On Estimators of a Spectral Density Function

Chengyin Wang

A Thesis
in the Department of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Mathematics) at
Concordia University
Montréal, Québec, Canada

May 2018

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Chengyin Wang**

Entitled: **On Estimators of a Spectral Density Function** and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Examiner

Dr. Arusharka Sen

_____ Examiner

Dr. Wei Sun

_____ Thesis Supervisor

Dr. Yogendra P. Chaubey

Approved by _____

Chair of Department or Graduate Program Director

_____

Dean of Faculty

Date _____

To my family, who got me this far.

. . .

Abstract

This thesis presents two main approaches to estimating the spectral density of a stationary time series, that are based on the classical periodogram. Both of these are related to the non-parametric density estimation. One is the kernel spectral density estimator while the other one is the Bernstein polynomial spectral density estimator. We have also introduced the method to determine the optimal smoothing parameters for estimating the spectral density of a stationary zero-mean process. Finally, the thesis concludes with a simulation study in order to examine the finite sample properties of the proposed spectral density estimators.

Acknowledgements

First of all, I am greatly indebted to my thesis supervisor, Dr. Yogendra Chaubey, who helped me with enrolling to the graduate program. Without his patient assistance and friendly encouragement, it would not be possible for me to learn lots of information about this topic in such a short period of time and complete this thesis without reducing its scholarly quality. His willingness to give me his time so generously has been much appreciated. Truly, without his painstaking efforts in revising and polishing my drafts, the completion of the present thesis would not have been possible. With his help, I earned valuable experience.

My sincere thanks should also go to all my professors of the Department of Mathematics and Statistics at the University of Concordia, because they taught me lots of required and related academic knowledge form their courses.

Last but not the least, I thank my beloved family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my classmates who gave me their help and time in listening to me and helping me work out my problems during the difficult course of the thesis.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Time Series

Definition of A Time Series

Time series is a set of observations $\{X_t, t \in \mathbb{Z}\}$, each one being recorded at a specific time t. It is a collection of sample values corresponding to different random variables. A time series can be discrete or continuous. A *discrete time series* is one in which the set of times at which observations are made is a discrete set. *Continuous time series* are obtained when observations are recorded continuously over time interval (see Brockwell & Davis, 1991). In general, if a time series contains a single variable is termed as $uni-variate$. Otherwise, It is termed as $multivariate$.

Time Series Analysis

Time series analysis is the procedure of fitting a time series to a proper model (see Hipel & McLeod, 1994). The basic objective of time series analysis is to un-

derstand the underlying context of the data through the use of stochastic models to forecast future values and to simulate based on previously observed values. Methods for time series analysis may be divided into two classes: frequency-domain method and time-domain method. The first one includes spectral-analysis and wavelet-analysis, and the other one includes auto-correlation and cross-correlation analysis. In real life, time series analysis are used in many different areas such as signal processing, mathematical finance, weather forecasting and so on.

1.2 Basic Concepts of Time Series

Most of the following materials are based on the textbooks of Priestley (1981) and Brockwell and Davis (1991).

Stationary Processes

In general, we have two types of the time series: stationary and non-stationary. **Stationary** series vary around a constant mean, neither decreasing nor increasing systematically over time, with a constant variance. **Non-stationary** series have systematic trends, such as linear, quadratic, and so on. In this thesis, we only focus on the stationary time series.

Definition 1. *Let* $\{X_t, t \in \mathbb{Z}\}$ *be a time series with* $\mathbb{E}(X_t^2) < \infty$*. The* **mean function** *of* $\{X_t\}$ *is*

$$\mu_X(t) = \mathbb{E}(X_t). \tag{1.2.1}$$

The **covariance function** *of* $\{X_t\}$ *is*

$$\begin{aligned} \gamma_X(r,s) &= Cov(X_r, X_s) \\ &= \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))], \quad r,s \in \mathbb{Z} \end{aligned} \tag{1.2.2}$$

Definition 2. *The time series* $\{X_t, t \in \mathbb{Z}\}$ *is* **(Weakly) stationary** *if*
(i) $\mu_X(t)$ *is independent of* t*,*
(ii) $\gamma_X(t+h,t)$ *is independent of* t *for each* h*.*

In other words, a stationary time series $\{X_t; t \in \mathbb{Z}\}$ must have these features: finite variance, constant first moment, and the second moment only depends on h and independent of t.

Remark 1. Strict stationary time series is defined by the condition that $(X_1, ..., X_n)$ and $(X_{1+h}, ..., X_{n+h})$ have the same joint distributions for all integers h and $n > 0$. Whenever term **stationary** is used we shall mean weakly stationary as in $Definition$ 2,(see Brockwell & Davis, 1991).

Definition 3. *If* $\{X_t, t \in \mathbb{Z}\}$ *is a stationary time series, then the* **auto-covariance function (ACVF)** *is defined by*

$$\begin{aligned} \gamma_X(h) &= Cov(X_{t+h}, X_t) \\ &= \mathbb{E}[(X_{t+h} - \mu_X(t+h))(X_t - \mu_X(t))], \quad t,h \in \mathbb{Z} \end{aligned} \tag{1.2.3}$$

Note that, for $h = 0$*, the auto-covariance reduces to the variance, that is*

$$\gamma_X(0) = \mathbb{E}(X_t - \mu_X(t))^2 = Var(X_t). \tag{1.2.4}$$

Time Series Models

The basic building block for all processes considered is the **white noise process**.

Definition 4. (White Noise Process) *A white noise process is a sequence* $\{\epsilon_t, t \in \mathbb{Z}\}$ *whose elements have zero mean and variance* σ^2*,*

$$\mathbb{E}(\epsilon_t) = 0, \;\; \mathbb{E}(\epsilon_t^2) = \sigma^2, \tag{1.2.5}$$

and for which the ϵ' *s are uncorrelated*

$$\mathbb{E}(\epsilon_t \epsilon_s) = 0 \;\; for \;\; t \neq s. \tag{1.2.6}$$

If ϵ*'s are independent, the sequence is called independent white noise process. Furthermore, if the* ϵ*'s are normally distributed*

$$\epsilon_t \sim N(0, \sigma^2) \tag{1.2.7}$$

is called Gaussian white noise process.

The selection of a proper model for time series data is very important as it reflects the underlying structure of the time series and this fitted model in turn is used for future forecasting. In the main, these models can have many forms and represent different stochastic processes. The most frequently used time series models in the literature are the Moving-Average (MA) Model, the Auto-Regressive (AR) Model, the Auto-Regressive Moving Average (ARMA) Model and the Auto-Regressive Integrated Moving Average (ARIMA) Model.

Definition 5. *(***Moving-Average Model)** *A p-th order moving average process, denoted* $MA(p)$ *is a stochastic process* $\{X_t\}$ *characterized by*

$$X_t = \epsilon_t + \sum_{i=1}^{p} \theta_i \epsilon_{t-i}, \tag{1.2.8}$$

where θ_i *is real number, and* ϵ_t *is white noise.*

Definition 6. (Auto-Regressive Model) *A q-th order auto-regressive process, denoted* $AR(q)$ *is a stochastic process* $\{X_t\}$ *characterized by*

$$X_t = \sum_{j=1}^{q} \phi_j X_{t-j} + \epsilon_t, \tag{1.2.9}$$

where ϕ_j *is any real number and the* ϵ_t *is an independent white noise process.*

The $ARMA$ model is a composite of AR and MA models.

Definition 7. (Auto-Regressive Moving Average Model) *An $ARMA$ model is a stochastic process $\{X_t\}$ characterized by:*

$$X_t = \sum_{j=1}^{q} \phi_j X_{t-j} + \sum_{i=1}^{p} \theta_i \epsilon_{t-i} + \epsilon_t. \tag{1.2.10}$$

where ϕ_j and θ_i are any real number and the ϵ_t is an independent white noise process.

Definition 8. (Auto-Regressive Integrated Moving Average Model) *The $ARIMA$ model is a generalization of an $ARMA$ model. $ARIMA$ models are applied in some cases where data is non-stationary. The model of $ARIMA(p, d, q)$ is defined as:*

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon_t, \tag{1.2.11}$$

where L is the lag operator, the ϕ_i are the parameters of the auto-regressive part of the model, and the θ_i are the parameter of the moving average part and the ϵ_t are error terms.

In the present thesis, we mainly consider non-parametric estimation of the spectral density function. The non-parametric smoothing methods provide a powerful methodology for estimating the spectral density function of stationary processes. The non-parametric approaches estimate the covariance or the spectrum of the process without assuming that the process has any particular structure. There exist a great many different non-parametric density estimation methods, for example, histogram estimation, kernel density estimation, maximum penalized likelihood estimators, etc.

1.3 Spectral Methods

The Fourier Transform

Let $\omega \in (-\pi, \pi)$ denote the *frequency*, and T is the *period* , that is, $T = \frac{2\pi}{\omega}$. Given a time series $\{X_t; t \in \mathbb{Z}\}$, its *Fourier transformation* is:

$$X(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} e^{i\omega t} X(t) \tag{1.3.1}$$

and the *inverse transform* is:

$$X(t) = \int_{-\pi}^{\pi} X(\omega) e^{-i\omega t} d\omega. \tag{1.3.2}$$

The Spectral Density Function

Definition 9. (Spectral Density Function) *Let $\{X_t; t \in \mathbb{Z}\}$ be a real value and zero-mean stationary time series with the auto-covariance function, $\gamma_X(h) = \gamma_X(-h) = Cov(X_t X_{t+h}) = \mathbb{E}(X_t X_{t+h})$, satisfying that*

$$\sum_{h=-\infty}^{+\infty} |\gamma_X(h)| < \infty, \; h \in \mathbb{Z}. \tag{1.3.3}$$

The **spectral density function** *of $\{X_t; t \in \mathbb{Z}\}$ can be written as*

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\omega} \gamma_X(h) && (1.3.4) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \cos(\omega h) \gamma_X(h) && (1.3.5) \\ &= \frac{1}{2\pi} \gamma_X(0) + \frac{1}{\pi} \sum_{h=1}^{\infty} cos(\omega h) \gamma_X(h) && (1.3.6) \end{aligned}$$

where $e^{i\omega} = \cos(\omega) + i \sin(\omega)$ and $i = \sqrt{-1}$.

There are some properties of the spectral density function:

1) $f_X(\omega)$ is even,

2) $f_X(\omega)$ is non-negative for all $\omega \in [-\pi, \pi]$, and $\int_{-\pi}^{\pi} f_X(\omega)d\omega = 1$,

3) $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\omega} f_X(\omega)d\omega = \int_{-\pi}^{\pi} \cos(h\omega) f_X(\omega)d\omega$.

Note that since cos is a periodic function with the period 2π, the range of values of the spectral density is determined by the value of $f_X(\omega)$ for $\omega \in [0, \pi]$. (see Brockwell & Davis, 1991)

1.4 Literature Review

Many techniques for spectral density have been established in the literature estimation. At the turn of the century, the most commonly used methods are based on the periodogram which was introduced by Arthur Schuster in 1898 when he was searching for hidden periodicities while studying sunspot data (see Schuster, 1898). Since the periodogram is inconsistent, then many scientists were concerned with the modification of the periodogram to find a consistent form such as Bartlett (1948), Daniell (1946), Grenander (1951), Parzen (1962) and so on.

Other alternative famous estimators, such as kernel density estimator (see Rosenblatt, 1956; Parzen, 1962), have received much attention at the same time. More recently, Kakizawa (2004, 2006) proposed the Bernstein polynomial approximation method to estimate a spectral density function by properly choosing different kernel estimates.

1.5 Thesis outline

The thesis is organized as follows: in Chapter 2, we define the algorithms for the several classical estimation methods. In Chapter 3, we provide details of the Bernstein spectral density estimator as well as the cross-validation for the selection of the smoothing parameter. In Chapter 4, simple transformation methods using kernel density estimator and Bernstein polynomial estimator on real line are explained as an alternative. In Chapter 5, a comparison of estimators given in the previous chapters is provided through simulation. In Chapter 6, we give summary of the thesis and provides some directions for future research work.

CHAPTER 2

Traditional Smoothed Estimators

2.1 Classical Estimator - The Periodogram

In practice, to estimate the spectral density function, we may replace the theoretical auto-covariances by the sample auto-covariances to equation (1.3.4). Thus, based on T observations $X_1, X_2, ..., X_T$ from X_t at equally spaced interval in time, the periodogram can be defined for all ω in the range $\omega \in [-\pi, \pi]$ by

$$I_T(\omega) = \frac{1}{2\pi T}|\sum_{t=1}^{T} e^{-it\omega}X_t|^2 \tag{2.1.1}$$

$$= \frac{1}{2\pi}\hat{\gamma}_T(0) + \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)\cos(\omega h), \tag{2.1.2}$$

where

$$\hat{\gamma}_T(h) = T^{-1}\sum_{t=1}^{T-|h|} X_t X_{t+|h|},\ |h| \leq T-1 \tag{2.1.3}$$

is the sample auto-covariance of lag h. We assume through out that $I_T(\omega)$ is continuous for all ω, this will certainly hold if $\hat{\gamma}(s)$ is absolutely summable. It shows that the function $I_T(\omega)$ is even and 2π periodic.

We have the following asymptotic distribution of the periodogram,

$$I_T(\omega) \sim \frac{f_X(\omega)\chi_2^2}{2}. \tag{2.1.4}$$

Since $\mathbb{E}(\chi_2^2) = 2$, the sample periodogram provides an unbiased estimate of the spectrum, $\lim_{T\to\infty} \mathbb{E}(I_T(\omega)) = f(\omega)$. However, the variance of $I_T(\omega)$ does not go to zero (see Brillinger, 1981). In fact,

$$Var(I_T(\omega)) = \begin{cases} 2f_X^2(0), & \omega = 0 \\ f^2(\omega), & \text{otherwise.} \end{cases}$$

Therefore, $I_T(\omega)$ is not a consistent estimator of $f_X(\omega)$in mean square (see Priestley, 1981).

Consistent estimators of $f_X(\omega)$ can be obtained by smoothing the periodogram. Let the Fourier frequencies of the sample be defined as $\omega_k = \frac{2\pi k}{T}$, $j \in \mathbb{Z}$. The basic idea of the simple smoothed periodogram can be expressed as the average neighboring value estimate:

$$f_X^*(\omega) = \frac{1}{n}\sum_k I_T(\omega_k). \tag{2.1.5}$$

Furthermore, there are several classical methods to have a consistent estimate. For example, Brillinger (1981) proposed to smooth the data $\{(\omega_k, I_T(\omega_k))\}, k = 1, 2...\frac{n-1}{2}$ directly through a weighted local average; another famous method is to smooth the log periodogram data $\{(\omega_k, log(I_T(\omega_k)))\}$ through the least square method (see Wahba, 1980), etc.

2.2 Kernel Smoothed Estimator

Kernel Density Estimator

The kernel method for density estimator was introduced by Rosenblatt (1956) and Parzen (1962) which is defined as follows,

Definition 10. *Let $\{X_t, t \in \mathbb{N}\}$ be a random sample from a continuous distribution function F with density function f, its kernel density estimator is*

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j) \tag{2.2.1}$$

$$= \frac{1}{nh}\sum_{j=1}^{n} K\Big(\frac{x - X_j}{h}\Big) \tag{2.2.2}$$

where K is a symmetric density with zero mean and unit variance, non-negative and real-valued kernel function.

The *bandwidth* $h > 0$ which depends on n such that $h \to 0$ and $nh \to \infty$ as the sample size $n \to \infty$. Several types of kernels are commonly used such as $Uniform$, *Parabolic*, *Biweight* and *Gaussian* and so on.

Remark 2. In particular, if we consider estimation of the density for circular data, i.e. an absolutely continuous circular density $f(\theta)$, $\theta \in [-\pi, \pi]$ which is 2π-periodic,

$$f(\theta) \geq 0 \ \ for \ \ \theta \in \mathbb{R} \ \ and \ \ \int_{-\pi}^{\pi} f(\theta)d\theta = 1 \tag{2.2.3}$$

Given a random sample $\{\theta_1, \theta_2, ..., \theta_n\}$ for the above density, thus the kernel density estimator may be written as (see Chaubey, 2017)

$$\hat{f}(\theta) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{\theta - \theta_i}{h}\Big). \tag{2.2.4}$$

Kernel Spectral Density Estimator

Kernel based estimator of the spectral density is the weighted sum of autocovariance, in which weights are decided by the *kernel K* and *bandwidth h*. More generally, we can consider **Lag Window Estimator**, that is,

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{s=1-T}^{T-1} \lambda(s)\hat{\gamma}(s)e^{-i\omega s} \tag{2.2.5}$$

(see Priestley, 1981), where $\{\lambda(.)\}$ is called lag window.

According to (see Priestley, 1981), if we apply the properties of Fourier transforms, the lag window density estimator may be written as an integrated version of periodogram,

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I_T(\theta)W(\omega - \theta)d\theta, \tag{2.2.6}$$

where $W(.)$ is the **spectral window**,

$$W(\theta) \;=\; \frac{1}{2\pi} \sum_{s=-(T-1)}^{T-1} \lambda(s)e^{-i\theta s} \tag{2.2.7}$$

(see Brockwell & Davis, 1991; Priestley, 1981)

Remark 3. For practical purposes, instead of the integral we will rather use the discrete sum over all Fourier frequency, which is

$$\hat{f}(\omega) \approx \frac{1}{2\pi} \sum_{j=-N}^{N} W(\omega - \omega_j)I_T(\omega_j). \tag{2.2.8}$$

where N is the largest integer less or equal to $\frac{T-1}{2}$.

There are lots of different possible lag windows that would fulfill the conditions to obtain a consistent estimate of the spectral density (see Priestley, p. 434). A rather convenient type of lag windows are the scale parameter windows (see Priestley, p. 446). These contain a parameter, the scale parameter, that in some way controls for the width of the window.

Definition 11. *$\lambda(.)$ is a scale parameter window if*

$$\lambda(s) = k(s/M) \tag{2.2.9}$$

where $k(.)$ is a lag window generator with $k(0) = 1$ and M is the scale parameter. Its Fourier transform

$$K(\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} k(u)e^{-iu\theta}du, \tag{2.2.10}$$

is called a spectral window generator.

Thus, the relationship between the $W(\theta)$ and $K(\theta)$ is

$$\begin{aligned} W(\theta) &= \frac{1}{2\pi}\sum_{s=-(T-1)}^{T-1}\lambda(s)e^{-i\theta s} \\ &= M[\frac{1}{2\pi M}\sum_{s=-\infty}^{\infty}k(s/M)e^{-i(s/M)M\theta}] \\ &= M\sum_{s=-\infty}^{\infty}K[M(\theta+2\pi s)] \\ &\sim MK(M\theta). \end{aligned} \tag{2.2.11}$$

Remark 4. However, in particular the Lomnicki-Zaremba window, the Whittle window, the Daniells window, and the 'exact' form of the Bartlett window cannot be written in the form (2.2.9), so that these windows are not of the 'scale parameter' form.

Therefore, the periodogram can be smoothed by convolving with a spectral kernel weight K to obtain the kernel spectral density estimator $f(\omega)$ from equation (2.2.6)

$$\hat{f}(\omega) = \int_{-\pi}^{\pi}\frac{1}{h}K[\frac{\omega-\theta}{h}]I_T(\theta)d\theta, \tag{2.2.12}$$

For the kernel estimation of spectral density, the following kernel windows are commonly used (see Priestly, 1981):

1. Truncated kernel

$$k(z) = \mathbb{1}(|z| \leq 1)$$

Its Fourier transform

$$K(u) = \frac{sinu}{\pi u}.$$

2. Bartlett kernel

$$k(z) = (1 - |z|)\mathbb{1}(|z| \leq 1)$$

Its Fourier transform

$$K(u) = \frac{1}{2\pi}\Big(\frac{sin(u/2)}{u/2}\Big)^2.$$

3. Daniell kernel

$$k(z) = \frac{sin(\pi z)}{\pi z}$$

Its Fourier transform

$$K(u) = \begin{cases} \dfrac{1}{2\pi}, & \text{if } |u| \leq \pi, \\ 0, & \text{otherwise.} \end{cases}$$

4. Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & \text{if } |z| \leq 1/2, \\ 2(1 - |z|)^3, & 1/2 \leq |z| < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Its Fourier transform

$$K(u) = \frac{3}{8\pi}\Big(\frac{sin(u/4)}{u/4}\Big)^4.$$

5. Quadratic-Spectral kernel (Epanechnikov)

$$k(z) = \frac{3}{(\pi z)^2}[\frac{\sin(\pi z)}{\pi z} - \cos(\pi z)]$$

Its Fourier transform

$$K(u) = \frac{3}{4\pi}[1 - (\frac{u}{\pi})^2]\mathbb{1}(|u| \leq \pi)$$

Bandwidth Selection

Smoothed estimators of the spectral density, relies on the choice of a bandwidth or lag number depending on the sample size. Selecting the bandwidth h is a very important step in estimating density function. If the bandwidth chosen is too large, the kernel density estimator is over-smoothed and key aspects of the true density may not be revealed. On the other hand, if the bandwidth is too small, then the kernel density estimator is under-smoothed (see Mugdadi and Jetter, 2010). The shape of the kernel must be specified but has little effect on the resulting estimate compared to the choice of h (see Silverman 1986). In summary, while there exists many bandwidth selectors, no rule comes with a guarantee that it will work satisfactorily in all cases. Different techniques and details have been proposed by Wand and Jones (1994).

CHAPTER 3

Alternative Smooth Estimator Based on Bernstein Polynomial

3.1 Introduction

Theorem 1. (Weierstrass Approximation Theorem) *Suppose f is a continues real-valued function defined on the real interval $[a, b]$. For every $\epsilon > 0$, there exists a polynomial p such that for all x in $[a, b]$, we have $|f(x) - p(x)| < \epsilon$, or equivalently, the supremum norm $||f - g|| < \epsilon$.*

It is well-known that the Bernstein polynomial is a useful tool for interpolating functions defined on a closed interval and can be used to approximate a density function defined on such an interval. Bernstein polynomials were first introduced by Bernstein (1912) who developed them as a probabilistic proof of the *Weierstrass* Approximation Theorem. He showed that for any continuous function can be uniformly approximated by Bernstein polynomials. The $Bernstein - Weierstrass$ Approximation Theorem assures that as the degree of the polynomial increase to infinity the Bernstein polynomial approximation

coverages uniformly to the true function. Bernstein-based approaches to other problems of non-parametric function have also been developed by different authors. Vitale (1975) considered the smooth estimate of a probability density function with a finite support $[0, 1]$. Babu et al. (2002) investigated the asymptotic properties of using Bernstein polynomials to approximate bounded and continuous density functions. Kakizawa (2006) considered the Bernstein polynomial can be used as a non-parametric prior for continuous densities. Then Kakizawa (2006) modified the Bernstein polynomial to estimate spectral density function.

3.2 Bernstein Probability Density Estimator

Let G be an any continuous function on the closed interval $[x_0, x_0 + \Delta]$, then the associated Bernstein polynomial is

$$B(x) = \sum_{j=0}^{m} G\Big(x_0 + \frac{j\Delta}{m}\Big) b_{j,m}\Big(\frac{x - x_0}{\Delta}\Big) \tag{3.2.1}$$

converges to $G(x)$ uniformly in $x \in [x_0, x_0 + \Delta]$ as $m \to \infty$, where

$$G(x) = \int_{x_0}^{x} g(u)du, \tag{3.2.2}$$

and

$$b_{j,m}(y) = \binom{m}{j} y^j (1 - y)^{m-j}, \;\; y \in [0, 1]. \tag{3.2.3}$$

Differentiating the formula of $B(x)$, we obtain

$$\begin{aligned} B'(x) &= \sum_{j=0}^{m} G\Big(x_0 + \frac{j\Delta}{m}\Big) \frac{d}{dx} b_{j,m}\Big(\frac{x - x_0}{\Delta}\Big) \\ &= \frac{m}{\Delta} \sum_{j=0}^{m-1} \Big\{ G\Big(x_0 + \frac{(j+1)\Delta}{m}\Big) - G\Big(x_0 + \frac{j\Delta}{m}\Big) \Big\} b_{j,m-1}\Big(\frac{x - x_0}{\Delta}\Big), \end{aligned} \tag{3.2.4}$$

$B'(x)$ converges to $G'(x) = g(x)$ uniformly in $x \in [x_0, x_0 + \Delta]$ as $m \to \infty$, where $g(x)$ is assumed to be continuous in $[x_0, x_0 + \Delta]$ (see Kakizawa, 2006).

Remark 5. In particular, without loss of generality, for the interval $[0,1]$, the Bernstein polynomial is

$$B(x) = \sum_{j=0}^{m} G\Big(\frac{j}{m}\Big)\binom{m}{j}x^j(1-x)^{m-j}, \tag{3.2.5}$$

which converges to $G(x)$ uniformly as m is large; and its derivative

$$B'(x) = m\sum_{j=0}^{m-1}\Big\{G\Big(\frac{j+1}{m}\Big) - G\Big(\frac{j}{m}\Big)\Big\}b_{j,m-1}(x) \tag{3.2.6}$$

converges to $g(x)$ under the same condition.

Note that the asymptotic properties of $B(x)$ and $B'(x)$ are examined in a paper by Babu, Canty and Chaubey (2002).

3.3 Bernstein Spectral Density Estimator

Let's extend the use of the Bernstein polynomials to the spectral density function on the closed interval $\theta \in [-\pi, \pi]$. Since the spectral density function is even and periodic, it is sufficient to confine ourselves to the study of the interval $\theta \in [0, \pi]$. Once an estimator $\hat{f}(\theta)$ of a spectral density function $f(\theta)$ is constructed for all $\theta \in [0, \pi]$, it can be extended to $\theta \in \mathbb{R}$, in such a way that we set $\hat{f}(\theta) = \hat{f}(|\omega|)$ when $\theta = \omega + 2\pi v$ for some $\omega \in (-\pi, \pi]$ and $v \in \mathbb{Z}$ (see Lorentz, 1986).

We may use the linear $transformation$ to θ:

$$x = \frac{\theta}{\pi} \in [0,1], \tag{3.3.1}$$

then apply the Bernstein polynomial approximation theory. It is possible that to apply the Bernstein polynomial for the spectral density $f(\theta)$ itself directly,

then we will have (see Kakizawa, 2006)

$$B(\theta) = \sum_{j=0}^{m} f\Big(\frac{\pi j}{m}\Big)\binom{m}{j}\Big(\frac{\theta}{\pi}\Big)^{j}\Big(1-\frac{\theta}{\pi}\Big)^{m-j} \tag{3.3.2}$$

$$\approx f(\theta), \quad \theta \in [0,\pi]. \tag{3.3.3}$$

We already know the property of the spectral density from the Chapter 1. Since the periodogram is inconsistent estimator of the spectral density $f(\omega)$, although it is, for each $\omega \in [-\pi,\pi]$, an asymptotically unbiased estimator of $f(\omega)$. Specifically, if $f(\omega)$ satisfies a Lipshitz condition of order 1, i.e. if $|f(\alpha) - f(\beta)| \leq C|\alpha-\beta|$ as $|\alpha-\beta| \to 0$ and C being a constant, then implies,

$$\mathbb{E}[I_T(\omega)] = \begin{cases} f(\omega) + O\Big(\dfrac{\log T}{T}\Big), & \text{if } \gamma = 1; \quad (3.3.4) \\ f(\omega) + O(T^{-\gamma}), & \text{if } \gamma < 1. \quad (3.3.5) \end{cases}$$

both uniformly in $\omega \in [-\pi,\pi]$ (see Hannan, 1970; Priestley, 1981).

Moreover, the periodogram is used to estimate the linear function of

$$L(f) = \int_{-\pi}^{\pi} A(\omega) f(\omega) d\omega \tag{3.3.6}$$

$$= \int_{-\pi}^{\pi} \frac{1}{2}\{A(\omega) + A(-\omega)\} f(\omega) d\omega \tag{3.3.7}$$

for given function $A(\omega)$. If we substitute the periodogram for the spectral density in this function, $L(I_T)$ or its discrete average at the points $\frac{2\pi j}{T}$ is an asymptotically unbiased, $T^{1/2}$-consistent and asymptotically normal estimator of $L(I_T)$. (see Brillinger, 1981; Hosoya and Taniguchi, 1982).

In particular, for the spectral distribution function, i.e. *cumulative spectrum* is:

$$F(\theta) = \int_{-\pi}^{\theta} f(\omega) d\omega, \tag{3.3.8}$$

then we set

$$F^{+}(\theta) = 2\int_{0}^{\theta} f(\omega) d\omega, \tag{3.3.9}$$

since $f(\omega)$ is even, which is naturally estimated by substituting the periodogram,

$$\begin{aligned}
\hat{F}_T^+(\theta) &= 2\int_0^\theta I_T(\omega)d\omega \\
&= 2\int_0^\theta \Big\{\frac{1}{2\pi}\hat{\gamma}_T(0) + \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)\cos(\omega h)\Big\}d\omega \\
&= 2\int_0^\theta \frac{1}{2\pi}\hat{\gamma}_T(0)d\omega + 2\int_0^\theta \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)\cos(\omega h)d\omega \\
&= \frac{\theta}{\pi}\hat{\gamma}_T(0) + \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)\int_0^\theta \cos(\omega h)d\omega \\
&= \frac{\theta}{\pi}\hat{\gamma}_T(0) + \frac{2}{\pi}\sum_{h=1}^{T-1}\frac{1}{h}\hat{\gamma}_T(h)\sin(\theta h). \qquad (3.3.10)
\end{aligned}$$

$\hat{F}_T^+(\theta)$ is an asymptotically unbiased estimate of $F^+(\theta)$, it follows that $\hat{F}_T^+(\theta)$ is a consistent estimate of $F^+(\theta)$ (see Priestly, 1981).

Then we can propose for $\theta \in [0, \pi]$,

$$\begin{aligned}
\hat{f}_{T,m}(\theta) &= \frac{m}{2\pi}\sum_{j=0}^{m-1}\Big\{\hat{F}_T^+\Big(\frac{(j+1)\pi}{m}\Big) - \hat{F}_T^+\Big(\frac{j\pi}{m}\Big)\Big\}\, b_{j,m-1}\Big(\frac{\theta}{\pi}\Big) \\
&= \frac{m}{\pi}\sum_{j=0}^{m-1} b_{j,m-1}\Big(\frac{\theta}{\pi}\Big)\int_{\frac{j\pi}{m}}^{\frac{(j+1)\pi}{m}} I_T(\theta)d\theta \\
&= \frac{m}{\pi}\sum_{j=0}^{m-1} b_{j,m-1}\Big(\frac{\theta}{\pi}\Big)\int_{\frac{j\pi}{m}}^{\frac{(j+1)\pi}{m}} \frac{1}{2\pi}\hat{\gamma}_T(0) + \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)cos(\theta h)]d\theta \\
&= \frac{m}{\pi}\sum_{j=0}^{m-1} b_{j,m-1}\Big(\frac{\theta}{\pi}\Big)\Big\{\frac{1}{2m}\hat{\gamma}_T(0) + \frac{1}{\pi}\sum_{h=1}^{T-1}\hat{\gamma}_T(h)\frac{1}{h}\sin\Big(\frac{(j+1)\pi}{m}\Big)\Big\}
\end{aligned} \qquad (3.3.11)$$

as an estimator of $f(\theta)$. (see Kakizawa, 2006)

3.4 Bernstein-Kernel Spectral Density Estimator

Recall that for Bernstein spectral density estimator:

$$\hat{f}(\theta) = \frac{m}{\pi} \sum_{j=0}^{m-1} b_{j,m-1}\Big(\frac{\theta}{\pi}\Big) \int_{\frac{j\pi}{m}}^{\frac{(j+1)\pi}{m}} I_T(\theta) d\theta \tag{3.4.1}$$

From the Chapter 2, the kernel spectral density function is:

$$\hat{f}_K(\omega) = \int_{-\pi}^{\pi} I_T(\theta) W(\omega - \theta) d\theta, \tag{3.4.2}$$

where $W(\omega) \approx MK(M\omega)$.

Then we define Bernstein's generalized kernel spectral density estimator

$$\hat{f}_{BGK}(\theta) = \sum_{j=0}^{m-1} b_{j,m-1}\Big(\frac{\theta}{\pi}\Big) \hat{f}_K(x_j) \tag{3.4.3}$$

$$= \int_{\frac{j\pi}{m}}^{\frac{(j+1)\pi}{m}} \phi(\lambda) I_T(\lambda) d\lambda, \ \theta \in [0, \pi] \tag{3.4.4}$$

(see Kakizawa, 2006)

where

$$\phi(\lambda) = 2m \sum_{j=0}^{m-1} K\{2m(\lambda - x_j)\} \mathbb{1}(\lambda; B) b_{j,m-1}(\frac{\theta}{\pi}), \tag{3.4.5}$$

and $\mathbb{1}(\lambda; B)$ is the indicator function,

$$\mathbb{1}(y; B) \begin{cases} 1; & \text{if } y \in B; \quad (3.4.6) \\ 0; & \text{otherwise.} \quad (3.4.7) \end{cases}$$

Example 1. For example, the Daniell-kernel spectral density estimator is defined as follows:

$$\hat{f}_{Daniell}(\lambda) = \int_{\lambda - \frac{\pi}{h}}^{\lambda + \frac{\pi}{h}} I_T(\omega) \frac{h}{2\pi} d\omega. \tag{3.4.8}$$

It may noted that this estimator at the frequency 0 and π but for the Daniell-kernel spectral density estimator at frequency $\frac{\pi}{2m}$ and $\pi - \frac{\pi}{2m}$:

$$\hat{f}(0) = \hat{f}_{Daniell}\Big(\frac{\pi}{2m}\Big) \tag{3.4.9}$$

and

$$\hat{f}(\pi) = \hat{f}_{Daniell}\left(\pi - \frac{\pi}{2m}\right) \tag{3.4.10}$$

Therefore,

$$\hat{f}(\lambda) = \sum_{j=0}^{m-1} b_{j,m-1}\left(\frac{\lambda}{\pi}\right)\hat{f}_{Daniell}(x_j) \tag{3.4.11}$$

where $x_j = -\pi + \frac{(j+1/2)\pi}{m}$ for all $j = 0, ..., m-1$.

3.5 The Selection of the Degree m

Beltrao and Bloomfield (1987) provide the first objective criterion for the selection of the parameter in the area of cross-validation methods. They argue that, by minimizing a cross-validation version of the log-likelihood function (CVLL), one will also minimize the mean square integrated error, which is what they propose as a theoretical figure of merit for a spectrum estimate. Since it only contains non-parametric estimates, then Hurvich (1985) extends to any estimate derived from the observed data. In particular, the class of estimates now include both Yule-Walker and periodogram-based type estimates. As Hurvich (1985) still wants to use Beltrao and Bloomfield's (1987) technique for the automatic smoothness parameter selection, he defines a leave-out-one spectrum version for any estimates. His main contribution is certainly the introduction of a method that allows for simultaneous and objective choice of both type of estimate and the corresponding smoothness parameters.

Hurvich (1985) presents three different forms of cross-validation methods: the cross-validated log-likelihood method of Beltrao and Bloomfield (1987), Stuetzle's smoothed estimate (SES, see Palmer (1983)) and an adaptation of the cross-validation mean squared error (CVMSE) method of Wahba and Wold (1975).

By introducing two generally applicable definitions of leave-out-one versions of the spectrum estimate he extends the applicability of the CVLL, SES and CVMSE techniques. Either of these definitions in conjunction with the CVLL, SES or CVMSE methods will yield an objective choice from a general class $\mathbb{C}$, where $\mathbb{C}$ includes any estimate whose leave-out-one versions is defined.

The distance measure, which Hurvich(1985) quite loosely denotes MISE, for the CVLL, SES, and CVMSE methods, respectively, are defined by,

$$MISE_1(\hat{f}) = \mathbb{E}\left\{\frac{1}{\tilde{N}}\sum_{p=1}^{\tilde{N}}(\frac{\hat{f}(\omega_p) - f(\omega_p)}{f(\omega_p)})^2\right\}, \tag{3.5.1}$$

$$MISE_2(\hat{f}) = \mathbb{E}\left\{\frac{1}{\tilde{N}}\sum_{p=1}^{\tilde{N}}(\hat{f}(\omega_p) - f(\omega_p))^2\right\}, \tag{3.5.2}$$

$$MISE_3(\hat{f}) = \mathbb{E}\left\{\frac{1}{\tilde{N}}\sum_{p=1}^{\tilde{N}}(\log\hat{f}(\omega_p) - \log f(\omega_p))^2\right\}, \tag{3.5.3}$$

The cross-validation estimates of $MISE_i(\hat{f})$, for $i = 1, 2, 3$ are

$$CVLL(\hat{f}) = \frac{1}{\tilde{N}}\left\{\sum_{p=1}^{\tilde{N}}\log\hat{f}^{-p}(\omega_p) + \frac{I(\omega_p)}{\hat{f}^{-p}(\omega_p)}\right\}, \tag{3.5.4}$$

$$SES(\hat{f}) = \frac{1}{\tilde{N}}\sum_{p=1}^{\tilde{N}}(\hat{f}^{-p}(\omega_p) - I(\omega_p))^2, \tag{3.5.5}$$

$$CVMSE(\hat{f}) = \frac{1}{\tilde{N}}\sum_{p=1}^{\tilde{N}}((\log\hat{f}^{-p}(\omega_p) - (\log I(\omega_p) + C))^2 - \frac{\pi^2}{6}) \tag{3.5.6}$$

where $C = 0.577216...$ is the Euler's constant and $\hat{f}^{-p}(\omega_p)$ is a general leave-out-one version of $\hat{f}$, such that $\hat{f}^{-p}(\omega_p)$ is approximately independent of $I(\omega_p)$ for each p. The independence is achieved by omitting $I(\omega_p)$ from the computation of $\hat{f}^{-p}(\omega_p)$.

In a first step, Hurvich (1985) defines the general leave-out-one spectrum estimate for any estimate that is a function of the sample auto-covariances $\hat{\gamma}(k)$ as defined before. In particular, this class of estimated includes both all non-parametric estimates and the Yule-Walker auto-regressive estimates. Let any estimate of this class be written as $\hat{f}(\omega, (\hat{\gamma}(k)))$.

$$\begin{aligned} I^{-p}(\omega) &= I(\omega) && \omega \notin ((\omega_{p-1}, \omega_{p+1}) \cup (\omega_{-p-1}, \omega_{-p+1})) \\ &= \theta_{1,\omega} I(\omega_{p-1}) + \theta_{2,\omega} I(\omega_{p+1}) && \omega \in (\omega_{p-1}, \omega_{p+1}) \\ &= I^{-p}(-\omega) && \omega \in (\omega_{-p-1}, \omega_{-p+1}) \end{aligned} \tag{3.5.7}$$

for $\omega \in [-\pi, \pi]$, where

$$\theta_{1,\omega} = 1 - \frac{\omega - \omega_{p-1}}{\omega_{p+1} - \omega_{p-1}} \tag{3.5.8}$$

and

$$\theta_{2,\omega} = \frac{\omega - \omega_{p-1}}{\omega_{p+1} - \omega_{p-1}}. \tag{3.5.9}$$

In general, the periodogram is only evaluated at the Fourier frequencies. If it is evaluated on a sufficiently fine grid, though, it completely determines the $\hat{\gamma}(k)$ sequence by

$$\hat{\gamma}(k) = \frac{2\pi}{n'} \sum_{k=0}^{n'-1} I(\omega'_k) e^{ir\omega'_k}, \tag{3.5.10}$$

where $n' = 2n$ and $\omega'_k = \frac{2\pi k}{n'}$. Here the ω'_k are defined on a grid exactly twice as finely spaced as the Fourier frequencies. Hurvich (1985) then defines the sequence $\hat{\gamma}(k)^{-p}$ by

$$\hat{\gamma}(k)^{-p} = \frac{2\pi}{n'} \sum_{k=0}^{n'-1} I^{-p}(\omega'_k) e^{ir\omega'_k}. \tag{3.5.11}$$

Finally he defines the general leave-out-one version of the spectrum estimate $\hat{f}^{-p}(\omega_p)$ for $1 \leq p \leq N$ as follows:

$$\hat{f}^{-p}(\omega_p) = \hat{f}(\omega_p; \hat{\gamma}(k)^{-p}). \tag{3.5.12}$$

It is important to note that $\hat{f}^{-p}(\omega_p)$ and $I(\omega_p)$ will be approximately independent for each p, as the computation of $\hat{f}^{-p}$ does not involve $I(\omega)$ for ω in the intervals $(\omega_{p-1}, \omega_{p+1})$ and $(\omega_{-p-1}, \omega_{-p+1})$.

Next, Hurvich (1985) defines a second general leave-out-one spectrum estimate which can be applied to any estimate and is denoted by $\hat{f}(\omega; x_t)$. First, he defines $\{J_k\}_{k=1}^n$, the Fourier transform of $\{x_t\}_{t=1}^n$, by

$$J_k = \frac{1}{n} \sum_{t=1}^{n} x_t e^{-i\omega_k t}. \tag{3.5.13}$$

This sequence completely determines the data sequence, through the relation

$$x_t = \sum_{k=1}^{n} J_k e^{i\omega_k t}. \tag{3.5.14}$$

Then, Hurvich (1985) defines the leave-out-one version of J_k, J_k^{-p}, for $1 \leq p \leq N$:

$$\begin{aligned} J_k^{-p} &= J_k & k \neq p, k \neq n-p \quad (3.5.15)\\ &= 0.5(J_{k-1} + J_{k+1}) & k = p, k = n-p \quad (3.5.16) \end{aligned}$$

and the leave-out-one-ω_p version of the data sequence $\left\{x_t^{-p}\right\}_{t=0}^{n-1}$ by

$$x_t^{-p} = \sum_{k=1}^{n} J_k^{-p} e^{-\omega_k t}. \tag{3.5.17}$$

Finally, the general leave-out-one spectrum estimate is defined as:

$$\hat{f}^{-p}(\omega_p) = \hat{f}(\omega_p; \left\{x_t^{-p}\right\}). \tag{3.5.18}$$

(see Hurvich, 1985; Fortin and Kuzmics, 2000)

CHAPTER 4

Transformation Based Smooth Estimators

Now, we are assuming that the density function $f_\Theta(\theta)$ is a continuous, circular random variable, Θ, is a non-negative continuous function such that (see Carnicero et al., 2018)

$$f_\Theta(\theta + 2\pi r) = f_\Theta(\theta),\ r \in \mathbb{Z}, \tag{4.0.1}$$

and

$$\int_{-\pi}^{\pi} f_\Theta(\theta) = 1. \tag{4.0.2}$$

4.1 Transformation Based Kernel Estimator

If we transform the angular data θ on $[-\pi, \pi]$ to the $[-\infty, \infty]$, the kernel density estimator on the real line may use the transformation:

$$x = \tan\left(\frac{\theta}{2}\right) \ \in \ [-\infty, \infty], \tag{4.1.1}$$

and the kernel density estimator of x is given by

$$\hat{p}(x) = \frac{1}{n}\sum_{j=1}^{n} K_h\Big(x - \tan\Big(\frac{\theta_j}{2}\Big)\Big). \tag{4.1.2}$$

Thus the transformation based kernel density estimator of $f(\theta)$ is given by

$$\hat{f}(\theta) = \frac{1}{1+\cos(\theta)}\hat{p}\Big\{\frac{\sin(\theta)}{1+\cos(\theta)}\Big\} \tag{4.1.3}$$

This is a motivation about transform the linear model to the periodic model. (see Chaubey, 2017)

4.2 Transformation Based Bernstein Polynomial Estimator

Babu and Chaubey (2006) consider estimating the distributions defined on a hyper-cube, extending the uni-variate Bernstein polynomials (see Babu, Canty and Chaubey, 2002; Vitale, 1973).
The interval $[0,1]$ can be mapped into the interval $[-\pi,\pi]$ through a one-to-one transformation, such as (see Chaubey, 2017),

$$\theta = 2\arctan\Big\{\frac{1}{c}\tan(\pi\Big(x-\frac{1}{2}\Big))\Big\},\ c\in\mathbb{Z}, \tag{4.2.1}$$

while, the backward transformation for $\theta\in[-\pi,\pi]$ is given by

$$x_c(\theta) = \frac{1}{2} + \frac{1}{\pi}\arctan\Big(c\tan\Big(\frac{\theta}{2}\Big)\Big),\ c\in\mathbb{Z} \tag{4.2.2}$$

The transforms the Bernstein polynomial to a periodic function given by

$$\begin{aligned}\hat{f}(\theta) &= B'\Big\{\frac{1}{2}+\frac{1}{\pi}\arctan\Big(c\tan\Big(\frac{\theta}{2}\Big)\Big)\Big\}\\ &= \frac{1}{2\pi}B'(x_c(\theta))\frac{c(1+\tan^2(\frac{\theta}{2}))}{1+c^2\tan^2(\frac{\theta}{2})}\end{aligned}$$

The function $x_c(\theta)$ is periodic according to the above definition because of the periodicity of arctan function. In particular, when $c = 1$ we will get the linear transformation as considered in Kakizawa (2006). For $c > 1$, the graph starts as concave and then becomes convex where as for $c < 1$, the nature of the graph is opposite.

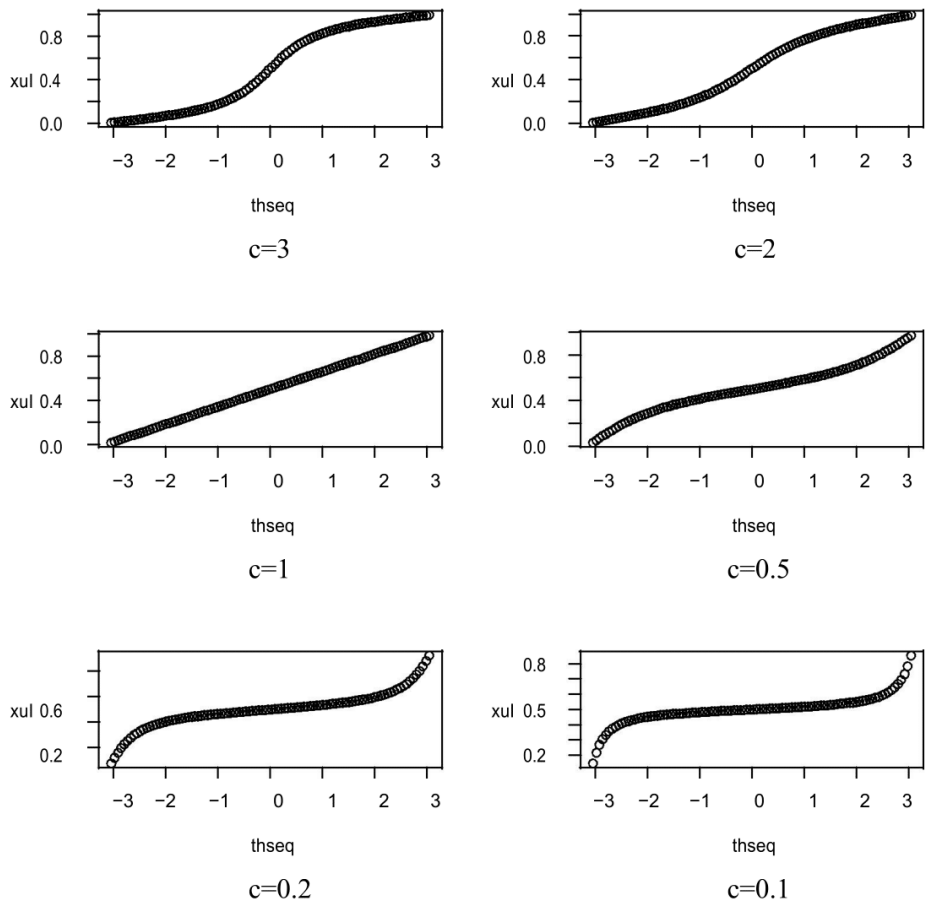

Figure 4.1: Graphs of arctan transformation for different c

In this thesis we only consider $c = 1$, however other values may be explored. This transformation yields a periodic density even for $c = 1$.

CHAPTER 5

A Simulation Study

5.1 Simulated Examples

In this section, we compare some estimators which is discussed in the previous chapters, for obtaining optimal parameters by means of a small simulation study. These procedures are applied to a set of AR and MA processes, selected such as to exhibit different shapes of the spectral densities. Three typical estimators are considered for comparison.

The first experiment was concerned with the **smoothing of the periodogram**. We are going to consider the *Modified Bartlett kernel*

$$\kappa(z) = \begin{cases} 1 - \dfrac{z}{q+1}\,, & \text{for } z = 1, 2, ..., q(\textit{the length of the window}), \\ 0, & \text{otherwise.} \end{cases}$$

and the *Quadratic-Spectral kernel (Epanechnikov)*

$$\kappa(z) = \frac{3}{(\pi z)^2}\Big\{\frac{\sin(\pi z)}{\pi z} - \cos(\pi z)\Big\}.$$

Many of the commonly used kernels in non-paramtric estimation, but the $Quadratic - Spectral\ kernel\ (Epanechnikov)$ is the optimal one since its optimality properties

in the density estimation setting (see Jerome and Donald, 1981).

The other experiment is about the **Bernstein polynomial estimator**. Here I will use the ideas presented in the previous section to generalize SES method (similar studies could be carried out for the $CVLL$ and $CVMSE$ methods) (see Hurvich, 1985). Since data-driven choices of m are not the subject of this thesis, we have fixed the interval of m from $T/2$ to T for convenience.

There are three different models we considered for the simulation study were from the $ARMA(a, b)$ model

$$X_t + \sum_{m=1}^{p} a_m X_{t-m} = \sum_{n=0}^{q} b_n Z_{t-n}, \quad \{Z_t\} \sim N(0,1) \tag{5.1.1}$$

given by

Example 1. AR(1) model: $X_t = -0.75X_{t-1} + Z_t$.

Example 2. AR(2) model: $X_t = 0.7X_{t-1} - 0.1X_{t-2} + Z_t$.

Example 3. MA(2) model: $X_t = -0.7Z_t + 0.1Z_{t-1}$.

These models are convenient because of their simplicity and the different spectra they represent. In general, the spectrum of the observed data can be expressed by:

$$f(\omega) = \frac{\sigma^2 |\sum_{n=0}^{q} b_n e^{-in\omega}|^2}{2\pi |1 - \sum_{m=1}^{p} a_m e^{-im\omega}|^2}, \tag{5.1.2}$$

Therefore, we could find the spectral density function for above models:

(1) $AR(1)$ model with the spectral density

$$f(\omega) = \frac{1}{2\pi |1 + 0.75e^{-i\omega}|^2} \tag{5.1.3}$$

(2) $AR(2)$ model with the spectral density

$$f(\omega) = \frac{1}{2\pi |1 - 0.7e^{-i\omega} + 0.1e^{-2i\omega}|^2} \tag{5.1.4}$$

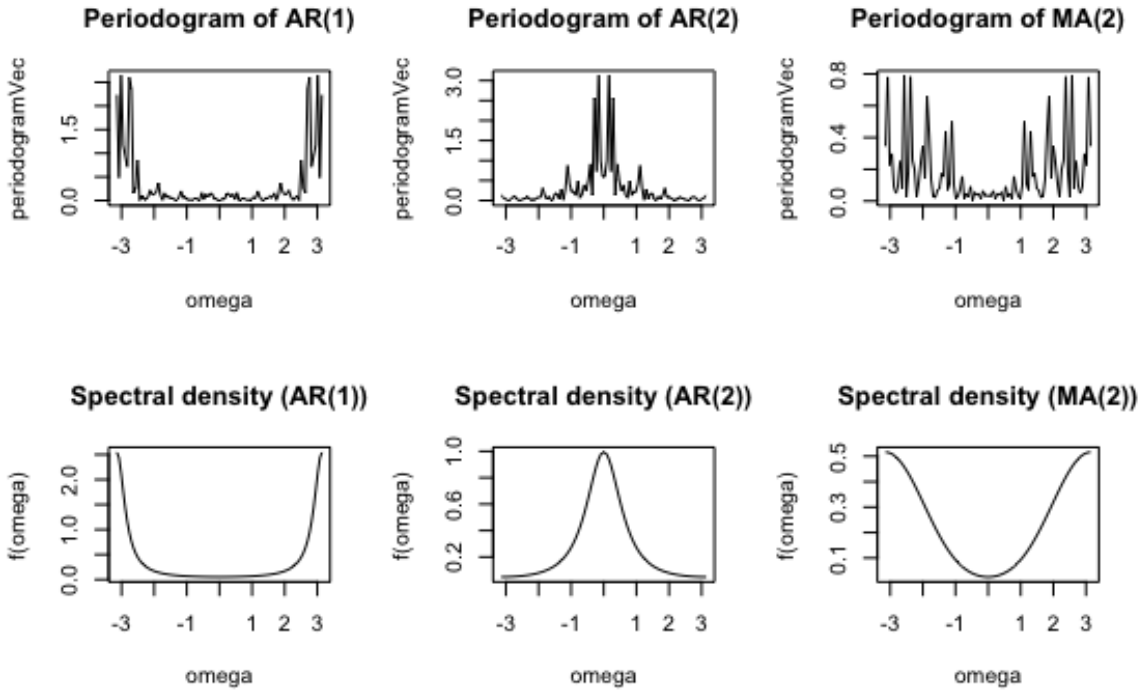

Figure 5.1: Spectral density & Periodogram of three models

(3) $MA(2)$ model with the spectral density

$$f(\omega) = \frac{|1 - 0.7e^{-i\omega} + 0.1e^{-2i\omega}|^2}{2\pi} \tag{5.1.5}$$

As a measure of performance, first we use the $ISE(Integrated\ Squared\ Error)$ as a criterion to compare the results of each estimator.

Definition 12. *The* **Integrated Squared Error** *or* **ISE** *of a spectral density estimate $\hat{f}$ is given by:*

$$ISE(\hat{f}) = \int_{-\pi}^{\pi} [\hat{f}(\lambda) - f(\lambda)]^2 d\lambda, \tag{5.1.6}$$

thus the ISE is the squared value of the distance between the estimated density and the true density, integrated over the support of the distribution.

Note that, from M pseudo-random samples of size T,

$$MISE(\hat{f}) \approx \frac{1}{M}\sum_{i=1}^{M} ISE_i(\hat{f}) \tag{5.1.7}$$

is a Monte Carlo approximation of $MISE(\hat{f})$, where $ISE_i(\hat{f})$ denotes the value of ISE caculated from the i^{th} randomly generated sample from f.

However, we have considered the following measure as goodness of the estimators of the simulated averages of ISE (see Guillen et al., 2003):

$$D = \frac{1}{n}\sum_{i=1}^{n}[\hat{f}(u) - f(u)]^2, \tag{5.1.8}$$

In order to see the distribution of these divergence measures, we repeat the data 1000 times for each sample size (T = 30, 60, 90). Even though this provides a limited study, it does give a relative comparison of the estimators based on the simulation. A small number of replications is chosen, specially for local comparison as the computing time required becomes enormous for larger replication.

Remark 6. For the data-driven method,

$$ISE(\hat{f}) \approx \frac{1}{\tilde{N}}\sum_{j=1}^{\tilde{N}}(\hat{f}(\omega_j) - f(\omega_j))^2 \tag{5.1.9}$$

where $\omega_j = \frac{2\pi j}{T}$ and $\tilde{N} = \frac{T-1}{2}$. (see Hurvich, 1985)

5.2 Global Comparison of the Estimator

The results of the global comparison for the three estimators is referred to *Table* 5.1 to *Table* 5.3. Smaller value of $MISE$ indicate better performance of the corresponding estimation method.

Table 5.1: Three Estimators with Simulated ISE for AR(1) model

Sample Size	ISE	Modified Bartlett	Quadratic	Bernstein
T=30	**Mean**	0.4706822	0.24657	0.3036055
	Sd	0.908447	0.7360596	0.6637804
	Median	0.1671616	0.1504551	0.1880033
T=60	**Mean**	0.5358472	0.1474206	0.1996915
	Sd	0.9443545	0.1351729	0.3390923
	Median	0.274227	0.1181745	0.118315
T=90	**Mean**	0.5735706	0.1322012	0.16835
	Sd	0.8066755	0.05799178	0.3908912
	Median	0.3320279	0.1167127	0.09840433

Table 5.2: Three Estimators with Simulated ISE for AR(2) model

Sample Size	ISE	Modified Bartlett	Quadratic	Bernstein
T=30	**Mean**	0.1067432	0.04290031	0.03130121
	Sd	0.1500781	0.0380971	0.03200272
	Median	0.06404788	0.03460981	0.02493879
T=60	**Mean**	0.1385082	0.02725096	0.02028
	Sd	0.1350772	0.02129694	0.01753757
	Median	0.09761639	0.02153008	0.01518008
T=90	**Mean**	0.1357193	0.02138214	0.01626558
	Sd	0.1091978	0.01544942	0.01688769
	Median	0.1059413	0.01679605	0.01212275

Table 5.3: Three Estimators with Simulated ISE for MA(2) model

Sample Size	ISE	Modified Bartlett	Quadratic	Bernstein
T=30	**Mean**	0.07181	0.01755306	0.02365606
	Sd	0.08117924	0.02765912	0.03458512
	Median	0.04891364	0.0103875	0.01643942
T=60	**Mean**	0.07643905	0.007452475	0.01624438
	Sd	0.9443545	0.009347472	0.01712805
	Median	0.274227	0.00464611	0.01518008
T=90	**Mean**	0.08468778	0.005281107	0.01626558
	Sd	0.05233537	0.00569866	0.01688769
	Median	0.07326121	0.003419347	0.01212275

According to these summary statistics showed in the tables, the Bernstein polynomial estimator and the Quadratic kernel estimator do better than another one in all the presented cases. Typically, the Bernstein polynomial method performs pretty much the same as the Quadratic kernel method for these three models. Then Let's consider the Local Comparison.

5.3 Local Comparison of the Estimator

If we are interested in non-parametric spectral density estimation at a single frequency, we need to minimize local distance measures such as MSE $(Mean\ Squared\ Error)$ and so on. We consider only fixed frequencies of the form $\omega_j = \frac{2\pi j}{T}$, where j is an integer. By the usual variance decomposition the MSE can be written as the sum of the squared bias and the variance.

$$\mathbb{E}(\hat{f}(\omega) - f(\omega))^2 = (\mathbb{E}(\hat{f}) - f(\omega))^2 + \mathbb{E}(\hat{f}(\omega) - \mathbb{E}(\hat{f}(\omega)))^2 \tag{5.3.1}$$

$$MSE(\hat{f}(\omega)) = BIAS^2(\hat{f}(\omega)) + VARS(\hat{f}(\omega)) \tag{5.3.2}$$

For various spectral bandwidths, given the true spectral density, it is possible to at least asymptotically assess bias and variance of the corresponding estimators. The bias as well as variance generally will depend on the spectral density and its derivatives and the form of the spectral window.

$$MSE(\hat{f}) = \frac{1}{T}\sum_{i=1}^{T}\left\{\hat{f}_i(\omega) - f(\omega)\right\}^2, \tag{5.3.3}$$

where $\hat{f}_i(\omega)$ is the density estimator of $f(\omega)$ based on the i^{th} replication. (see Priestley, 1981)

The plots with the simulation outcomes are given at the following.

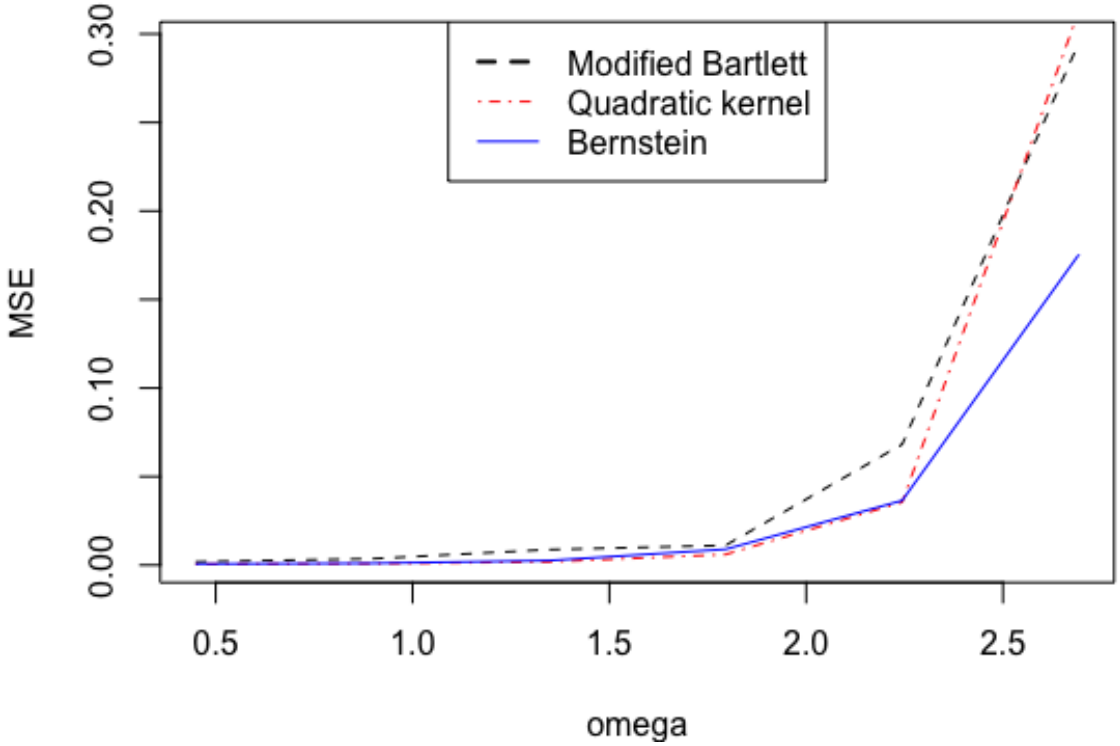

Figure 5.2: MSE for AR(1) model (T=30)

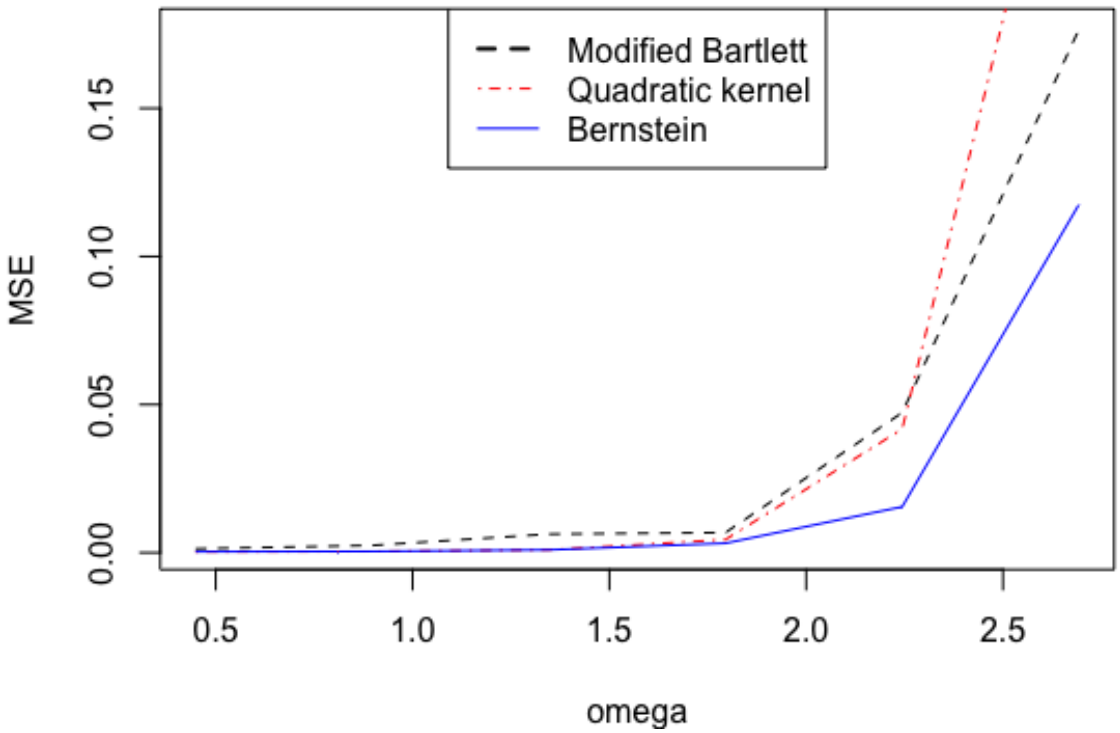

Figure 5.3: MSE for AR(1) model (T=60)

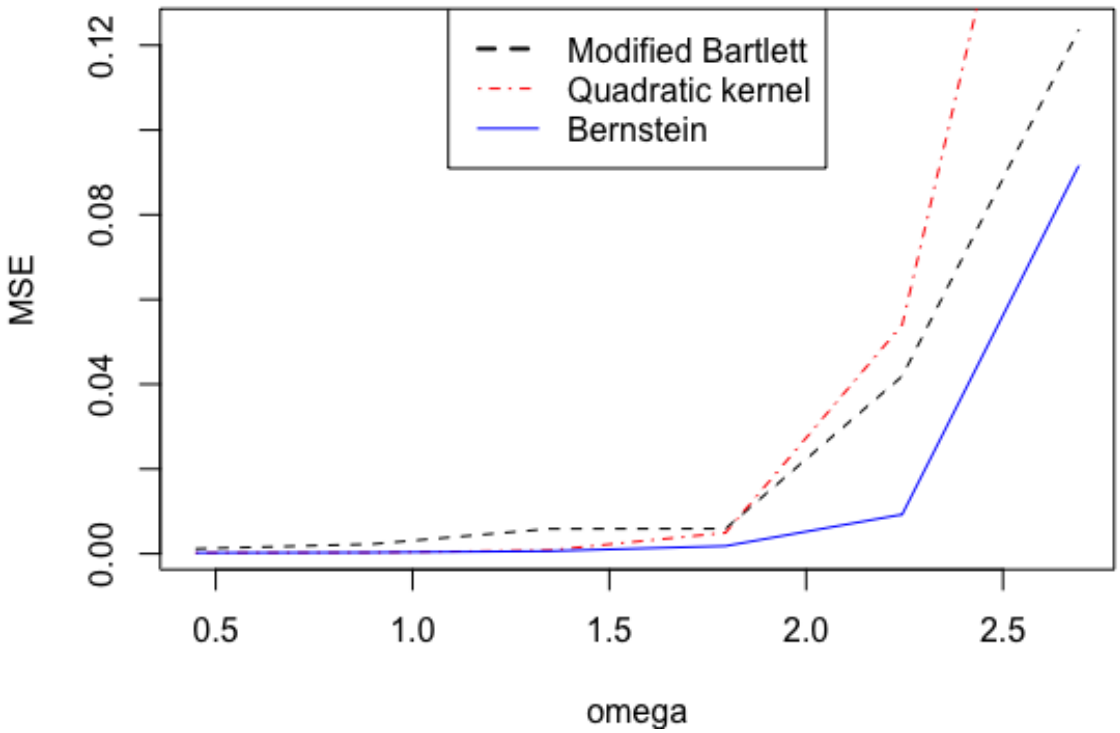

Figure 5.4: MSE for AR(1) model (T=90)

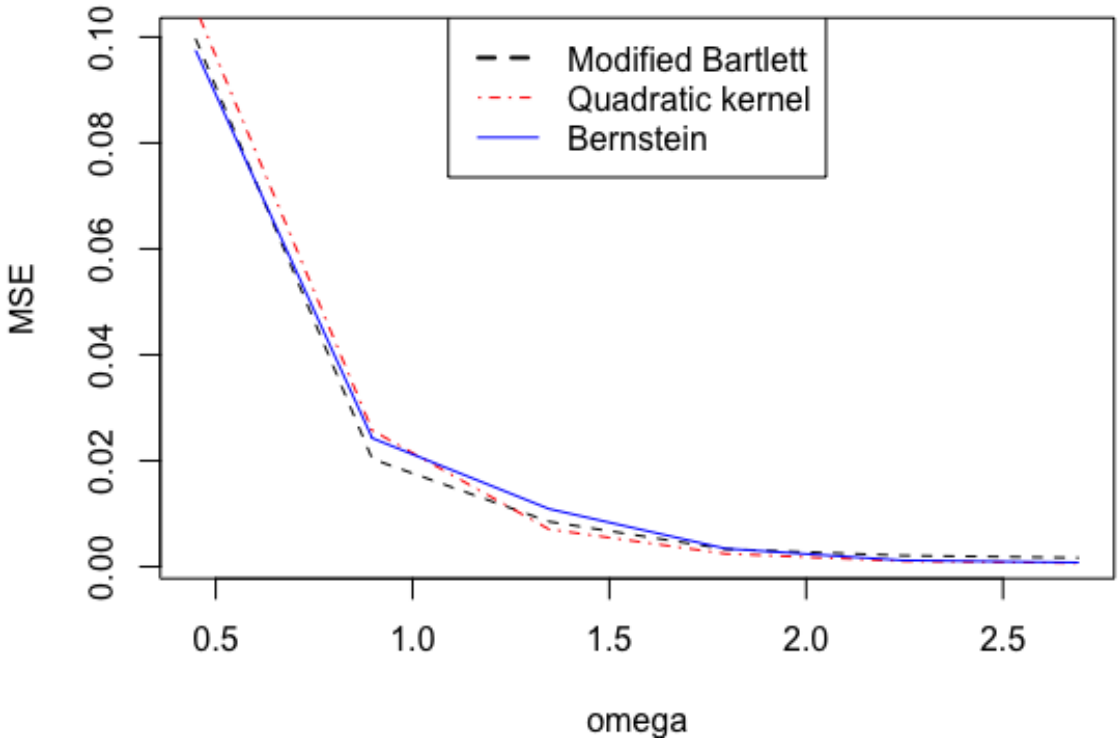

Figure 5.5: MSE for AR(2) model (T=30)

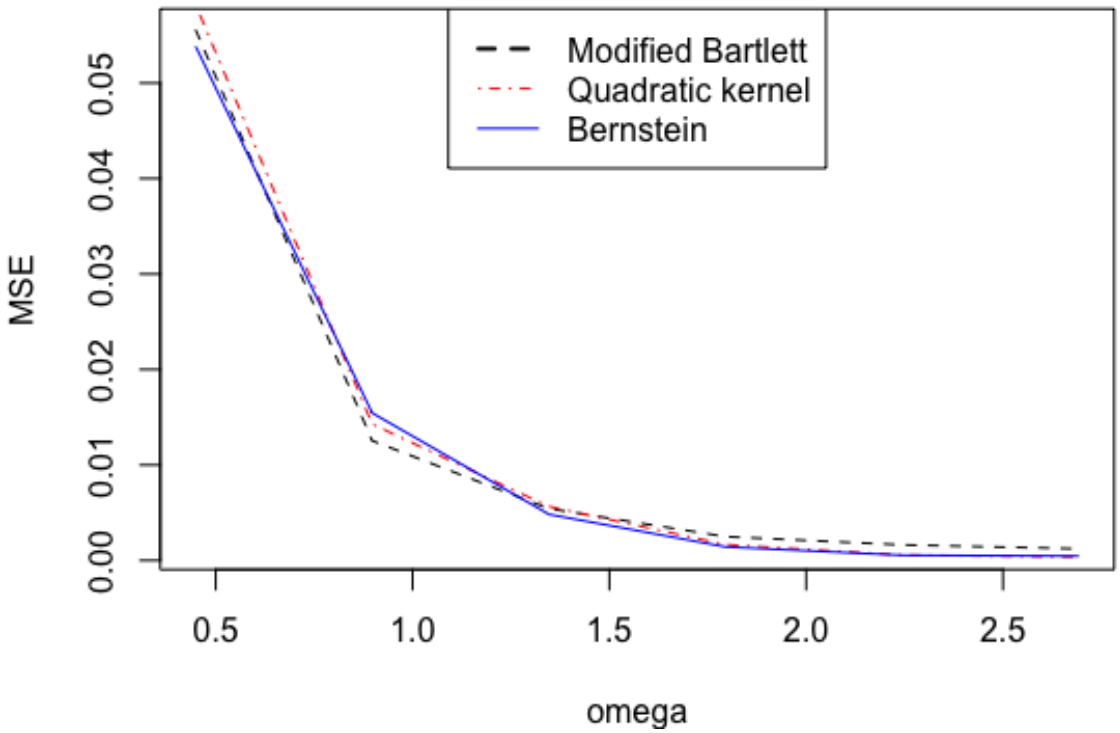

Figure 5.6: MSE for AR(2) model (T=60)

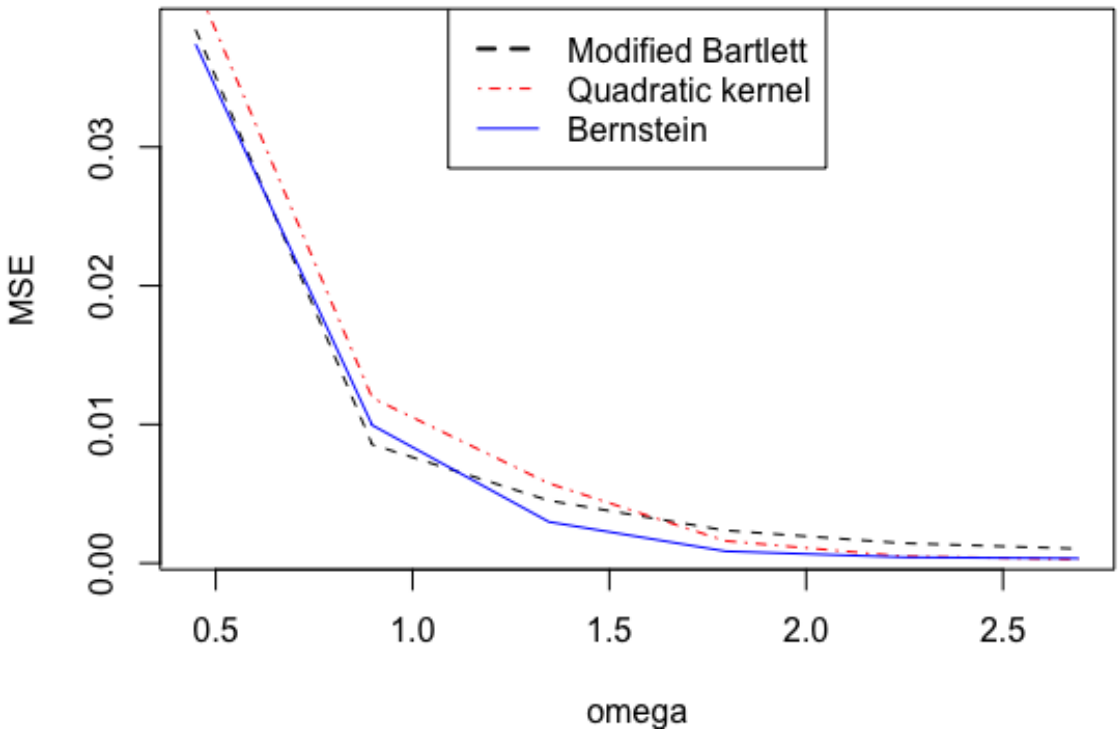

Figure 5.7: MSE for AR(2) model (T=90)

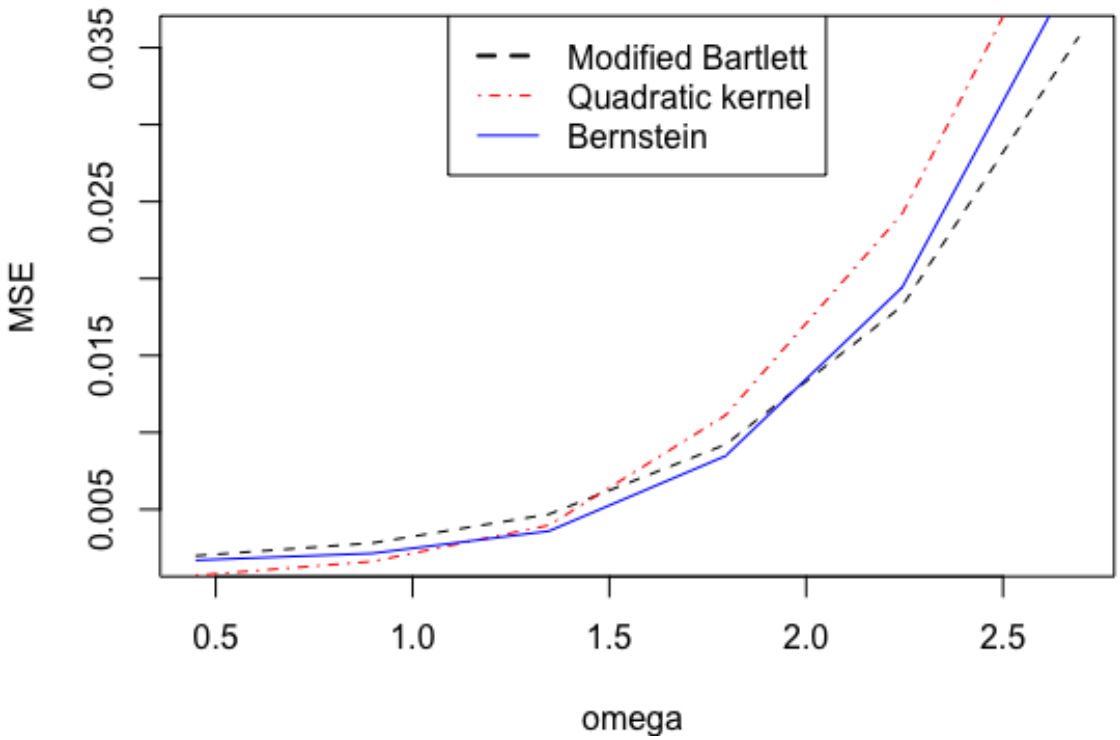

Figure 5.8: MSE for MA(2) model (T=30)

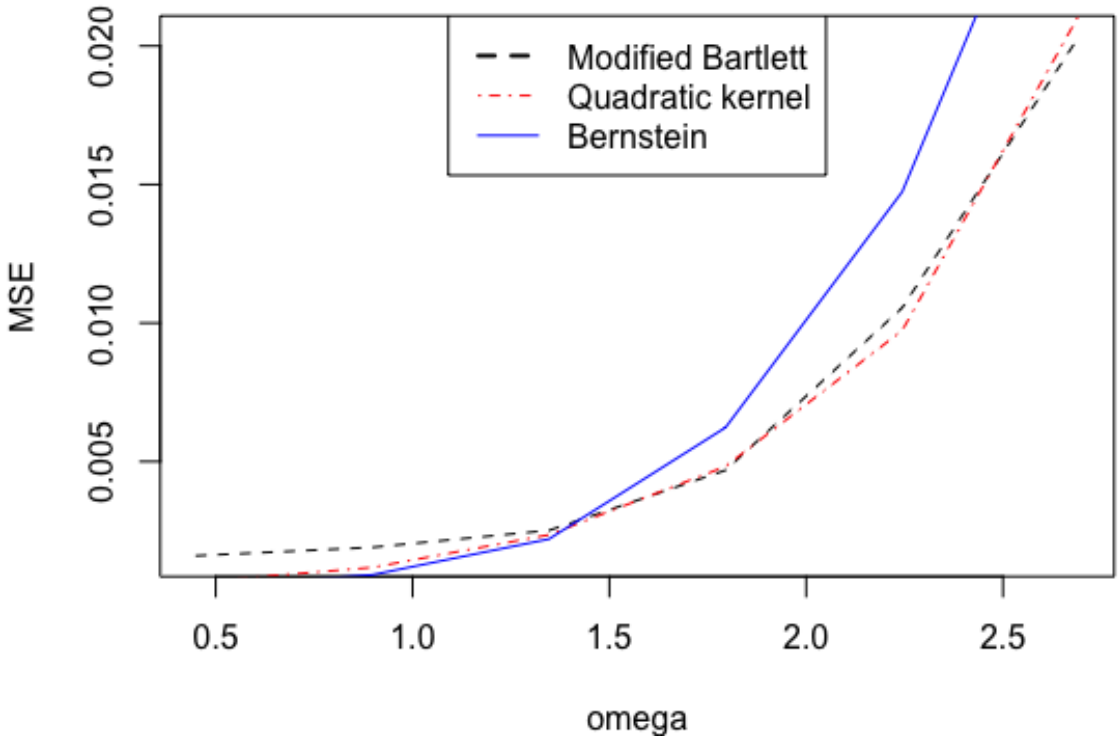

Figure 5.9: MSE for MA(2) model (T=60)

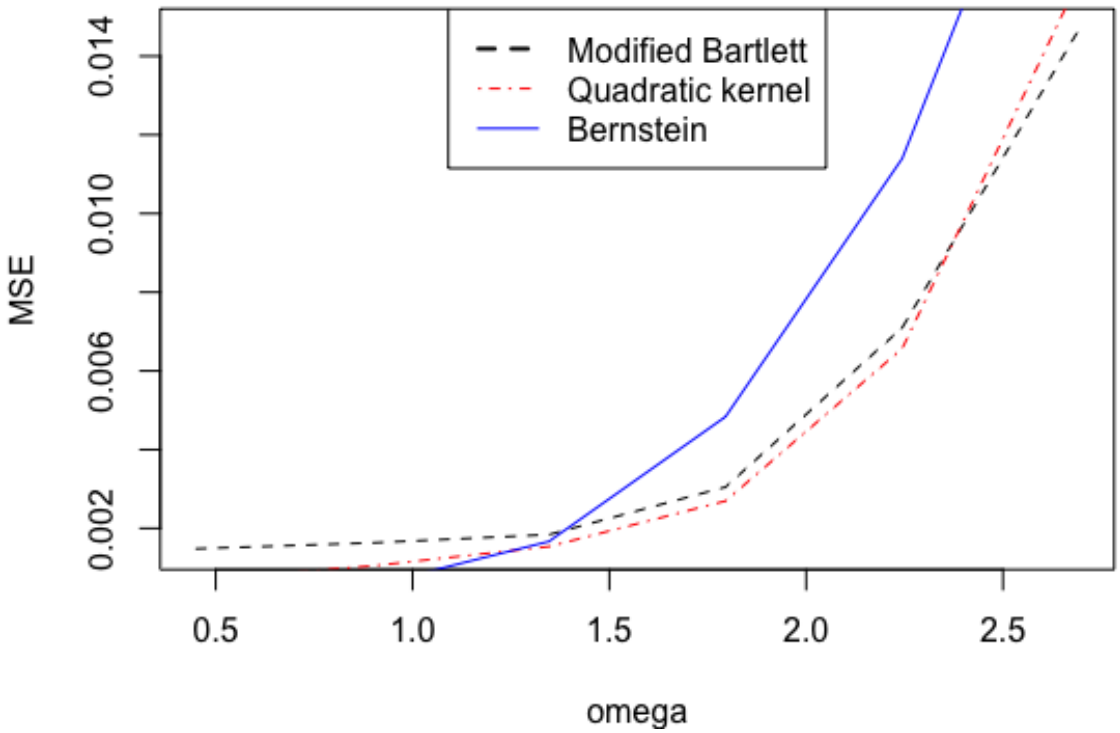

Figure 5.10: MSE for MA(2) model (T=90)

Focussing on the *Figure* 5.2 to *Figure* 5.4, we can see that at the first half the Bernstein estimator is pretty much similar to the Quadratic kernel estimator in terms of MSE. However, we can find that Bernstein estimator is better than the Quadratic kernel estimator when estimating the tail. Further looking at the *Figure* 5.5 to *Figure* 5.7, these three estimators seem very similar to each other, but as the T is getting larger, the Bernstein estimator seems better than others. *Figure* 5.8 to *Figure* 5.10 is for $MA(2)$ model, here is different than other two models, we can clearly see that kernel estimator is better than Bernstein estimator in this model.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

In this thesis we have compared several estimations of the spectral density and reviewed some methods for determining optimal scale parameters for non-parametric spectral window. These are cross-validation based estimates following Hurvich (1985) and Beltrao & Bloomfield (1987).

A small simulation study indicates that all of the three estimators considered here are quite good, if the sample size T is large. However, we conclude that Bernstein estimator is the best choice for $AR(1)$ model and $AR(2)$ model, and the transformed kernel estimation with quadratic kernel is the best for $MA(2)$ model in our case. The estimators are found to achieve good mean square error in our simulation study.

6.2 Future Work

In the future, we may consider larger sample sizes and focus on other spectral density models such $ARMA$ and $ARIMA$, that are of importance in the area of time series. It may be very useful if we could apply the Bernstein-Kernel polynomial estimators to the circular case. We could also consider extensions to spherical data or data defined on the torus. Such approaches could be implemented by exploring generalizations of multivariate Bernstein polynomial density estimators as in e.g. Babu and Chaubey (2006).

Another problem relevant to the transformation based Bernstein polynomial estimator is the determination of 'optimal' value of c. This may be explored based on the asymptotic Mean Integrated Squared Error. This is usually achieved by considering the leading term of an appropriate expansion of $MISE$.

APPENDIX A

The R-Code of the program

A.1 MISE of Bernstein

```
ISECal<-function(){
T <- 30
x1<-arima.sim(model=list(ar=-0.75),n=T)
j <- 1:ceiling((T-1)/2)
omega_j <- (2*pi*j)/T

IOmegaJ <- function(xVector,omega) {
    T <- length(xVector)
    lagh <- 1:(T-1)
    acvfPart <- acvf(xVector,h=(T-1))[-1]
    cosPart <- cos(lagh*omega)
    fun <- 1/(2*pi)*acvf(xVector,h=0)+1/pi*sum(acvfPart*cosPart)
    return(fun)
  }
periodogramVec<-  sapply(omega_j,IOmegaJ,xVector=x1)
f <- function(omegaVector) 1/(2*pi*(1+1.5*cos(omegaVector)+9/16))
```

```
fVvalue <- sapply(omega_j,f)

kseq <- 1:T
omega_k <- 2*pi*kseq/T
tseq<- 1:T
J_kreal<-function(k,xVector){
    T<-length(xVector)
    tseq<-seq(1,T)
    csvector<-cos(2*pi*k*tseq/T)
    return(sum(csvector*xVector)/T)
  }
J_kvalR<- sapply(kseq, J_kreal, xVector=x1)

J_kIm<-function(k,xVector){
    T<-length(xVector)
    tseq<-seq(1,T)
    sinvector <- -sin(2*pi*k*tseq/T)
    return(sum(sinvector*xVector)/T)
  }
J_kvalI<- sapply(kseq, J_kIm, xVector=x1)

J_kminsonevalR<-sapply((kseq-1),J_kreal,xVector=x1)
J_kminsonevalI<-sapply((kseq-1),J_kIm,xVector=x1)

J_kplusonevalR <- sapply((kseq+1),J_kreal,xVector=x1)
J_kplusonevalI <- sapply((kseq+1),J_kIm,xVector=x1)

J_minsjR <- 1/2 *(J_kminsonevalR+J_kplusonevalR)
J_minsjI <- 1/2 *(J_kminsonevalI+J_kplusonevalI)
```

```
x_minsj.real <- function(t){
    kseq<-seq(1,T)
    cosTerm <- cos(2*pi*kseq*t/T)
    sinTerm <- sin(2*pi*kseq*t/T)
    val <- sum(J_minsjR*cosTerm-J_minsjI*sinTerm)
    return(val)}
x_minsjval.real <- sapply(tseq, x_minsj.real)

x_minsj.im <- function(t){
    kseq<-seq(1,T)
    cosTerm <- cos(2*pi*kseq*t/T)
    sinTerm <- sin(2*pi*kseq*t/T)
    val <- sum(J_minsjR*sinTerm+J_minsjI*cosTerm)
    return(val)}
x_minsjval.im <- sapply(tseq, x_minsj.im)
hHat <- function(xVector, omega){
    N <- length(xVector)
    lagTerm <- 1:(N-1)
    jInverse <- 1/lagTerm
    acvfTerm <- acvf(xVector, h = (N-1))[-1]
    sinTerm <- sin(omega*lagTerm)
    val <- ((omega+pi)/(2*pi))*acvf(xVector, h=0)
 + (1/pi)*sum(jInverse*acvfTerm*sinTerm)
    return(val)
  }
BsmdenWithM<-function(omega,xVector,m = m)
  {
    N <- length(xVector)
```

```
    m <- m
    jSequence <- 0:(m-1)
    jPiMSeqUpper <- -pi+((jSequence + 1)*2*pi)/m
    jPiMSeqLower <- -pi+(jSequence*2*pi/m)

    hUpper <- sapply(jPiMSeqUpper, hHat, xVector = xVector)
    hLower <- sapply(jPiMSeqLower, hHat, xVector = xVector)
    hTerm <- hUpper - hLower
    bTerm <- dbinom(jSequence, size=m-1, prob = (omega+pi)/(2*pi))

    val <- m/(2*pi)*sum(hTerm*bTerm)
    return(val)
  }

bdsmVector <- function(m)
{sapply(omega_j, BsmdenWithM, xVector = x_minsjval.real,m=m)}
bdsmVectorIn <- function(m)
{sapply(omega_j, BsmdenWithM, xVector = x1,m=m)}
c <- ceiling(T/log(T))

periodogramVecNew<- sapply(omega_j,IOmegaJ,xVector=x_minsjval.real)
value <- function(m){
    T <- length(x_minsjval.real)
    j <- 1:ceiling((T-1)/2)
    val<- (1/ceiling((T-1)/2))*sum(log(bdsmVector(m))
+periodogramVecNew/bdsmVector(m))
    return(val)
  }
 simplify=FALSE) )
```

```
mSequence <- c:(T/2)
valueOfIse <- sapply(mSequence,value)

optimalM <- mSequence[which.min(valueOfIse)]
ISEvalue <- function(m){
    T <- length(x1)
    j <- 1:T
    val<- (1/T)*sum((bdsmVectorIn(m)-fVvalue)^2)
    return(val)
  }
FINALVALUE <- ISEvalue(optimalM)
FINALVALUE
}
N <- 1000
SOL <- replicate(N,ISECal())
sum(SOL)/N
```

A.2 MSE of Bernstein

```
T <- 30
omegaVec <- seq(0,pi,length.out = 8)
f <- function(omegaVector) 1/(2*pi*(1+1.5*cos(omegaVector)+9/16))
fVvalue <- function(omegaVector) sapply(omegaVector,f)
j <- 1:ceiling((T-1)/2)
omega_j <- (2*pi*j)/T
IOmegaJ <- function(xVector,omega) {
    T <- length(xVector)
    lagh <- 1:(T-1)
    acvfPart <- acvf(xVector,h=(T-1))[-1]
    cosPart <- cos(lagh*omega)
    fun <- 1/(2*pi)*acvf(xVector,h=0)+1/pi*sum(acvfPart*cosPart)
    return(fun)
  }
periodogramVec<- function(xVector)
 sapply(omega_j,IOmegaJ,xVector=xVector)
kseq <- 1:T
omega_k <- 2*pi*kseq/T
tseq<- 1:T
J_kreal<-function(k,xVector){
    T<-length(xVector)
    tseq<-seq(1,T)
    csvector<-cos(2*pi*k*tseq/T)
    return(sum(csvector*xVector)/T)
  }
J_kvalR<- function(xVector) sapply(kseq, J_kreal, xVector=xVector)
```

```
J_kIm<-function(k,xVector){
    T<-length(xVector)
    tseq<-seq(1,T)
    sinvector <- -sin(2*pi*k*tseq/T)
    return(sum(sinvector*xVector)/T)
  }
J_kvalI<- function(xVector) sapply(kseq, J_kIm, xVector=xVector)

J_kminsonevalR<-function(xVector)
sapply((kseq-1),J_kreal,xVector=xVector)
J_kminsonevalI<-function(xVector)
sapply((kseq-1),J_kIm,xVector=xVector)

J_kplusonevalR <-function(xVector)
sapply((kseq+1),J_kreal,xVector=xVector)
J_kplusonevalI <- function(xVector)
sapply((kseq+1),J_kIm,xVector=xVector)

J_minsjR <-function(xVector)
 1/2 *(J_kminsonevalR(xVector)+J_kplusonevalR(xVector))
J_minsjI <-function(xVector)
 1/2 *(J_kminsonevalI(xVector)+J_kplusonevalI(xVector))

x_minsj.real <- function(t,xVector){
    kseq<-seq(1,T)
    cosTerm <- cos(2*pi*kseq*t/T)
    sinTerm <- sin(2*pi*kseq*t/T)
    val <- sum(J_minsjR(xVector)*cosTerm-J_minsjI(xVector)*sinTerm)
    return(val)}
```

```
hHat <- function(xVector, omega){
    N <- length(xVector)
    lagTerm <- 1:(N-1)
    jInverse <- 1/lagTerm
    acvfTerm <- acvf(xVector, h = (N-1))[-1]
    sinTerm <- sin(omega*lagTerm)
    val <- ((omega+pi)/(2*pi))*acvf(xVector, h=0)
+ (1/pi)*sum(jInverse*acvfTerm*sinTerm)
    return(val)
  }
BsmdenWithM<-function(omega,xVector,m = m)
  {
    N <- length(xVector)
    jSequence <- 0:(m-1)
    jPiMSeqUpper <- -pi+((jSequence + 1)*2*pi)/m
    jPiMSeqLower <- -pi+(jSequence*2*pi/m)

    hUpper <- sapply(jPiMSeqUpper, hHat, xVector = xVector)
    hLower <- sapply(jPiMSeqLower, hHat, xVector = xVector)
    hTerm <- hUpper - hLower
    bTerm <- dbinom(jSequence, size=m-1, prob = (omega+pi)/(2*pi))

    val <- m/(2*pi)*sum(hTerm*bTerm)
    return(val)
  }

bdsmVector <- function(m,xVector)
{sapply(omega_j, BsmdenWithM, xVector =xVector,m=m)}
```

APPENDIX A: THE R-CODE OF THE PROGRAM

```
bdsmVectorIn <- function(xVector,m)
{sapply(omega_j, BsmdenWithM, xVector = xVector,m=m)}
BernsteinVec<- function(m,xVector)
 sapply(omegaVec,BsmdenWithM,m=m,xVector=xVector)

simTime <- 1000
BMse <- replicate(simTime,{
  x1<-arima.sim(model=list(ar=-0.75),n=T)
  x_minsjval.real <- sapply(tseq, x_minsj.real, x1)
  value <- function(m){
  j <- 1:ceiling((T-1)/2)
  val<- (1/ceiling((T-1)/2))
*sum((bdsmVector(m,xVector=x_minsjval.real)-periodogramVec(xVector=x1))
  return(val)
  }
  mSequence <- (0.5*T):T
  valueOfIse <- sapply(mSequence,value)
  optimalM <- mSequence[which.min(valueOfIse)]
  BernsteinMse <- (BernsteinVec(optimalM,x1)-fVvalue(omegaVec))^2
  print(BernsteinMse)
})

c(sum(BMse[2,])/simTime,sum(BMse[3,])/simTime,sum(BMse[4,])/simTime,
sum(BMse[5,])/simTime,sum(BMse[6,])/simTime,sum(BMse[7,])/simTime)
```

References

[1] Babu, G. J., Canty, A. J. and Chaubey, Y. P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, **105**, 377–392.

[2] Babu, G. J. and Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on a hyper-cube using Bernstein polynomials for dependent random vectors. *Statistics and Probability Letters*, **76**, 959–969.

[3] Bartlett, M. S. (1948). Smoothing periodograms from time-series with continuous spectra. *Nature*, **161,** 686–687.

[4] Beltrao, K. and Bloomfield, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *Journal of Time Series Analysis*, **8**(1), 21–38.

[5] Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theory and Methods* (Second Edition), Springer Series in Statistics.

[6] Carnicero, J. A., Wiper, M. P. and Ausin, M. C. (2018). Density estimation of circular data with Bernstein polynomials. *Hacettepe Journal of Mathematics and Statistics*, **47**(2), 273–286.

[7] Chaubey, Y. P. (2017). On smooth density estimation for circular data. *Invited Paper*, Presented at 61st World Statistics Congress, Marrakech July 16–21, 2017.

[8] Daniell, P. J. (1946). Discussion on 'Symposium on auto-correlation in time series'. *Supplement to the Journal of the Royal Statistical Society*, **8**, 88-90.

[9] Fortin, I. and Kuzmics, C. (2000). Optimal window width choice in spectral density estimation. *Journal of Statistical Computation and Simulation*, **67**, 109–131.

[10] Grenander, U. (1951). On empirical spectral analysis of stochastic processes. *Arkiv for Matematik*, **1**, 503–531.

[11] Guillen, M., Bolance, C. and Nielsen, J. P. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, **32**, 19–36.

[12] Hipel, K. W. and McLeod, A. I. (1994) *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, Amsterdam and New York.

[13] Hosoya. Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, **10**, 132–153.

[14] Hurvich, C. M. (1985). Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *Journal of the American Statistical Association*, **80**(392), 933–940.

[15] Jerome, S. and Donald, Y. (1981). Asymptotically optimum kernels for density estimation at a point. *The Annals of Statistics*, **9**(2), 334–336.

[16] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *Journal of Nonparametric Statistics*, **16**, 709–729.

[17] Kakizawa, Y. (2006). Bernstein polynomial estimation of a spectral density. *Journal of Time Series Analysis*, **27**(2), 253–287.

[18] Lorentz, G. (1986). *Bernstein polynomials* (Second Edition), Chelsea Publishing, New York.

[19] Mugdadi, A. and Jetter, J. (2010). A simulation study for the bandwidth selection in the kernel density estimation based on the exact and the asymptotic MISE. *Pak. J. Statist*, **26**, 239-265.

[20] Palmer, F. (1983). Adaptive spectral smoothing. *Technical Report ONR*, **30**, Massachusetts Institute of Technology, Statistics Center.

[21] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**(3), 1065–1076.

[22] Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press. London.

[23] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**(56), 832–837.

[24] Schuster, A. (1898). On the investigation on hidden periodicties with application to a supposed 26-day period of meteorological phenomena. *Terrestrial Magnestism*, **3,** 13–41.

[25] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, United Kingdom.

[26] Vitale, R. A. (1973). A Bernstein polynomial approach to density estimation. *Communications in Statistics*, **2,** 493–506.

[27] Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, **75**, 122–132.

[28] Wahba, G. and Wold, S. (1975). Periodic splines for spectral density estimation: the use of cross validation for determining the degree of smoothing. *Communications in Statistics*, **4**, 125–141.

[29] Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing*. Chapman and Hall, London, United Kingdom.