maximum likelihood estimators ...................... 9
      2.4.1 Whittle estimators ............................................. 10
      2.4.2 Maximum likelihood estimators ........................... 10
  2.5 Predictions ............................................................... 11
LIST OF TABLES

Table 4.1  Best models for average wind speed in Ames ................................. 24
Table 4.2  RMSE of predictions of average wind speed for ARTFIMA, ARFIMA with
          observed p, q .................................................................................. 26
Table 4.3  RMSE of predictions of average wind speed for ARTFIMA, ARFIMA with
          p=0, q=0 ....................................................................................... 26
Table 4.4  Best models for maximum temperature in Ames ............................ 29
Table 4.5  RMSE of predictions of maximum temperature for ARTFIMA, ARFIMA
          with observed p, q ........................................................................... 30
Table 4.6  RMSE of predictions of maximum temperature for ARTFIMA, ARFIMA
          with p=0, q=0 .................................................................................. 31
Table 4.7  Best models for Crude Oil price .................................................... 34
Table 4.8  RMSE of predictions of price of crude oil for ARTFIMA, ARFIMA with
          observed p, q ................................................................................... 35
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 4.1</td>
<td>Daily average wind speed (in MPH) in Ames from 1988 to 2018</td>
<td>23</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Periodogram of average wind speed data (circles), along with fitted ART-FIMA (light green line) and ARFIMA (light red line) spectrum with observed p and q</td>
<td>25</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Periodogram of average wind speed data (circles), along with fitted ART-FIMA (light green line) and ARFIMA (light red line) spectrum with p=0 and q=0</td>
<td>26</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Non-Standardized residuals of ARTFIMA with p=0, q=3</td>
<td>27</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Daily maximum temperature (in degree Fahrenheit) in Ames from 1988 to 2018</td>
<td>29</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Periodogram of maximum temperature data (circles), along with fitted ART-FIMA (light green line) and ARFIMA (light red line) spectrum with observed p and q</td>
<td>30</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Periodogram of maximum temperature data (circles), along with fitted ART-FIMA (light green line) and ARFIMA (light red line) spectrum with p=0 and q=0</td>
<td>31</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Non-Standardized residuals of ARTFIMA with p=0, q=0</td>
<td>32</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Crude Oil daily price from 1986 to 2018 in dollars per barrel</td>
<td>33</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Periodogram of price of crude oil (circles), along with fitted ARTFIMA (light green line) and ARFIMA (light red line) spectrum with observed p and q</td>
<td>35</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Non-Standardized residuals of ARTFIMA with p=2, q=0</td>
<td>36</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere thanks for those who helped me with various aspects of conducting research and the writing of this creative component.

First and foremost, I would like to thank Dr. Farzad Sabzikar for accepting me as his student for this creative component. His continuous guidance, patience and support throughout this research helped to finish this creative component successfully. His immense knowledge and words of encouragement inspired and renewed my hopes for completing my graduate education. I would also like to thank my committee members Dr. Cindy Yu and Dr. Lily Wang for their contributions to this work.

Finally, I would like to thank all the course instructors in the department, my family, friends, and colleagues for helping me and inspiring me for the past two years.
ABSTRACT

The objective of this creative component is to learn the ARTFIMA time series model and use it to fit real world data and compare the model with ARFIMA model in terms of goodness of fit and predictions. The ARTFIMA model is a short range dependent time series that exhibits semi-long range dependency. For small values of tempering parameter $\lambda > 0$, the spectral density behaves like a power law at low frequencies, and it remains bounded as frequency reaches to zero. ARTFIMA can be extended to Autoregressive Fractionally Integrated Moving Average (ARFIMA) when the tempered parameter $\lambda$ is zero, and to Autoregressive Moving Average (ARMA, which is most commonly used time series model) when both tempered parameter $\lambda$ and difference parameter $d$ are zero. We can consider ARTFIMA model as a generalized time series model, where as both ARFIMA and ARMA are the extensions of ARTFIMA.
CHAPTER 1. Introduction

A time series is a series of data in a set, where each data point is observed at a specific time \( t \). A discrete time series data is a set where the data are observed at discrete time points, whereas a continuous time series data is a set where the data are observed continuously over some time interval. Time series is a broad field of study. It is used in Engineering, Economics, Hydrology, Geology and more. Time series analysis is used to recognize the seasonality/trends in the data, use the data to fit a model with appropriate parameters and predict the future values.

The most commonly used model in time series analysis is Auto Regressive Moving Average (ARMA) model. This is mainly because of it’s simplistic theory. We also have more complex models like Auto Regressive Integrated Moving Average (ARIMA), Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to fit a time series data. Recently a new time series model which is called Auto Regressive Tempered Fractionally Integrated Moving Average (ARTFIMA) is introduced in (Sabzikar et al. (2015). The main objective of this creative component is to use the ARTFIMA model to fit the real world data and to compare it with ARFIMA model in terms of goodness of the fit also predictions.

This creative component has three chapters. In chapter 2 the basic time series model ARMA is introduced along with ARFIMA model. The properties causality, invertibility, autocovariance, spectral density of both ARMA and ARFIMA are discussed in this chapter. The definitions of long memory and short memory processes are defined in chapter 2. Also the parameter estimation using Maximum Likelihood Estimation and Whittle Estimation are presented along with the predictions using ARMA and ARFIMA.

Chapter 3 is dedicated for ARTFIMA model. In this chapter ARTFIMA model is defined along with it’s properties causality, invertibility, covariance function and spectral density. Parameter estimation of ARTFIMA using ML and whittle are presented. In chapter 4 the package artfima
is introduced and some of the functions in that package to fit the model. In this chapter the real world data is used to fit ARTFIMA model. For each dataset a best model is found for the data using functions in the package. ARTFIMA and ARFIMA models are used to fit the data and compared by plotting log-frequency and log-periodogram along with the fitted spectral densities. Future values are predicted by dividing the data into training and testing sets. Root Mean Squared Error (RMSE) is used to judge the performance of the predictions. An appropriate model for the data is picked by analyzing the residuals. Finally, in chapter 5 a brief summary of the creative component and some discussion are presented.
CHAPTER 2. ARMA and ARFIMA models

2.1 ARMA model

ARMA model is one of the important time series defined in terms of linear difference equations with constant coefficients (Brockwell, P. J. and Davis, R. A. (1991)). The linear structure of ARMA processes leads to a very simple theory of linear prediction which is discussed later in this chapter.

2.1.1 Definition

The time series \( \{X_t\} \) is an ARMA(p,q) process, if it is stationary and satisfies

\[
\phi(B)X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, \ldots
\]

(2.1)

where \( \phi \) and \( \theta \) are the \( p^{th} \) and \( q^{th} \) degree polynomials

\[
\phi(z) = 1 - \phi_1z - \phi_2z^2 - \ldots - \phipz^p
\]

\[
\theta(z) = 1 + \theta_1z + \theta_2z^2 + \ldots + \theta_qz^q
\]

respectively and \( B \) is a backshift operator which is defined by

\[
B^jZ_t = Z_{t-j}, \quad j = 0, \pm 1, \pm 2, \ldots
\]

(2.2)

The process \( \{Z_t\} \) is a white noise with mean 0 and variance \( \sigma^2 \). The polynomials in the definition are referred to as autoregressive (for \( \phi \)) and moving average (for \( \theta \)) of the difference equations (2.1).
2.1.2 Causality and Invertibility

2.1.2.1 Causality

An ARMA(p,q) process defined in 2.1.1 is said to be causal if there exists a sequence of constants \( \psi \) such that \( \sum_{j=0}^{\infty} |\psi_j| < \infty \) and

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots
\] (2.3)

In other words the process \( \{X_t\} \) is causal if \( X_t \) is expressed in terms of \( \{Z_s, s \leq t\} \).

2.1.2.2 Invertibility

An ARMA(p,q) process defined in 2.1.1 is said to be invertible if there exists a sequence of constants \( \pi_j \) such that \( \sum_{j=0}^{\infty} |\pi_j| < \infty \) and

\[
Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots
\] (2.4)

In other words the process \( \{X_t\} \) is invertible if \( Z_t \) is expressed in terms of \( \{X_s, s \leq t\} \).

2.1.3 Autocovariance

For any observed time series \( \{X_t\} \), the sample mean of the time series is

\[
\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t,
\]

the sample autocovariance function is

\[
\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}) (x_t - \bar{x}), \quad -n < h < n.
\] (2.5)

There are two different methods to calculate the autocovariance function of ARMA(p,q) process.

First Method: The autocovariance function \( \gamma(\cdot) \) of the causal ARMA(p,q) process defined in 2.1.1 is

\[
\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|},
\] (2.6)
where
\[
\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z) \quad \text{for } |z| \leq 1.
\] (2.7)

In order to find the coefficients \(\psi_j\) we can rewrite (2.7) as \(\psi(z)\phi(z) = \theta(z)\) and equate the coefficients of \(z^j\) to obtain
\[
\psi_j - \sum_{0<h\leq j} \phi_h \psi_{j-h} = \theta_j, \quad 0 \leq j < \max(p, q + 1),
\]
and
\[
\psi_j - \sum_{0<h\leq j} \phi_h \psi_{j-h} = 0, \quad j \geq \max(p, q + 1).
\]

Here we define \(\theta_0 = 1, \theta_j = 0 \quad \text{for} \quad j > q \quad \text{and} \quad \phi_j = 0 \quad \text{for} \quad j > p.

The above equations can be solved successively for \(\psi_0, \psi_1, \ldots\) Thus giving
\[
\psi_0 = 1,
\]
\[
\psi_1 = \theta_1 + \phi_1,
\]
\[
\psi_2 = \theta_2 + \phi_2 + \theta_1 \phi_1 + \phi_1^2, \ldots.
\] (2.8)

**Second Method**: A different method of computing the autocovariance function \(\gamma(\cdot)\) of the causal ARMA\((p, q)\) defined in 2.1.1 is based on the difference equations for \(\gamma(h), k = 0, 1, 2, \ldots\) which can be obtained by multiplying each side of the equation (2.1) by \(\{X_{t-h}\}\) and taking expectations on both sides, gives
\[
E\left[(X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p}) X_{t-h}\right] =
\]
\[
E\left[(Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}) X_{t-h}\right] \]
by solving the above equation we get
\[
\gamma(h) - \phi_1 \gamma(h-1) - \ldots - \phi_p \gamma(h-p) = \sigma^2 \sum_{h \leq j \leq q} \theta_j \psi_{j-h}, 0 \leq h < m, \quad \text{(2.9)}
\]
and
\[
\gamma(h) - \phi_1 \gamma(h-1) - \phi_2 \gamma(h-2) - \ldots - \phi_p \gamma(h-p) = 0, \quad h \geq m, \quad \text{(2.10)}
\]
here \(m = \max(p, q+1)\), \(\psi_j = 0\) for \(j < 0\), \(\theta_0 = 1, \theta_j = 0 \quad \text{for} \quad j \notin 0, 1, \ldots, q\).
2.1.4 Spectral Density

The spectral density of a stationary time series \( \{X_t\} \) is the function \( f(\cdot) \) defined by

\[
f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma(h), \quad -\pi < \omega < \pi.
\]

(2.11)

**Spectral Density of an ARMA(p,q) Process:** Let \( \{X_t\} \) be an ARMA\((p,q)\) process satisfying

\[
\phi(B)X_t = \theta(B)Z_t, \quad Z_t \sim \text{WN}(0,\sigma^2),
\]

where \( \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \) and \( \theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q \) have no common zeros and \( \phi(z) \) has no zeros on the unit circle. Then \( X_t \) has spectral density

\[
f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right|^2, \quad -\pi \leq \omega \leq \pi.
\]

(2.12)

The above equation is the ratio of trigonometric polynomials. Hence the spectral density of ARMA\((p,q)\) is often called as **rational spectral density**.

### 2.2 ARFIMA model

For any real number \( d > -1 \), the difference operator \( \nabla^d = (1 - B)^d \) is defined by means of binomial expansion,

\[
\nabla^d = \sum_{j=0}^{\infty} \pi_j B^j,
\]

(2.13)

where

\[
\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\leq j} \frac{k-1-d}{k}, \quad j = 0,1,2,\ldots,
\]

(2.14)

and \( \Gamma(\cdot) \) is the gamma function.

#### 2.2.1 Definition

The time series \( \{X_t\} \) is said to be an ARFIMA\((p,d,q)\) process with parameter \( d \in (-\frac{1}{2}, \frac{1}{2}) \), if it is stationary and satisfies the difference equations,

\[
\phi(B)\nabla^d X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, \ldots,
\]

(2.15)
where $Z_t$ is a WN($0, \sigma^2$) sequence, and $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$ and $\theta(z) = 1 + \theta z + \theta_2 z^2 + \ldots + \theta_q z^q$ are polynomials of degrees $p$ and $q$ respectively with no common zeros.

The process $\{X_t\}$ can be regarded as an ARMA($p, q$) driven by fractionally integrated noise if

$$\phi(B)X_t = \theta(B)Y_t, \quad Y_t = \nabla^{-d}Z_t.$$  \hfill (2.16)

### 2.2.2 Causality and Invertibility

If $\phi(B) \neq 0$ for $|z| = 1$, then there exists a unique and stationary solution for the process $\{X_t\}$ defined in 2.2.1, and it is given by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \nabla^{-d}Z_{t-j} \quad (2.17)$$

where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$.

#### 2.2.2.1 Causality

A solution $\{X_t\}$ of equations (2.15) is said to be causal if it can be represented as

$$X_t = \sum_{j=0}^{\infty} a_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots, \quad \sum_{j=0}^{\infty} a_j^2 < \infty. \quad (2.18)$$

Another simple definition of causality of ARFIMA($p, q$) model is:

The solution $\{X_t\}$ is causal iff $\phi(z) \neq 0$ for $|z| \leq 1$.

#### 2.2.2.2 Invertibility

A solution $\{X_t\}$ of equations (2.15) is said to be invertible if it can be represented as

$$Z_t = \sum_{j=0}^{\infty} b_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots, \quad \sum_{j=0}^{\infty} b_j^2 < \infty. \quad (2.19)$$

Another simple definition of invertibility of ARFIMA($p, q$) model is:

The solution $\{X_t\}$ is invertible iff $\theta(z) \neq 0$ for $|z| \leq 1$. 
2.2.3 Autocovariance

If \{X_t\} is causal, then \(\phi(z) \neq 0\) for \(|z| \leq 1\). This would let us write \(X_t\) as

\[
X_t = \psi(B)Y_t = \sum_{j=0}^{\infty} \psi_j Y_{t-j}
\]

(2.20)

where \(Y_t = \nabla^{-d}Z_t\).

If \(\gamma_y(\cdot)\) is the autocovariance of \(\{Y_t\}\), then the autocovariance of \(\{X_t\}\) becomes

\[
\gamma_x(h) = \sum_k \sum_j \psi_j \psi_{j+k} \gamma_y(h-k) = \sum_k \tilde{\gamma}(k) \gamma_y(h-k)
\]

(2.21)

where \(\tilde{\gamma}(k) = \sum_j \psi_j \psi_{j+k}\) is the autocovariance function of ARMA(p,q) defined in 2.1.1 with \(\sigma^2 = 1\).

2.2.4 Spectral Density

The spectral density of (2.20) can be written as

\[
f_X(\omega) = |\psi(e^{-i\omega})|^2 f_Y(\omega)
\]

where \(f_Y(\omega)\) is the spectral density of \(\{Y_t\}\), which is given by

\[
f_Y(\omega) = \frac{\sigma^2}{2\pi} |1 - e^{-i\omega}|^{-2d}
\]

and \(\psi(z)\) is defined as the ratio of MA and AR polynomials. Hence the spectral density of ARFIMA\((p,d,q)\) defined in 2.2.1 is

\[
f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} |1 - e^{-i\omega}|^{-2d}, \quad -\pi \leq \omega \leq \pi.
\]

(2.22)

2.3 Long-Memory and Short-Memory Processes

**Definition 1**: A stationary process \(\{X_t\}\) with autocovariance function \(\gamma(k)\) is said to be Long Memory (LM) if

\[
\sum_{k \in \mathbb{Z}} |\gamma(k)| = \infty.
\]

(2.23)
Short Memory (SM) if
\[ \sum_{k \in \mathbb{Z}} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k \in \mathbb{Z}} \gamma(k) > 0. \] (2.24)

**Definition 2**: A stationary process \( \{X_t\} \) with autocorrelation function \( \rho(k) \) is said to be Long Memory (LM) if
\[ \rho(k) \sim C k^{2d-1} \quad \text{as} \quad k \to 0 \] (2.25)
where \( C \neq 0 \) and \( d < 0.5 \).

Short Memory (SM) if
\[ |\rho(k)| \leq C r^{-k}, \quad k = 1, 2, 3, \ldots \] (2.26)
where \( C > 0 \) and \( 0 < r < 1 \). Here in Short Memory process the autocorrelation function is geometrically bounded.

In general an ARMA\((p, q)\) process \( \{X_t\} \) with parameters \( p \) and \( q \) is referred to as Short Memory (SM) process where as an ARIMA\((p, d, q)/ARFIMA(p, d, q)\) process \( \{X_t\} \) with parameters \( p, d, q \) is referred to as Long Memory (LM) process when \( d \in (0, \frac{1}{2}) \).

### 2.4 Whittle and Maximum likelihood estimators

Let \( \Theta \) be a subset of \( \mathbb{R}^p \), and \( Z_t \sim \text{WN}(0, \sigma^2) \) be a white noise process. For \( \theta \in \Theta \), let \( \psi_j(\theta), j = 0, 1, 2, \ldots \), be sequences of real numbers. Suppose the observations \( X_1, X_2, \ldots, X_n \) are from a linear (moving-average) process
\[ X_t = \sum_{j=0}^{\infty} \psi_j(\theta)Z_{t-j}, \quad \psi_0(\theta) = 1, \quad t \in \mathbb{Z} \] (2.27)
with \( \sum_{k=0}^{\infty} \psi_k^2(\theta) < \infty \) and \( \theta \in \Theta \).

The spectral density of the process \( \{X_t\} \) defined in (2.27) is given by
\[ f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \psi_j(\theta)e^{-ij\omega} \right|^2, \] (2.28)
where \( \theta \in \Theta \) and \( \omega \in \pi \).
The periodogram of the process \( \{X_t\} \) is

\[
I_X(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-it\omega} \right|^2.
\] (2.29)

The integrated weighted periodogram is

\[
Q_X(\theta) = \int_{\Pi} I_X(\omega) \sum_{j=0}^{\infty} \psi_j(\theta)e^{-ij\omega} d\omega = \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{n} a_{t-j}(\theta)X_tX_j,
\] (2.30)

where \( a_t(\theta) = \frac{1}{2\pi} \int_{\Pi} \frac{e^{-it\omega}}{\sum_{j=0}^{\infty} \psi_j(\theta)e^{-ij\omega}} d\omega, \quad t \in \mathbb{Z}, \theta \in \Theta. \)

The equation (2.30) is considered the integrated weighted periodogram. This quadratic form is useful for developing inference procedures on the observed data.

Let \( X_n = (X_1, \ldots, X_n) \), and let the objective function be

\[
\Lambda_n(X_n; \sigma, \theta) = \frac{1}{2\sigma^2} Q_X(\theta) + \log \sigma.
\] (2.31)

2.4.1 Whittle estimators

The whittle estimators of the true parameter values of \( \theta \), and \( \sigma \) based on \( X_n \) are defined as

\[
(\hat{\sigma}_n, \hat{\theta}_n) = \arg\min_{(\sigma, \theta) \in \Omega} \Lambda_n(X_n; \sigma, \theta).
\] (2.32)

From the above, it is clear that the estimators for \( \theta \) and \( \sigma \) are

\[
\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_X(\theta), \quad \hat{\sigma}^2 = Q_X(\hat{\theta}_n).
\] (2.33)

In finding the whittle estimators of the true parameter values, we make no assumption of Gaussianity on \( \{X_t\} \). These estimators were first discussed by Whittle, P. (1953).

2.4.2 Maximum likelihood estimators

In order to obtain the maximum likelihood estimators of the true parameter values, we assume the process \( \{X_t\} \) is gaussian and write the likelihood function \( \mathcal{L}_n \)

\[
\mathcal{L}_n = (2\pi\sigma^2)^{-n/2}(r_0 \times \ldots \times r_{n-1})^{-1/2} \exp \left\{ \frac{1}{2\sigma^2} \sum_{j=1}^{n} (X_j - \hat{X}_j)^2/r_{j-1} \right\},
\] (2.34)
and maximize $L_n$ with respect to $\theta$ and $\sigma^2$.

In the above equation $\hat{X}_j, j = 1, 2, \ldots, n$ are the one-step predictors and $r_{j-1} = E(X_j - \hat{X}_j)^2/\sigma^2$.

Upon maximizing the likelihood function in (2.34) we get

$$\hat{\sigma}^2 = n^{-1} S(\tilde{\theta}),$$

(2.35)

where $S(\tilde{\theta}) = \sum_{j=1}^{n} (X_j - \hat{X}_j)^2/r_{j-1}$ and $\tilde{\theta}$ is the value of $\theta$ which minimizes

$$l(\theta) = \ln(S(\theta)/n) + n^{-1} \sum_{j=1}^{n} \ln(r_{j-1}).$$

(2.36)

### 2.5 Predictions

#### 2.5.1 Prediction when the infinite past is given

Let $X_1, X_2, \ldots, X_n$ be generated by a stationary process

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \Psi(B) Z_t, \quad t \in \mathbb{Z},$$

where

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j,$$

(2.37)

and $Z_t$ are identically distributed uncorrelated variables with zero mean and finite variance $\sigma^2 = \text{Var}(Z_t), \sum \psi_j^2 < \infty$, and $B^j Z_t = Z_{t-j}$. Here we would like to predict $X_{n+k}$ for some $k \geq 1$. One of the simplest methods is to use the past values of $X_t(t \leq n)$ with suitable weights, $\beta_{j,k}$ on $X_t$. Hence the predicted $X_{n+k}$ can be given as

$$\hat{X}_{n+k} = \sum_{j=0}^{\infty} \beta_{j,k} X_{n-j}.$$  

(2.38)

If $X_t$ is invertible, then

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

with

$$\sum_{j=0}^{\infty} \pi_j z^j = \Psi^{-1}(z) = \left( \sum_{j=0}^{\infty} \psi_j z^j \right)^{-1}, \quad (|z| \leq 1),$$

(2.39)
the $\sigma$-algebra generated by both $X_t(t \leq n)$ and $Z_t(t \leq n)$ is same. So, $\hat{X}_{n+k}$ can be written as

$$
\hat{X}_{n+k} = \sum_{j=0}^{\infty} \alpha_{j,k} Z_{n-j}.
$$

(2.40)

In order to check the correctness of the predictions, mean squared prediction error is used

$$
\text{MSE}(k) = \text{E}[(\hat{X}_{t+k} - X_{t+k})^2].
$$

(2.41)

The best linear predictor minimizes the above equation. First we calculate the difference between $\hat{X}_{n+k}$ and $X_{n+k}$. Since $Z_t$ are uncorrelated, we have

$$
X_{n+k} - \hat{X}_{n+k} = \sum_{j=0}^{\infty} \psi_j Z_{n+k-j} - \sum_{j=0}^{\infty} \alpha_{j,k} Z_{n-j},
$$

which results in $\sum_{j=0}^{\infty} [\psi_{j+k} - \alpha_{j,k}] Z_{n-j} + \sum_{j=0}^{k-1} \psi_j Z_{n+k-j}$.

Hence,

$$
\text{MSE}(k) = \sigma^2 \sum_{j=0}^{\infty} [\psi_{j+k} - \alpha_{j,k}]^2 + \sigma^2 \sum_{j=0}^{k-1} \psi_j^2.
$$

(2.42)

The minimum MSE is achieved when $\alpha_{j,k} = \psi_{j+k}$ ($j = 0, 1, \ldots$). Thus, the optimal linear predictor is

$$
\hat{X}_{n+k} = \sum_{j=0}^{\infty} \psi_{j+k} Z_{n-j},
$$

(2.43)

and the optimal mean squared error is

$$
\text{MSE}(k) = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2.
$$

(2.44)

Equation (2.43) is not computable directly because $Z_t's, t \leq n$ are not observable. Since we assumed $X_t$ is invertible, the optimal weights in (2.38) can be obtained from

$$
\Psi^{(k)}(z)\Psi^{-1}(z) = \sum_{j=0}^{\infty} \beta_{j,k} z^k \quad (|z| \leq 1),
$$

where $\Psi^{(k)}(z) = \sum_{j=0}^{\infty} \psi_{j+k} z^j$.
2.5.2 Prediction when the finite past is given

The optimal linear prediction when a finite past is given with observations $X_1, X_2, \ldots, X_n$ is of the form

$$\hat{X}_{n+k} = \sum_{j=1}^{n} \varphi_{n,j}(k)X_{n-j+1} = [\varphi(n;k)]^T X(n)$$

(2.45)

with $X(n) = (X_n, X_{n-1}, \ldots, X_1)^T$ and $\varphi(n;k) = (\varphi_{n1}(k), \varphi_{n2}(k), \ldots, \varphi_{nn}(k))^T$ such that the MSE of $\hat{X}_{n+k}$ is minimized. From Brockwell and Davis (2016), we can obtain the optimal coefficients $\varphi_{n,j}$ from the autocovariances by the following

$$\gamma_X(k+s-1) = \varphi_{n1}(k)\gamma_X(s-1) + \ldots + \varphi_{nn}(k)\gamma_X(s-n), \quad s = 1, 2, \ldots, n$$

In the matrix notation we can write

$$\gamma_X(n;k) = (\gamma_X(k), \gamma_X(k+1), \ldots, \gamma_X(k+n-1))^T$$

and $\sum_n = [\gamma_X(i-j)]_{i,j=1,2,\ldots,n}$. This implies $\varphi(n;k) = \sum_n^{-1}\gamma_X(n;k)$. Hence the eq (2.45) is

$$\hat{X}_{n+k} = [\gamma_X(n;k)]^T \sum_n^{-1} X(n)$$

(2.46)

the mean squared prediction error is $\text{MSE}(k) = \gamma_X(0) - \gamma_X(n;k)\sum_n^{-1}\gamma_X(n;k)$

2.5.2.1 Partial autocorrelation

Partial autocorrelation is used in calculating the coefficients $\varphi(n;k)$ for $k \geq 2$ can be obtained recursively by repeated conditioning and insertion of corresponding one-step forecasts. To obtain all the coefficients we use Durbin-Levinson algorithm.

If $\hat{X}_{n+1}(2,n)$ denotes the best linear prediction of $X_n$ given $X_2, \ldots, X_n$ and $\hat{X}_1(2,n)$ denotes the best linear prediction of $X_1$ given $X_2, \ldots, X_n$, then the Partial autocorrelation at lag $n$ is defined as $\text{corr}(X_{n+1} - \hat{X}_{n+1}(2,n), X_1 - \hat{X}_1(2,n))$. This correlation is equal to the coefficient of $X_1$ in the forecast of $X_{n+1}$ in other words

$$\text{corr}(X_{n+1} - \hat{X}_{n+1}(2,n), X_1 - \hat{X}_1(2,n)) = \varphi_{nn}(1).$$

(2.47)

we use Durbin-Levinson algorithm (Brockwell, P. J. and Davis, R. A. (1991)) to calculate $\varphi(n;1)$
2.5.3 Forecasting of ARMA and ARFIMA processes

Fractional ARIMA (ARFIMA) processes are convenient when it comes to prediction because of the involvement of difference equations. ARFIMA processes are defined in terms of difference equations. This makes the computation of optimal prediction coefficients and prediction errors simple. We can write ARFIMA(p,d,q) process with \( d \) between \(-1/2\) and \(1/2\) as

\[
X_t = \psi(B)Z_t = (1 - B)^{-d} \frac{\theta(B)}{\phi(B)} Z_t
\]

with

\[
\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \sum_{j=0}^{\infty} \left(-d\right)^j \frac{1}{\phi_1 z - \ldots - \phi_p z^p} 1 + \theta_1 z + \ldots + \theta_q z^q.
\]

(2.48)

If \( d \neq 0 \) then our predictions would be for ARFIMA(p,d,q) processes. In contrast, if \( d = 0 \) our predictions would be for ARMA(p,q) processes.

The optimal coefficients \( \beta_{j,k} \), in (2.38) are defined by

\[
\Psi^{(k)}(z) \Psi^{-1}(z) = \sum_{j=0}^{\infty} \beta_{j,k} z^j
\]

\[
= \sum_{j=0}^{\infty} \psi_{j+k} z^j \sum_{j=0}^{\infty} \left(-d\right)^j \frac{1}{\phi_1 z - \ldots - \phi_p z^p} 1 + \theta_1 z + \ldots + \theta_q z^q.
\]

(2.49)

By using the notation \( \alpha_k(B) = \sum_{j=0}^{k-1} \psi_j B^j \), an alternative expressions for \( \beta_{j,k} \) can be obtained from

\[
\hat{X}_{n+k} = \sum_{j=k}^{\infty} \psi_j \hat{z}_{n+k-j}
\]

\[
= X_{n+k} - \sum_{j=0}^{\infty} \psi_j \hat{z}_{n+k-j}
\]

\[
= B^{-k} \left[ 1 - \alpha_k(B) \phi(B) \theta(B) (1 - B)^d \right] X_n
\]

\[
= \sum_{j=0}^{\infty} \beta_{j,k} x_{n-j} = \beta_k(B) X_n,
\]

which implies

\[
\sum_{j=0}^{\infty} \beta_{j,k} z^j = z^{-k} \left\{ 1 - z^{-k} \sum_{j=0}^{k-1} \psi_j z^j \sum_{j=0}^{\infty} \pi_j z^j \right\},
\]

which yields \( \beta_{j,k} = \sum_{i=0}^{k-1} \pi_i \psi_{k+j-i} \).
By using Durbin-Levinson algorithm to obtain $\varphi(n;k)$, we can get predictions based on the finite past.
CHAPTER 3. ARTFIMA model

The ARTFIMA model is first introduced in Sabzikar, Meerschaert, and Chen (2014). This model extends the Tempered Fractional Integrated (TFI) model from Giraitis, Kokoszka and Leipus (2000). According to Giraitis, Koul, and Surgailis (2012), TFI time series exhibit a semi-long dependence. Depending on the tempering parameter, the covariance function echoes long range dependence for many lags, but finally it decays exponentially fast.

The tempered fractional difference (TFD) operator is defined by

\[
\nabla^{d,\lambda} X_t = \left(1 - e^{-\lambda B}\right)^d X_t = \sum_{j=0}^{\infty} \eta_j^{d,\lambda} X_{t-j},
\]

where \(d, \lambda > 0\), \(B\) is the back shift operator, and

\[
\eta_j^{d,\lambda} = (-1)^j \left(\begin{array}{c} d \\ j \end{array}\right) e^{-\lambda j}
\]

where \(\left(\begin{array}{c} d \\ j \end{array}\right) = \frac{\Gamma(1+d)}{j!\Gamma(1+d-j)}\), and \(\Gamma(\cdot)\) is a gamma function and it is given by \(\Gamma(t) = \int_0^{\infty} e^{-x} x^{t-1} dx\).

We can extend (3.1) to a non-integer values of \(d < 0\), by using the property of Gamma function, \(\Gamma(d) = (d - 1)\Gamma(d - 1)\). If \(\lambda = 0\) the equation (3.1) reduces to the equation (2.13).

3.1 Definition

The time series \(\{X_t\}\) is said to be an ARTFIMA\((p, d, \lambda, q)\) process with parameters \(d \notin \mathbb{Z}\) and \(\lambda > 0\), if it satisfies the difference equations,

\[
\phi(B)\nabla^{d,\lambda} X_t = \theta(B)Z_t, \quad t \in \mathbb{Z},
\]

where \(Z_t\) is i.i.d white noise sequence with mean 0 and constant variance (i.e., \(E[Z_t] = 0\) and \(E[Z_t^2] = \sigma^2\), \(d \notin \mathbb{Z}\)), \(\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p\), and \(\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q\) are polynomials with degrees \(p, q \geq 0\) and no common zeros.
From the above definition we can say that \( \{X_t\} \) is ARTFIMA\((p,d,\lambda,q)\) iff \( Y_t = \nabla^{d,\lambda} X_t \) is an ARMA\((p,q)\) time series
\[
Y_t - \sum_{j=1}^{p} \phi_j Y_{t-j} = Z_t + \sum_{i=1}^{q} Z_{t-i}, \quad t \in \mathbb{Z},
\]
where \( \{Z_t\} \) is a white noise sequence for \( t \in \mathbb{Z} \).

### 3.2 Causality and Invertibility

In order to define *causality* and *invertibility* we assume that the AR and MA polynomials have no common zeros and
\[
|\phi(z)| > 0 \quad \text{and} \quad |\theta(z)| > 0 \quad \text{for} \quad |z| \leq 1 \quad (3.5)
\]

#### 3.2.1 Causality

Let \( \{X_t\} \) be an ARTFIMA\((p,d,\lambda,q)\) time series defined in 3.1 and it also satisfies the assumptions in (3.5). Then \( \{X_t\} \) is *causal*, that is it can be represented as
\[
X_t = \sum_{j=0}^{\infty} \psi_j^{d,\lambda} Z_{t-j}, \quad \text{where} \quad \sum_{j=0}^{\infty} |\psi_j^{d,\lambda}| < \infty. \quad (3.6)
\]

#### 3.2.2 Invertibility

Let \( \{X_t\} \) be an ARTFIMA\((p,d,\lambda,q)\) time series defined in 3.1 and it also satisfies the assumptions in (3.5). Then \( \{X_t\} \) is *invertible*, that is it can be represented as
\[
Z_t = \sum_{j=0}^{\infty} \pi_j^{d,\lambda} X_{t-j}, \quad \text{where} \quad \sum_{j=0}^{\infty} |\pi_j^{d,\lambda}| < \infty. \quad (3.7)
\]

### 3.3 Spectral Density

Let \( \{X_t\} \) be an ARTFIMA\((p,d,\lambda,q)\) time series defined in 3.1 and it also satisfies the assumptions in (3.5). Then the spectral density of \( \{X_t\} \) is given by
\[
f_X(\nu) = \frac{\sigma^2}{2\pi} |1 - e^{-(\lambda + i\nu)}|^{-2d} \frac{|\theta(e^{-i\nu})|^2}{|\phi(e^{-i\nu})|^2}, \quad \text{for} \quad -\pi \leq \nu \leq \pi, \quad (3.8)
\]
where \( \sigma^2 \) is the variance of white noise.
3.4 Covariance

Before computing the covariance of ARTFIMA\((p, d, \lambda, q)\) process, first we recall an important form of the spectral density for the ARMA\((p, q)\) process defined in 2.1.1. Under the assumptions in (3.5), \(\phi(x)\) can be written as

\[
\phi(x) = \prod_{j=1}^{p} (1 - \rho_j x)
\]

where \(\rho_1, \rho_2, \ldots, \rho_p\) are the complex numbers such that \(|\rho_n| < 1\) for \(n = 1, 2, \ldots, p\) (according to Sowell, F. (1992) [Section 4]). The spectral density of the ARMA\((p, q)\) process can be written as

\[
f_X(\nu) = \frac{\sigma^2}{2\pi} \frac{|\theta(\eta)|^2}{|\phi(\eta)|^2} \prod_{j=1}^{p} (1 - \rho_j \eta)^{-1}(1 - \rho_j \eta^{-1})^{-1},
\]

where \(\eta = e^{-i\nu}\). According to Sowell, F. (1992) [Section 4], the spectral density of ARMA\((p, q)\) process defined in 2.1.1 can also be written as

\[
f_X(\nu) = \frac{\sigma^2}{2\pi} \sum_{l=-q}^{q} \psi(l) \eta^l \sum_{j=1}^{p} \eta^p \zeta_j \left(\frac{\rho_j^{2p}}{(1 - \rho_j \eta)} - \frac{1}{(1 - \rho_j^{-1})}\right),
\]

where

\[
\psi(l) = \sum_{s=\max[0,l]}^{\min[q,q+l]} \theta_s \theta_{s-l},
\]

and

\[
\zeta_j = \frac{\sigma^2}{2\pi} \left[\rho_j \prod_{i=1}^{p} (1 - \rho_i \rho_j) \prod_{m \neq j \leq m \leq p} (\rho_j - \rho_m)^{-1}\right].
\]

Suppose \(\{X_t\}\) is an ARTFIMA\((p, d, \lambda, q)\) time series defined in 3.1 and it also satisfies the assumptions in (3.5). Using the notation in (3.10), the covariance function of \(\{X_t\}\) is

\[
\gamma_X(k) = \frac{\sigma^2}{2\pi} \sum_{l=-q}^{q} \sum_{j=1}^{p} \psi(l) \zeta_j C(d, \lambda, k - l - p, \rho_j),
\]

where

\[
C(d, \lambda, h, \rho) = \rho^{2p} \sum_{m=0}^{\infty} \rho^m e^{-\lambda(h-m)} \Gamma(h - m + d) \frac{\Gamma(h - m + d; h - m + 1; e^{-2\lambda})}{\Gamma(d)\Gamma(0)} + \sum_{n=1}^{\infty} \rho^n e^{-\lambda(h+n)} \Gamma(h + n + d) \frac{\Gamma(h + n + d; h + n + 1; e^{-2\lambda})}{\Gamma(d)\Gamma(h + n + 1)}
\]
where the Gaussian hypergeometric function

$$
\pFq{2}{1}{\sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+j)\Gamma(j+1)}z^j} = 1 + \frac{a \cdot b}{c \cdot 1}z + \frac{a(a+1)b(b+1)}{c(c+1)\cdot 1 \cdot 2}z^2 + \ldots
$$
is defined for all complex numbers $a$ and $b$, all complex $|z| < 1$ and real $c$ not a negative integer.

This is discussed in Sabzikar, Mcleod, and Meerschaert (2018).

### 3.5 Parameter Estimation

From (3.8), the spectral density of ARTFIMA($p,d,\lambda,q$) can be written as

$$
f_X(\nu; \theta) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{-i\nu})}{\phi(e^{-i\nu})} \right|^2 \left| 1 - e^{-(\lambda+i\nu)} \right|^{-2d} = \frac{\sigma^2}{2\pi} K(\nu, \theta), \quad \text{for} \quad \nu \in (-\pi, \pi), \quad (3.12)
$$

where $\sigma > 0$, and $\theta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, d, \lambda)$.

Suppose $X = (X_1, X_2, \ldots, X_N)$ are the observed values of the ARTFIMA($p,d,\lambda,q$) time series with a sample size $N$. We can write the periodogram of $X$ as

$$
I_X(\nu) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} X_t e^{it\nu} \right|^2. \quad (3.13)
$$

Define a quadratic form as below

$$
Q_X(\theta) = \int_{-\pi}^{\pi} \frac{I_X(\nu)}{K(\nu, \theta)} d\nu, \quad (3.14)
$$

and

$$
D_N(X, \sigma, \theta) = \frac{1}{2\sigma^2} Q_X(\theta) + \log \sigma. \quad (3.15)
$$

Suppose $\sigma_0$ and $\theta_0$ are the true parameter values of $\sigma \in (0, \infty)$ and $\theta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, d, \lambda) \in \Xi$, where $\Xi = \mathbb{R}^{p+q+1} \times (0, \infty)$. Let $\Omega = (0, \infty) \times \Xi$, the Whittle estimators are as given below

#### 3.5.1 Whittle estimators

The Whittle estimators of $\sigma_0$ and $\theta_0$ of ARTFIMA($p,d,\lambda,q$) are given as

$$
(\hat{\sigma}_N, \hat{\theta}_N) = \arg\min \{ D_N(X, \sigma, \theta; (\sigma, \theta) \in \Omega) \}, \quad (3.16)
$$
which gives \( \theta_N = \arg\min_{\theta \in \Xi} Q_X(\theta) \) and \( \bar{\sigma}_N^2 = Q_X(\theta_N) \).

To obtain consistency of the parameters, we assume the parameter space is restricted to a compact set \( \Omega_0 \subset \Omega \). We chose this set so that our true parameters \((\sigma_0, \theta_0)\) is an interior point.

Over the compact space \( \Omega_0 \), the Whittle estimators (3.16) are strongly consistent (Sabzikar, Mcleod, and Meerschaert (2018)).

\[
\lim_{N \to \infty} \bar{\theta}_N = \theta_0 \quad a.s. \\
\lim_{N \to \infty} \bar{\sigma}_N^2(\bar{\theta}_N) = \sigma_0^2 \quad a.s.
\] (3.17)

The Whittle estimators are asymptotically normal over the compact space \( \Omega_0 \). That is,

\[
\sqrt{N}(\bar{\theta}_N - \theta_0) \xrightarrow{d} N(0, W^{-1}),
\] (3.18)

where

\[
W = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \frac{\partial \log K(\nu, \theta_0)}{\partial \theta} \right\} \left\{ \frac{\partial \log K(\nu, \theta_0)}{\partial \theta} \right\}' d\nu.
\] (3.19)

3.5.2 Maximum likelihood estimators

Given the observed \( X = (X_1, X_2, \ldots, X_N) \), the maximum likelihood estimator \( \hat{\theta}_N \) for \( \theta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, d, \lambda) \) can be calculated by taking the logarithm of the likelihood function

\[
l(X_1, X_2, \ldots, X_N) = (2\pi\sigma^2)^{-N/2}|G_N|^{-1/2}\exp\left[\frac{-X_N' G_N^{-1} X_N}{2\sigma^2}\right]
\] (3.20)

where \( G_N = \frac{1}{\sigma^2} E[X_N X_N'] \) and \( |G_N| \) is the determinant of \( G_N \).

Under the same assumptions made in 3.5.1 the maximum likelihood estimators are strongly consistent and are given by

\[
\lim_{N \to \infty} \hat{\theta}_N = \theta_0 \quad a.s. \\
\lim_{N \to \infty} \hat{\sigma}_N^2(\hat{\theta}_N) = \sigma_0^2 \quad a.s.
\] (3.21)

Using the same assumptions in 3.5.1, the maximum likelihood estimators are asymptotically normal. That is,

\[
\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, W^{-1})
\] (3.22)
where $W$ is given by (3.19)

CHAPTER 4. ARTFIMA package and Applications

In this chapter real datasets are used to fit ARTFIMA\((p,d,\lambda,q)\) time series model using the R package \texttt{artfima}. Authors of this package are A.I.McLeod, Mark M. Meerschaert, and Farzad Sabzikar (2016). This package is freely available on CRAN (available at \url{https://CRAN.R-project.org/package=artfima}). Also in this chapter, we present the best models to fit the datasets along with the predictions and how well the models ARFIMA and ARTFIMA perform with predictions. RMSE is used to test the accuracy of the predictions. In order to find the RMSE of the predictions the data are divided into training and testing data. Training data are used to fit ARFIMA, and ARTFIMA models. The obtained results are then used to predict the next values in the time series. Once we predict, we use the new values to compare to the testing data and get RMSE (Root Mean Squared Error) of our predictions for both ARFIMA and ARTFIMA models.

The function \texttt{bestModels} (in the package) is used to find the best models (ARTFIMA, ARFIMA, ARIMA) along with the parameters \(p, q\) based on AIC and BIC criterion. By using \texttt{best_glpmodels}, we can find the best model with respective parameters for a specific (ARTFIMA, ARFIMA, ARIMA) model. In this function one needs to give the required model ARIMA/ARFIMA/ARTFIMA and maximum order of AR component (\(p\)) and maximum order of MA component (\(q\)). For each combination of \(p\) and \(q\), Log-Likelihood, AIC, and BIC values are calculated and reported by the function. By observing the AIC and BIC values we can chose a best possible \(p\) and \(q\) for the model. The function \texttt{predict.artfima} is used to predict the next \(n\) subsequent values in the time series.
4.0.1 Daily Average Wind Speed in Ames

This data contains the average wind speed (in MPH) in Ames recorded daily from 1988 to 2018. The data are available at https://www.ncdc.noaa.gov/cdo-web/datasets. Figure 4.1 shows the time series plot of the wind speed from 1988 to 2018 in Ames.

![Recorded Average Wind Speed in Ames (1998-2018)](image)

Figure 4.1 Daily average wind speed (in MPH) in Ames from 1988 to 2018

First the stationarity of the data are tested by Augmented Dickey-Fuller Test. The null hypothesis of this test is that a unit root of a univariate time series \( \{X_t\} \) exists (equivalently, \( \{X_t\} \) is non-stationary time series). The alternative hypothesis is the time series \( \{X_t\} \) is stationary. Large \( p \) values indicate non-stationarity in the data, small \( p \) values indicate the data are stationary. From the below output the \( p \) value is smaller than 0.01. Assuming a significance level 0.01, we can conclude that the data are stationary.

```r
> adf.test(ames_Wind)

Augmented Dickey–Fuller Test

data: ames_Wind
Dickey–Fuller = -10.372, Lag order = 19, p-value = 0.01
alternative hypothesis: stationary
```
Warning message:
In adf.test(ames_Wind) : p-value smaller than printed p-value

But the figure 4.1 clearly says otherwise. We can observe a strong seasonality in the data with little to no significant trend. We can observe a significant seasonal mean variation and heteroscedasticity with respect to each season. The data are made stationary by subtracting the seasonal mean and dividing by the seasonal standard deviation. The standardized data is used for further analysis.

Table 4.1 shows the best models for the data obtained by using bestModels. The table contains the best models according to AIC and BIC criterion. From the table it is evident that the best model according to AIC is ARFIMA(1,0,2) and according to BIC is ARFIMA(0,0,1). If we observe closely for most models the AIC and BIC values are approximately equal.

| Table 4.1 Best models for average wind speed in Ames |
|-----------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| AIC models | Best | 2nd Best | 3rd Best | 4th Best |
| ARFIMA(1,0,2) | 20136.67 | 20139.59 | 20141.03 | 20141.22 |
| ARFIMA(2,0,0) | 20141.03 | 20141.22 | 0.113 | 0.103 |
| ARFIMA(0,0,1) | 20141.03 | 20141.22 | 0.113 | 0.103 |
| ARFIMA(0,0,3) | 20141.22 | 0.103 | 0.103 | 0.103 |
| AIC | 1.000 | 0.232 | 0.113 | 0.103 |
| p(AIC) | 1.000 | 0.232 | 0.113 | 0.103 |
| BIC | 20168.64 | 20172.36 | 20174.10 | 20176.34 |
| BIC | 1.000 | 0.155 | 0.065 | 0.021 |

Using best_glp_models the best ARTFIMA model is ARTFIMA(0,0,3) with AIC 20136.98, ARFIMA(0,0,1) with BIC 20183.08. When we observe closely the AIC value of ARTFIMA(0,0,3) is approximately equal to all other AIC values in the table 4.1. Using the package we fit the data with p=0, q=3 for ARTFIMA and p=1, q=2 for ARFIMA models using the Whittle estimator. Here the parameter d is fixed to 5/6 for ARTFIMA (from Kolmogorov scaling) model.

The resulted parameter fits for ARTFIMA are $d = 0.8333$ (standard error = 0.000), $\lambda = 0.13691$ (standard error = 0.0399), $\theta = \{0.4067, 0.2808, 0.0487\}$, and for ARFIMA $d = 0.10559$ (standard error = 0.0245), $\phi = \{0.64529\}$, $\theta = \{0.43218, 0.21308\}$. 
The Figure 4.2 shows the periodogram and fitted ARTFIMA\((p,d,\lambda,q)\) and ARFIMA\((p,d,q)\) spectral density function for the data with fixed \(d\) (\(d=5/6\)) and observed parameters. By looking at the figure the ARTFIMA spectral density follows the data at low frequencies because of the tempering parameter. Whereas, the spectral density of the ARFIMA is linear at low frequencies. The same figure is plotted for ARTFIMA\((p,d,\lambda,q)\) and ARFIMA\((p,d,q)\) but with \(p=0\), \(q=0\) for both ARTFIMA and ARFIMA models. The observed \(\lambda\) for ARTFIMA when \(p=0\), \(q=0\), \(d=5/6\) is 2.9944. Figure 4.3 shows the periodogram and fitted spectral density lines for \(p=0\) and \(q=0\) with fixed \(d\).

![Figure 4.2](image)

Figure 4.2  Periodogram of average wind speed data (circles), along with fitted ARTFIMA (light green line) and ARFIMA (light red line) spectrum with observed \(p\) and \(q\)

### 4.0.1.1 Predictions

By using the function `predict.artfima` in the package `artfima` the future values of the time series can be predicted. The data are divided into train and test datasets. Training dataset is used to fit a model, the estimated parameters are used to predict the future values. The RMSE values for both ARTFIMA and ARFIMA in tables 4.2 are same where as the RMSE when \(p\) and \(q\) are zeroes for both ARTFIMA and ARFIMA in table 4.3 are higher than the RMSE in the table 4.2.
Figure 4.3  Periodogram of average wind speed data (circles), along with fitted ARTFIMA (light green line) and ARFIMA (light red line) spectrum with p=0 and q=0

Table 4.2  RMSE of predictions of average wind speed for ARTFIMA, ARFIMA with observed p, q

<table>
<thead>
<tr>
<th></th>
<th>ARTFIMA</th>
<th>ARFIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.9993055</td>
<td>0.9992338</td>
</tr>
</tbody>
</table>

An appropriate model is picked by further analyzing the residuals for each fit. That is the residuals of each ARTFIMA and ARFIMA with observed p, q, d, λ (for ARTFIMA) and p=0, q=0, d, λ are analyzed. The residuals of ARTFIMA model with p=0, q=3, d = 0.833, λ = 0.13691 appear to be independent and normal. Figure 4.4 shows the non standardized residuals of ARTFIMA(0,0.833,0.13691,3). The independence is tested by Ljung-Box test with null hypothesis being all the residuals are independent, and alternative hypothesis being the residuals are not independent. Below is the output of the test. From the output the p value is 0.9546. Assuming a significance level 0.5, the p value is greater than the significance level. Hence we fail to reject the null hypothesis and conclude that the residuals are independent.

Table 4.3  RMSE of predictions of average wind speed for ARTFIMA, ARFIMA with p=0, q=0

<table>
<thead>
<tr>
<th></th>
<th>ARTFIMA</th>
<th>ARFIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>10.34087</td>
<td>10.32894</td>
</tr>
</tbody>
</table>
> Box.test(residuals_WINDSpeed, type = "Ljung-Box")

Box–Ljung test

data: residuals_WINDSpeed

X-squared = 0.0032341, df = 1, p-value = 0.9546

Figure 4.4 Non-Standardized residuals of ARTFIMA with p=0, q=3

From spectral density plots, RMSE values of the predictions, and residual analysis, ARTFIMA model with p=0 and q=3 provides a reasonable fit to the data.
4.0.2 Daily Maximum Temperature in Ames

This data contains the maximum temperature (in degree Fahrenheit) in Ames recorded daily from 1988 to 2018. The data are available at https://www.ncdc.noaa.gov/cdo-web/datasets. Figure 4.5 shows the time series plot of the maximum temperatures from 1988 to 2018 in Ames.

From Augmented Dickey-Fuller test (with alternative hypothesis as the data are stationary, null hypothesis as the data are not stationary) the $p$ value is smaller than 0.01 (from the below output). Assuming a significance level 0.01, we reject the null hypothesis and we conclude that the data are stationary.

```r
> adf.test(ames_MAXt)

Augmented Dickey-Fuller Test

data: ames_MAXt
Dickey-Fuller = -5.0388, Lag order = 19, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ames_MAXt) : p-value smaller than printed p-value
```

But when we closely observe the time time series plot fig 4.5, we see there is a strong seasonality in the data. There is no evidence of trend in this data, but there is a significant mean variation and heteroscedasticity with respect to each season. Hence we subtract the seasonal mean and divide by the seasonal standard deviation. This would standardize the time series. The analysis is done using the standardized time series data.

Table 4.4 shows the best models for the data obtained by using bestModels. The table contains the best models according to AIC and BIC criterion. According to AIC criterion the best model is ARTFIMA with $p=1$, $q=1$ with AIC = 16662.65. According to BIC criterion the best model is ARFIMA with $p=1$, $q=1$ with BIC = 16697.50.
Figure 4.5  Daily maximum temperature (in degree Fahrenheit) in Ames from 1988 to 2018

Table 4.4  Best models for maximum temperature in Ames

<table>
<thead>
<tr>
<th></th>
<th>Best</th>
<th>2nd Best</th>
<th>3rd Best</th>
<th>4th Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC models</td>
<td>ARTFIMA(1,0,1)</td>
<td>ARFIMA(1,0,1)</td>
<td>ARFIMA(2,0,1)</td>
<td>ARFIMA(1,0,2)</td>
</tr>
<tr>
<td>AIC</td>
<td>16707.03</td>
<td>16707.24</td>
<td>16707.96</td>
<td>16708.00</td>
</tr>
<tr>
<td>p(AIC)</td>
<td>1.000</td>
<td>0.899</td>
<td>0.628</td>
<td>0.615</td>
</tr>
<tr>
<td>BIC models</td>
<td>ARFIMA(1,0,1)</td>
<td>ARFIMA(2,0,0)</td>
<td>ARTFIMA(1,0,1)</td>
<td>ARFIMA(2,0,1)</td>
</tr>
<tr>
<td>BIC</td>
<td>16741.75</td>
<td>16743.38</td>
<td>16748.44</td>
<td>16749.37</td>
</tr>
<tr>
<td>p(BIC)</td>
<td>1.000</td>
<td>0.442</td>
<td>0.035</td>
<td>0.022</td>
</tr>
</tbody>
</table>

In this analysis we use AIC criterion to choose the best models and fit the data. Using the package we fit the data with \( p=1, q=1 \) for ARTFIMA and \( p=1, q=1 \) for ARFIMA models using the Whittle estimator.

The estimated parameter fits for ARTFIMA are \( d = 0.20214 \) (standard error = 0.01980), \( \lambda = 0.0046692 \) (standard error = 0.00390), \( \phi = \{0.33920\}, \theta = \{-0.13667\} \), and for ARFIMA \( d = 0.18995 \) (standard error = 0.01980), \( \phi = \{0.35518\}, \theta = \{-0.13228\} \).

The Figure 4.6 shows the periodogram and fitted ARTFIMA\((p,d,\lambda,q)\) and ARFIMA\((p,d,q)\) spectral density function for the data with estimated parameters. There is a clear distinction between ARTFIMA and ARFIMA models in the figure. The spectral density of ARTFIMA follows the data at low frequencies but the spectral density of ARFIMA is linear. Having a tempering parameter \( \lambda \) in ARTFIMA causes a deviation of spectral density at low frequencies. The same figure
is plotted for ARTFIMA(p,d,λ,q) and ARFIMA(p,d,q) but with p=0, q=0 for both ARTFIMA and ARFIMA models. The estimated d and λ for ARTFIMA when p=0, q=0 are 1.14029 (standard error = 0.04800) and 0.52687 (standard error = 0.0352) respectively. The estimated d for ARFIMA with p=0, q=0 is 0.4895 (standard error = 1.223115e-07). Figure 4.7 shows the periodogram and fitted spectral density lines for p=0 and q=0 with estimated d for ARFIMA and ARTFIMA.

![Figure 4.6 Periodogram of maximum temperature data (circles), along with fitted ARTFIMA (light green line) and ARFIMA (light red line) spectrum with observed p and q](image)

**4.0.2.1 Predictions**

The tables 4.5 and 4.6 provide the RMSE values of the predictions for ARTFIMA and ARFIMA with observed p, q and with p=0 , q=0. The RMSE value of ARTFIMA in both cases is lower than the RMSE of ARFIMA. Therefore, we can conclude that the predictions using ARTFIMA are more accurate than the predictions using ARFIMA.

<table>
<thead>
<tr>
<th>ARTFIMA</th>
<th>ARFIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.003099</td>
</tr>
</tbody>
</table>
Figure 4.7  Periodogram of maximum temperature data (circles), along with fitted ART-FIMA (light green line) and ARFIMA (light red line) spectrum with p=0 and q=0

Table 4.6  RMSE of predictions of maximum temperature for ARTFIMA, ARFIMA with p=0, q=0

<table>
<thead>
<tr>
<th></th>
<th>ARTFIMA</th>
<th>ARFIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.004642</td>
<td>1.003142</td>
</tr>
</tbody>
</table>

The residuals for each ARTFIMA and ARFIMA model are analyzed to pick the appropriate model for the data. The residuals of ARTFIMA model with p=1, q=1 and p=0, q=0 appear to be independent and normal. Figure 4.8 shows the non standardized residuals of ARTFIMA(0, 1.14029, 0.52687, 0). By Ljung-Box test the independence of residuals can be tested. Below is the output of the test for ARTFIMA (p=0, q=0). The p value is greater than 0.05. Hence we fail to reject the null hypothesis of the residuals are independent and conclude that the residuals appear to be independent.

> Box.test(maxtemp_ARTFIMA0$res, type = "Ljung-Box")

Box-Ljung test

**data:** maxtemp_ARTFIMA0$res
X-squared = 0.67933, df = 1, p-value = 0.4098

Figure 4.8  Non-Standardized residuals of ARTFIMA with p=0, q=0

From the figures 4.6, 4.7, the tables 4.5, 4.6, and residual diagnostics ARTFIMA with p=0, q=0 (the simplest model) provides an appropriate fit to the data.
4.0.3 Crude Oil Prices

This data contains the price (Dollars per Barrel) of Crude oil recorded daily in Cushing, Oklahoma from 1986 to 2018. The data are available at https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm. Figure 4.9 shows the plot of the daily prices from 1986 to 2018.

From the time series plot, we can observe that there is a trend in the data. We test the stationarity of the data by a formal Augmented Dickey-Fuller test. Below is the output from the test. Here the null hypothesis is that the data are not stationary and the alternative hypothesis is the data are stationary. Assuming a significance level 0.05, the observed $p$ value is 0.2831 which is way larger than 0.05. Hence we fail to reject the null hypothesis and conclude that the data are not stationary. The data is made stationary by taking one difference using the function `diff`.

```r
> adf.test(crudeOil_Price)

Augmented Dickey-Fuller Test

data: crudeOil_Price
Dickey-Fuller = -2.697, Lag order = 20, p-value = 0.2831
alternative hypothesis: stationary
```
Table 4.7 shows the best models for the data obtained by using bestModels. The table contains the best models according to AIC and BIC criterion. From the table it is evident that the best model according to AIC is ARIMA(2,0,2) and according to BIC is ARIMA(0,0,1). If we observe closely for most models the AIC and BIC values are approximately equal. The best ARFIMA model is ARFIMA(2,0,0) with AIC = 25393.46.

<table>
<thead>
<tr>
<th></th>
<th>Best</th>
<th>2nd Best</th>
<th>3rd Best</th>
<th>4th Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC models</td>
<td>ARIMA(2,0,2)</td>
<td>ARIMA(4,0,0)</td>
<td>ARIMA(2,0,0)</td>
<td>ARIMA(3,0,0)</td>
</tr>
<tr>
<td>AIC</td>
<td>25379.90</td>
<td>25391.98</td>
<td>25393.46</td>
<td>25393.55</td>
</tr>
<tr>
<td>p(AIC)</td>
<td>1.000</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>BIC models</td>
<td>ARIMA(0,0,1)</td>
<td>ARIMA(1,0,0)</td>
<td>ARIMA(2,0,2)</td>
<td>ARIMA(2,0,0)</td>
</tr>
<tr>
<td>BIC</td>
<td>25419.40</td>
<td>25420.22</td>
<td>25422.00</td>
<td>25422.10</td>
</tr>
<tr>
<td>p(BIC)</td>
<td>1.000</td>
<td>0.667</td>
<td>0.273</td>
<td>0.260</td>
</tr>
</tbody>
</table>

Using best_glpm_models the best ARTFIMA model is ARTFIMA(2,0,0) with AIC 25394.93 and ARTFIMA(0,0,0) with BIC 25427.96. The AIC value of ARTFIMA(2,0,0) is less compared to all other AIC values in the table 4.7.

Using the package we fit the data with p=2, q=0 for ARTFIMA and p=2, q=0 for ARFIMA models using the Whittle estimator. The resulted parameter fits for ARTFIMA are \( d = 0.0335, \lambda = 0.134, \phi = \{-0.0713, -0.0434\} \), and for AFIMA are \( d = 0.0259, \phi = \{-0.0680, -0.0434\} \).

The figure 4.10 shows the periodogram and fitted ARTFIMA(p,d,\lambda,q) and ARFIMA(p,d,q) spectral density function for the data and observed parameters. Here the spectral density of both ARTFIMA and ARFIMA follows the data closely. Here for this stationary data we obtained an adequate fit by including the AR component with p=2, like we got from the bestModels and best_glpm_models.
Figure 4.10  Periodogram of price of crude oil (circles), along with fitted ARTFIMA (light green line) and ARFIMA (light red line) spectrum with observed p and q

4.0.3.1 Predictions

The tables 4.8 provide the RMSE values of the predictions for ARTFIMA and ARFIMA with observed p and q, that is p=2 and q=0. The prediction RMSE value of ARTFIMA is very less than that of ARFIMA.

Table 4.8  RMSE of predictions of price of crude oil for ARTFIMA, ARFIMA with observed p, q

<table>
<thead>
<tr>
<th></th>
<th>ARTFIMA</th>
<th>ARFIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.976987</td>
<td>0.9771284</td>
</tr>
</tbody>
</table>

The residuals for each ARTFIMA and ARFIMA (with p=2, q=0) model are analyzed to pick the appropriate model for the data. The residuals of both ARTFIMA and ARFIMA appear to be independent. Figure 4.11 shows the non standardized residuals of ARTFIMA with p=2, q=0. The independence of residuals is tested by Ljung-Box test. From the output below, we can see that the p value of the test is 0.9265 which is greater than 0.05. Hence we fail to reject the null hypothesis (residuals are independent) and conclude that the residuals are independent.

> Box.test(crudeoil_ARTFIMA$res)
Box–Pierce test

data: crudeoIL_ARTFIMA08.res

X-squared = 0.0085156, df = 1, p-value = 0.9265

Figure 4.11 Non-Standardized residuals of ARTFIMA with p=2, q=0

From the figure 4.10 both ARTFIMA and ARFIMA models seems appropriate. Whereas, the prediction RMSE of ARTFIMA is less than the prediction RMSE of ARFIMA. From the residual dignostics along with the spectral density plots and RMSE of predictions ARTFIMA with p=2, q=0 provides a reasonable fit to the data.
This creative component discusses the basic time series model ARMA, the complex models ARFIMA and ARTFIMA. The properties (causality, invertibility, covariance, spectral density) of ARTFIMA are presented in detail. The goal of this creative component is to learn about ARTFIMA model and use the model to fit real world data. The package `artfima` helps to fit ARMA, ARFIMA, and ARTFIMA models. The function `bestModels` in the package allows to pick a best model for the data, the function `best_glp_models` helps to find a specific (ARMA/ ARFIMA/ARTFIMA) model with appropriate parameters. The performance of the predictions are judged using RMSE for ARFIMA and ARTFIMA. In most cases ARTFIMA outperformed ARFIMA in terms of the predictions also in terms of AIC and BIC values. In general the ARTFIMA model seems appropriate for hydrology, geology and atmospheric data. Although ARTFIMA fits for different types of data, in some situations TFI (Tempered Frational Integrated) model is appropriate for fitting the model and predicting the future values. One of the limitations of this study is that the residuals of all the best models provided by the package are not independent. We need to further analyze the residuals to pick the best model for the data. Another limitation is regarding the residuals provided by the package. The residuals are not recognized by any other packages. The residuals are of type numeric in this package. We have to convert the residuals to time series object in order to further analyze them. Another limitation is with AIC and BIC values. As we observed in chapter 4 most of the times the AIC and BIC values for the best models are approximately equal. Eventhough having approximate AIC and BIC values would allow us to chose a simple model over complex model, providing the best models with different AIC and BIC values would give us an opportunity to choose different models.
REFERENCES


A.I. McLeod, Mark M. Meerschaert, and Farzad Sabzikar (2016) *artfima: ARTFIMA Model Estimation* https://CRAN.R-project.org/package=artfima,