이 학 석 사 학 위 논 문

Nonparametric Spectral Density Estimation for Time Series Data with Missing Values

결측값이 있는 시계열 자료의 비모수 스펙트럴 밀도함수 추정

2013년 2월

서울대학교 자연과학대학원

통계학과

박 수 진

Nonparametric Spectral Density Estimation for Time Series Data with Missing Values

by

Sujin Park

A Thesis
submitted in fulfillment of the requirement
for the degree of
Master of Science
in
Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University
February, 2013

Abstract

Spectral density estimation is important in many different areas including astronomy, radio communications, and geophysics. Although many spectral density estimation techniques assume equally spaced complete data, in most cases, real data include missing values. To solve this problem we suggest a new nonparametric spectral density estimation procedure with missing data. Our algorithm is based on the self-consistency and relation between autocovariance and spectral density function, thus it is intuitive and simple. Also, it can be applied with any nonparametric spectral density estimation method such as kernel estimators, wavelet regression, and spline estimators. The practical performance of the proposed method is reported through simulation study.

Keywords: Missing data, nonparametric spectral density estimation, periodogram, time series.

Student Number: 2011-20243

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spectral density estimation is important in many different areas including astronomy, radio communications, radars, and geophysics. Most existing methods for estimating the spectral density function assume equally spaced complete data. However, real data collected from these fields usually contain missing values. In this case, the discrete Fourier transform (DFT) cannot be applied to calculate the periodogram which is an essential factor for many useful nonparametric spectral analysis techniques.

To deal with this problem, various techniques for the estimation of spectral density with missing date have been proposed. For instance, the Lomb-Scargle method (Scargle, 1982) can be applied to irregularly spaced data. In missing-data case, the Lomb-Scargle periodogram can be obtained by imputing missing values with the mean of observed data. However, this time-domain mean imputation may destroy high frequency information in data. In addition, nonparametric spectral density estimation algorithm via the EM algorithm was suggested by Li *et al.* (2005). A shortcoming of this algorithm is that it works well only for strong spectral peaks but not in low spectrum energy regions.

In this paper we consider new nonparametric spectral density estimation

for time series data with missing values occurring in arbitrary patterns. Since the spectral density is the frequency-domain representation of a time series that is directly related to the autocovariance in the time-domain representation, the spectral density and the autocovariance function contain the same information, but express it in different ways. In this respect, it seems natural to impute missing values using estimated spectral density. Through our approach reflecting this concept, we can nearly achieve the true spectrum with no loss of information. The best part of this method is that it is extremely easily and diversely applicable. Although we explain our approach with kernel estimators, it can be combined with any other nonparametric spectrum estimation techniques for complete data such as wavelet regression or spline estimators. Moreover, from these aspects of the proposed method, we can infer the possibility of an extension of the suggested algorithm to the high dimensional problem.

After finishing this paper, we accidentally found that there was a very similar algorithm proposed by Lee and Zhu (2009). However, we would like to say that our idea had been developed independently from theirs.

The remainder of this paper is organized as follows. In Section 2, background materials required to derive a new method are reviewed. In Section 3, we propose a nonparametric algorithm for the missing data spectral estimation problem. Simulation results are reported in Section 4. Finally, we provide concluding remarks in Section 5.

Chapter 2

Review of Spectral Analysis

2.1 Spectral Density

Here we introduce the property of spectral density, which is the basic concept needed for studying time series via spectral analysis. If the autocovariance function, $\gamma(h)$, of a stationary process satisfies

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

for the lag h, then it has the representation

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\omega h} f(\omega) d\omega, \qquad h = 0, \pm 1, \pm 2, ...,$$

as the inverse transform of the spectral density, where $f(\omega)$ is the spectral density function of the form

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h}, \qquad -1/2 \leq \omega \leq 1/2.$$

These fact should make sure that, when the summability condition is satisfied, the autocovariance function $\gamma(h)$ and the spectral density function $f(\omega)$

give the same information. In the language of advanced calculus, the autocovariance and spectral density are Fourier transform pairs and this relation is unique.

We will focus on the estimation of the spectral density which is the frequency-domain characterization of a time series. Mathematically, the spectral density is defined for both negative and positive frequencies. However, due to symmetry of the function and its repeating pattern for frequencies outside the range $-1/2$ to $1/2$, we only need to be concerned with frequencies between 0 and $1/2$.

An example of spectral analysis for a given time series of an autoregressive (AR) process of order 3 by using the autocovariance function and spectrum (*i.e.*, the estimated spectral density) is provided in Figure 2.1.

2.2 Definitions of DFT and Periodogram

The definitions of discrete Fourier transform (DFT) and periodogram described by Shumway and Stoffer (2005) are as follows.

Definition 2.2.1. (Discrete Fourier transform) *Given data $x_1, ..., x_n$, we define the discrete Fourier transform (DFT) to be*

$$d(\omega_j) = n^{-1/2} \sum_{k=1}^{n} x_k e^{-2\pi i \omega_j k}$$

for $j = 0, 1, ..., n-1$, where the frequencies $\omega_j = j/n$ are called the Fourier or fundamental frequencies.

If n is a highly composite integer, the DFT can be computed by the fast Fourier transform (FFT) introduced in Cooley and Tukey (1965).

The periodogram is a function that displays information about the periodic components of a time series. Any time series can be expressed as a sum of

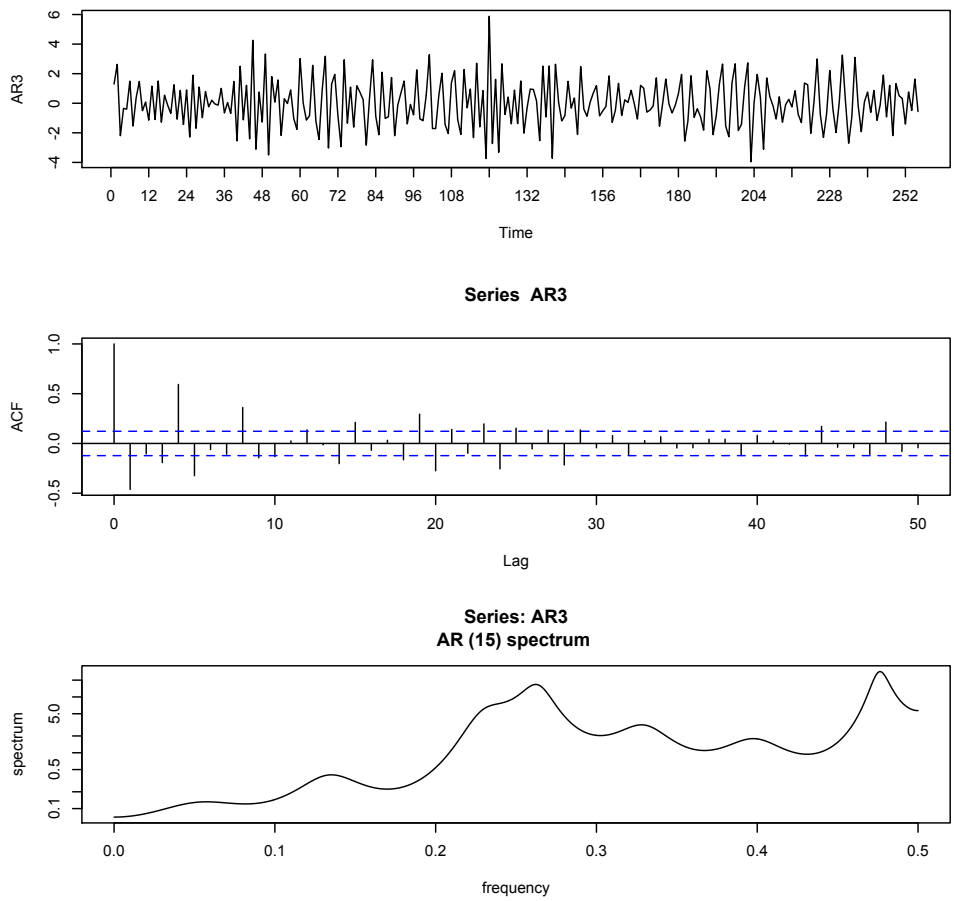

FIGURE 2.1. *Plot of AR(3) series (top), its autocovariance function (middle), and the estimated spectral density from AR fit (bottom).*

cosine and sine waves oscillating at the fundamental frequencies $\omega_j = j/n$, with $j = 1, 2, ..., n/2$. The periodogram means the relative strengths of the various frequencies for explaining the variation in the time series.

Definition 2.2.2. (Periodogram) *Given data $x_1, ..., x_n$, we define the periodogram to be*

$$I(\omega_j) = |d(\omega_j)|^2$$

for $j = 0, 1, 2, ..., n-1$.

Note that $I(0) = n\overline{x}^2$, where $\overline{x}$ is the sample mean. In addition, we can write the DFT as

$$d(\omega_j) = n^{-1/2} \sum_{k=1}^{n} (x_k - \overline{x}) e^{-2\pi i \omega_j k}$$

for $j \neq 0$. Thus, for $j \neq 0$,

$$\begin{aligned} I(\omega_j) = |d(\omega_j)|^2 &= n^{-1} \sum_{k=1}^{n} \sum_{l=1}^{n} (x_k - \overline{x})(x_l - \overline{x}) e^{-2\pi i \omega_j (k-l)} \\ &= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{k+|h|} - \overline{x})(x_k - \overline{x}) e^{-2\pi i \omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}. \end{aligned}$$

The periodogram is a sample estimate of a population function called the spectral density, which is a frequency-domain characterization of a population stationary time series. We are now ready to tie together the periodogram, which is the sample-based concept, with the spectral density, which is the population-based concept.

The raw periodogram is a rough sample estimate of the population spectral density. Figure 2.2 shows the raw periodogram of AR(3) series. The estimate

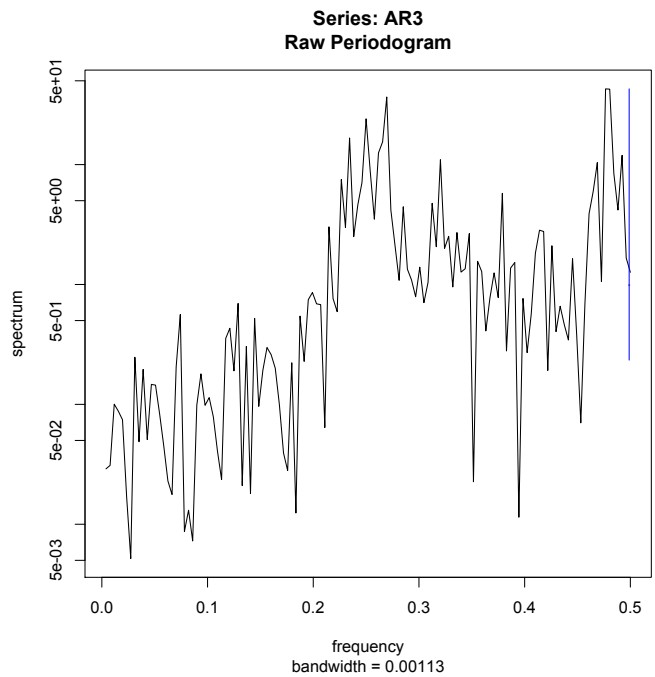

FIGURE 2.2. *The raw periodogram of AR(3) series.*

is rough in part, because we only use the discrete fundamental frequencies for the periodogram whereas the spectral density is defined over a continuum of frequencies. One possible improvement to the periodogram estimate of the spectral density is to smooth it using centered moving averages. An additional smoothing can be created using tapering methods which weight the ends of the series less than the center of the data. An alternative approach to smoothing the periodogram is a parametric estimation approach based on the fact that any stationary time series can be approximated by an AR model of some order. In this case a suitable AR model is found, and then the spectral density is estimated as the spectral density for that estimated AR model. However, in this paper, we will deal with the nonparametric smoothing method.

2.3 Smoothing Method

Though there are several methods for smoothing a periodogram, in fact, it is merely a centered moving average procedure with a few possible modifications. For time series, the Daniell kernel with parameter m is a centered moving average which creates a smoothed value at time t by averaging all values between times $t-m$ and $t+m$. For example, the Daniell kernel with $m=2$,

$$\hat{x}_t = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5}.$$

When we smooth a periodogram, we implement it across a frequency interval rather than a time interval. Thus, the smoothed value $\hat{I}(\omega_j)$ is a weighted average of periodogram values for frequencies in the range $(j-m)/n$ to $(j+m)/n$ when we use the Daniell kernel with parameter m to smooth a periodogram. To generalize this range, we will introduce the bandwidth as follows.

There are $L = 2m+1$ fundamental frequency values in the range $(j-m)/n$ to $(j+m)/n$, the range of values used for smoothing. The bandwidth for the smoothed periodogram is defined as

$$B_\omega = \frac{L}{n}.$$

This is a measure of the width of the frequency interval used for smoothing the periodogram. When unequal weights are used in the smoothing, the bandwidth definition is modified. Denote the smoothed periodogram value at $\omega_j = j/n$ as

$$\hat{I}(\omega_j) = \sum_{k=-m}^{+m} h_k I(\omega_j + \frac{k}{n}).$$

The h_k are the possibly unequal weights used in the smoothing. The bandwidth formula is then modified to

$$B_\omega = \frac{L_h}{n} = \frac{1/\sum h_k^2}{n}.$$

The result of smoothing method is suggested in Figure 2.3. We can see the rough shape of periodogram at the top of the figure which we find it using only discrete fundamental frequencies without smoothing. As we apply the Daniell kernel with parameter 4 to the sample data repeatedly, we can get smoother shape of periodograms such as we can find in the middle and bottom of the figure.

We will apply a kernel smoothing in our simulation study and it will be described later.

2.4 Self-Consistency

One important thing when we deal with signal data is that retaining much information as possible while compressing the size of data. In this sense, self-consistency was illustrated as a fundamental and powerful concept in statistics by Tarpey and Flury (1996).

Suppose we want to approximate the distribution of a random vector $\mathbf{Y}$ by another random vector $\mathbf{Z}$ which has much simpler structure. The mean squared error $E\|\mathbf{Y}-\mathbf{Z}\|^2$ can be one measure to verify how well $\mathbf{Y}$ is estimated by $\mathbf{Z}$. In the case of mean squared error, the approximation can always be improved using $E[\mathbf{Y}|\mathbf{Z}]$ instead of $\mathbf{Z}$. It is because, for any function g, $E\|\mathbf{Y}-E[\mathbf{Y}|\mathbf{Z}]\|^2 \leq E\|\mathbf{Y}-g(\mathbf{Z})\|^2$. This leads to the definition of self-cosistency as follows.

Definition 2.4.1. (Self-consistency) *For two jointly distributed random vectors* $\mathbf{Y}$ *and* $\mathbf{Z}$, $\mathbf{Z}$ *is self-consistent for* $\mathbf{Y}$ *if*

$$E[\mathbf{Y}|\mathbf{Z}] = \mathbf{Z} \qquad \textit{almost surely.}$$

Based on this concept, some fruitful outcomes were developed such as EM algorithm, complete data score function and incomplete data maximum likelihood estimator. Since we would like to estimate the spectral density function

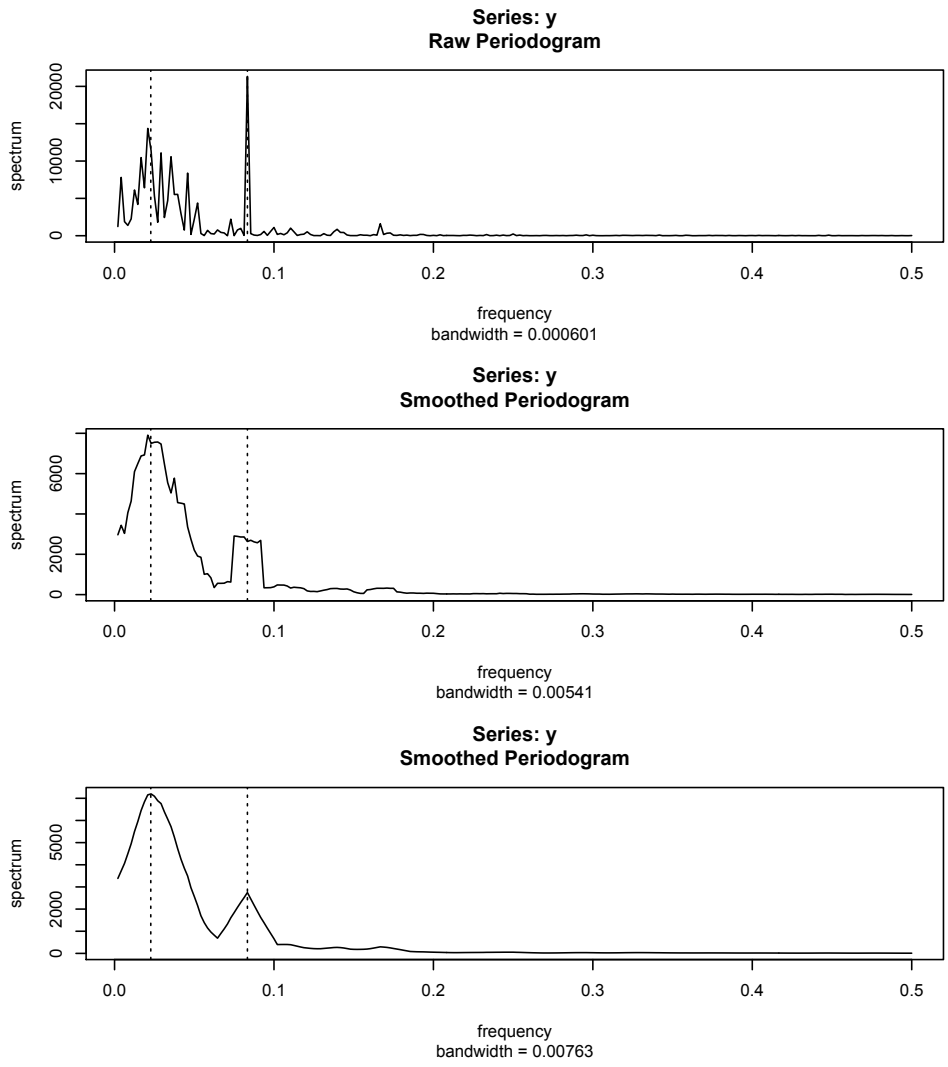

FIGURE 2.3. *The periodogram of fish recruitment series,* $n = 453$ *(top) and smoothed periodograms using the Daniell kernel with* $m = 4$ *(middle), using two passes of the Daniell kernel with* $m = 4$ *(bottom).*

from incomplete data, this might give an idea to estimate it more accurately. Therefore, like other applications, we also use this self-consistency concept to achieve our algorithm.

Chapter 3

The Proposed Method

Let $Y_t, t = 1, 2, ..., n$ be a stationary time series with the autocovariance function $\gamma(h) = Cov(Y_t, Y_{t+h})$. Suppose that the data have missing values, $\mathbf{Y} = (\mathbf{Y_o}, \mathbf{Y_m})$, where $\mathbf{Y_o}$ is observed data and $\mathbf{Y_m}$ is missing data. Our goal is to estimate the spectral density function f in this case. Since missing data are present, we describe a method to impute data $\mathbf{Y_m}$ such that the data after imputation has the same spectral density as the original data. The proposed algorithm assumes that we already have a smoothing procedure for estimating $f(\omega)$ from the periodogram $I(\omega)$ computed using complete data.

To improve the accuracy of estimation we adopt self-consistency concept. Denote the corresponding estimate for f obtained by complete data as $\hat{f}_c$. And the estimate for f from missing data denotes as $\hat{f}_m$. Then we can obtain $\hat{f}_c$ from the following self-consistent equation:

$$E[\hat{f}_c|\mathbf{Y_o}, f = \hat{f}_m] = \hat{f}_m.$$

The solution of this equation can be obtained via an iterative process and it will be described next. However, there exists a problem. The conditional expectation cannot usually be calculated analytically, and we could use Monte Carlo approximation to solve this problem. Then, the proposed method are as

follows.

We start with an initial estimate of $\hat{f}^{(0)}$ by imputing $\mathbf{Y_m}$ with the average of $\mathbf{Y_o}$ and applying the complete data procedure for estimating $f(\omega)$, and then iterate between multiple imputation and estimation of $\hat{f}^{(k)}$ for $k = 1, 2, ...$:

1. Compute the autocovariance function $\hat{\gamma}^{(k-1)}$ from $\hat{f}^{(k-1)}$.

2. Perform multiple imputation as, for $l = 1, 2, ..., L$, simulate $\mathbf{Y_{m}}_{(l)}$ from $\mathbf{Y_m}|\mathbf{Y_o}$ using $\hat{\gamma}^{(k-1)}$.

3. Compute $I_{(l)}$ and $\hat{f}_{(l)}$ for $l = 1, 2, ..., L$.

4. Compute $\hat{f}^{(k)}$ as the average of all $\hat{f}_{(l)}$; i.e.,

$$\hat{f}^{(k)}(\omega) = \sum_{l=1}^{L} \hat{f}_{(l)}(\omega).$$

5. Iterate steps 1 to 4 until $\hat{f}^{(k)}(\omega)$ converges.

As we briefly referred in introduction, this algorithm has a great advantage on applicability. In step 3, we can use any nonparametric estimation procedure to $I_{(l)}$ to calculate $\hat{f}_{(l)}$. Thus, it can be coupled with wavelet regression or spline estimators though we use kernel estimators. Also, we could expect this algorithm to be extended to high dimensional problem. This might be our future work.

Chapter 4

Simulation Study

To verify the performance of the proposed method, a simulation study was conducted. Two test examples were used and they were from the $ARMA(p, q)$ model

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

where a_t as iid $N(0, 1)$ with parameters given by

- **Example 1.** AR(3) with $\phi_1 = 0.9, \phi_2 = 0.7$ and $\phi_3 = 0.1$.
- **Example 2.** MA(3) with $\theta_1 = 0.7, \theta_2 = 0.9$ and $\theta_3 = 0.3$.

Two spectral density functions of these examples are displayed in Figure 4.1. Our purpose is estimating spectral density function as accurately as possible in each case when there exists some missing values. Both have the same sample size $n = 1024$. Then for each sample, missing data were introduced with two patterns. Firstly, 10% of observations were removed randomly and secondly, 30% of them were removed in the same way. We used the Gaussian kernel to obtain $\hat{f}$ from the periodogram I. The bandwidth of a local Gaussian kernel regression estimate was selected by using the direct plug-in methodology by Ruppert *et al.* (1995). We calculated two mean squared errors $(MSE) =$

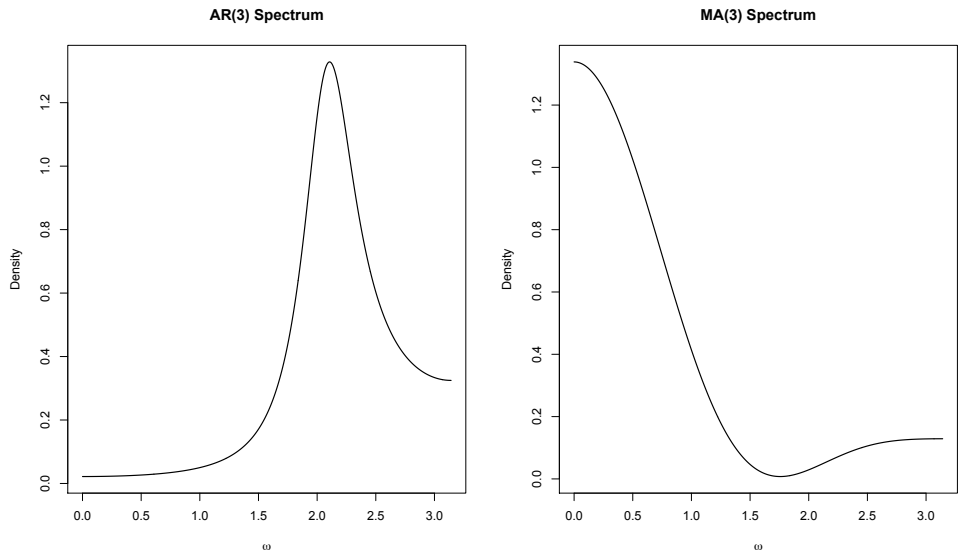

FIGURE 4.1. *The two spectral densities used in the simulation study.*

$\frac{1}{n}\sum_{j=0}^{n-1}(\hat{f}_j - f_j)^2$ of the spectrum estimates obtained by using missing data and complete data for comparison. In this case, f_j denotes $f(w_j)$. Table 4.1 gives information on $\sqrt{MSE}$ of the estimated spectral density using the proposed method with complete data and missing data. One could see that the value of $\sqrt{MSE}$ obtained by the proposed method with missing data is similar to that of complete data. Also we can find that the shape of estimated spectrum is approximated with the true one in Figure 4.2 and 4.3. That is, it works well even in the case of existing many missing values.

case	AR(3)	MA(3)
complete data	0.05779 (0.05779)	0.06038 (0.06038)
missing data	0.06404 (0.06081)	0.06825 (0.05536)

TABLE 4.1. $\sqrt{MSE}$ *of the estimated spectral density using the proposed method with complete data and missing data. Numbers without parentheses are the case of 10%-missing. Numbers in parentheses are the case of 30%-missing.*

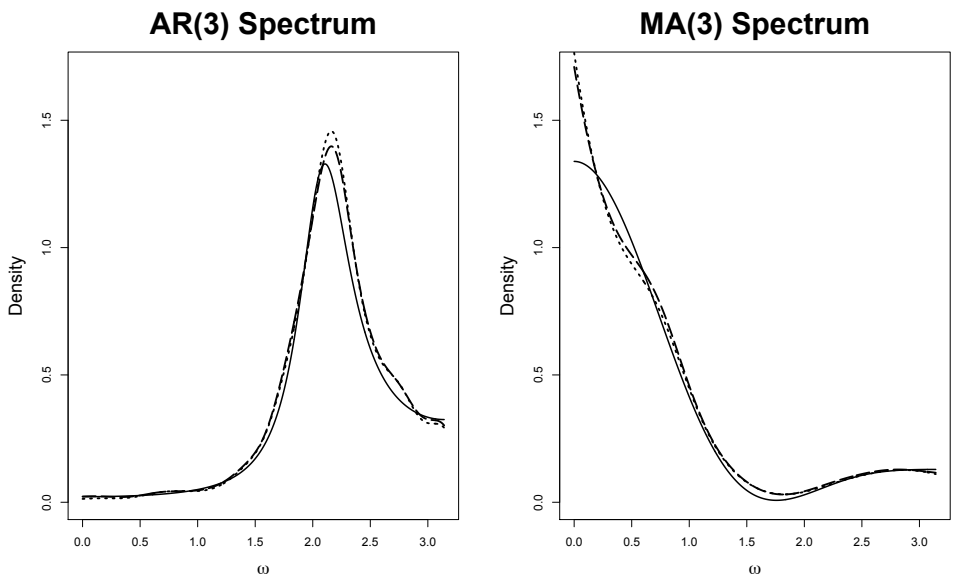

FIGURE 4.2. *The results that 10% of observations were removed randomly. Solid line is the spectral density used in the simulation study and long-dashed line is the estimated one through the suggested algorithm with complete data. Lastly, dotted line is the estimated one with missing data.*

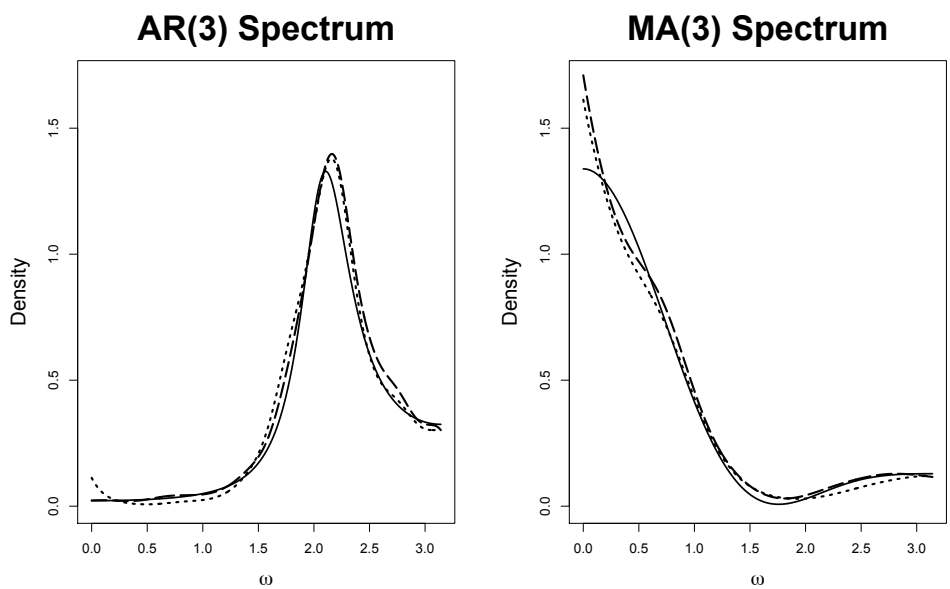

FIGURE 4.3. *The results that 30% of observations were removed randomly. Solid line is the spectral density used in the simulation study and long-dashed line is the estimated one through the suggested algorithm with complete data. Lastly, dotted line is the estimated one with missing data.*

Chapter 5

Conclusions

In this paper we introduce our new nonparametric method for estimating spectral density for time series with missing values. By imposing self-consistency concept, we try to estimate spectral density function more precisely using missing data. The performance of the proposed algorithm were proven in the simulation study. We can find that it works well even there exists quite many missing values. In addition, it is very flexible, in the sense that it can be combined with any complete data nonparametric spectrum estimation methods. Also, we expect that this algorithm can be developed to handle high dimensional problem. Therefore, there are still remaining works.

References

Brillinger, D. R. (1974). Fourier analysis of stationary processes, *Proceedings of the IEEE*, **62**, 1628–1643.

Lee, T. C. M. and Zhu, Z. (2009). Nonparametric spectral density estimation with missing observations, In *Proceedings of IEEE International Conference on Acustics, Speech and Signal Processing; ICASSP 2009*, 3041–3044.

Lee, T. C. M. (2001). A stabilized bandwidth selection method for kernel smoothing of the periodogram, *Signal Processing*, **8**, 419–430.

Li, J., Wang, Y., Stoica, P. and Marzetta, T. L. (2005). Nonparametric spectral analysis with missing data via the EM algorithm, *Digital Signal Processing*, **15**, 191–206.

Scargle, J. D. (1982). Studies in astronomical time series analysis II: Statistical aspects of spectral analysis of unevenly space data, *The Astrophysical Journal*, **263**, 835–853.

Shumway, R. H. and Stoffer, D. S. (2005). *Time Series Analysis and Its Applications With R Examples*, 2nd ed., Springer.

Tarpey, T. and Flury, B. (1996). Self-consistency: A fundamental concept in statistics, *Statistical Science*, **11**, 299–243.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association*, **90**, 1257–1270.

국문초록

기상학이나 물리학에서 시계열 자료가 가지는 주기성분은 매우 중요한 정보를 가지고 있으며 이는 스펙트럴 밀도함수를 통해서 분석할 수 있다. 많은 스펙트럴 밀도함수 추정 방법들이 등간격으로 관찰된 완전한 시계열 자료를 가정하고 있지만, 실제 데이터 상에서는 대부분 결측값이 존재하게 된다. 이를 해결하기 위해 본 논문에서는 새로운 비모수 스펙트럴 밀도함수 추정 방법을 제안한다. 이는 웨이블렛이나 스플라인과 같이 다양한 방법들과 함께 적용될 수 있으며 일차원 자료뿐만 아니라 고차원 자료에까지 확장될 수 있을 것으로 기대된다.

주 요 어 : 결측값, 비모수 스펙트럴 밀도함수 추정, 시계열, 주기도
학　　번 : 2011-20243

감사의 글

대학원에 진학해 어느덧 2년이라는 시간이 흘러 연구실 생활을 마무리 짓는 시간이 왔습니다. 도종환님의 시 <흔들리며 피는 꽃>에 보면 '흔들리지 않고 피는 꽃이 어디 있으랴. 이 세상 그 어떤 아름다운 꽃들도 다 흔들리면서 피었나니. 흔들리면서 줄기를 곧게 세웠나니.' 라는 구절이 있습니다. 대학원 생활을 하며 수없이 흔들려 온 부족한 저를 이 시와 같이 항상 응원해 주신 감사한 분들이 생각납니다.

우선 언제나 저를 지지해주신 부모님과 가족들에게 고맙고 사랑한다고 전하고 싶습니다. 부모님께서 제멋대로인 저의 선택을 믿고 응원해주셨기에 씩씩하게 생활할 수 있었습니다. 공부한다는 핑계로 떠넘겼던 저의 귀찮은 부탁에도 언제나 불평 없이 도와준 언니, 언니는 정말 세상 어디에도 없는 착한 언니야. 그리고 귀여운 막내 수미, 몇 년 먼저 겪어본 사회생활이 생각만큼 쉽지 않아 뒤따라올 네가 걱정이 되기도 하지만, 우리 가족이 나를 응원해줬듯이 네가 어떤 선택을 하든 항상 응원할게.

정말 많이 감사하고, 죄송스럽기도 한 오희석 교수님. 취업준비를 핑계로 논문에 온전히 몰두하지 못해 죄송스럽기만 합니다. 항상 제 편에서 격려해주시고 위로해주셔서 면담이 끝날 때마다 가슴이 따뜻해졌습니다. 교수님 말씀처럼 항상 밝은 모습으로 주어진 일에 최선을 다하도록 하겠습니다. 앞으로도 부끄럽지 않은 제자가 되도록 노력할테니 계속 응원해주시고, 지켜봐 주세요.

우리 연구실 가족들에게도 감사한 마음뿐입니다. 카리스마 동익오빠, 오빠가 도와주셨기에 제 시간에 논문을 마칠 수 있었습니다. 오빠는 정말 R신이세요. 하나부터 열까지 다 배우고 싶은 정란언니, 언니의 똑소리나는 일처리는 정말 닮고 싶었어요. 아무것도 모르는 저의 초보적인 질문에도 항상 친절히 알려주셔서 감사했습니다. 우리 연구실의 여신 예지언니, 언니가 있어서 너무 든든했어요. 앞으로 계획하신 모든 일들이 꼭 이루어지도록 응원하겠습니다. 모르는 게 없는 연구실 짱 민수오빠, 연구실에 오빠가 없으면 왠지 모르게

불안해질 정도로 오빠는 연구실의 기둥이에요. 모든 일을 책임지시느라 고생 많으셨고, 앞으로도 파이팅 해주세요. 함께 연구실 생활을 하지는 못했지만 선배로서 여러가지 조언을 아끼지 않았던 상우오빠, 현경이. 학교생활이나 취업준비를 하는데 있어 정말 많은 도움이 되었습니다. 그리고 사랑하는 연구실 동기 서현언니, 용민오빠. 두 사람이 있어 연구실 생활이 외롭지 않고 즐거웠습니다. 졸업하더라도 각자의 위치에서 반짝반짝 빛날 거라고 생각해요. 자주 연락하고, 자주 만납시다. 유일한 연구실 후배, 승민이. 같은 연구실에 있으면서도 얘기할 기회가 많지 않아 아쉬웠어. 남은 일년도 지금처럼 잘해 나갈 거라고 믿어.

대학원에서 만난 소중한 인연, 우리 동기들. 함께 연구실을 썼을 때만큼 자주 보지는 못했지만, 같은 공간에 있는 것만으로도 큰 힘이 되었습니다. 특히 민진언니, 지희, 윤실이. 대학원에서도 이렇게 마음 맞는 친구들을 만날 수 있다는 것이 신기하고 감사했습니다. 그대들과 함께여서 항상 즐거웠고, 힘든 나날 속에서도 웃음을 잃지 않을 수 있었습니다. 우리 또 조만간 떠납시다.

자주 보지는 못하지만 꾸준히 연락하며 지내는 예은이, 현진언니, 은경언니. 가끔 보는 그 만남이 삶의 활력이 되었습니다. 그리고 저와 비슷한 길을 걷고 있는 선민이. 함께 힘든 상황에서도 언제나 나를 먼저 생각해주고 응원해주는 너에게 너무 미안하고 고마웠어. 우리 앞으로도 함께 힘내자. 우리는 '대기만성'형이잖아.

저의 십년지기 친구들 은순이, 수연이, 하늬. 기쁠 때나 슬플 때나 항상 제일 먼저 생각나는 은하수 친구들이 있어 나는 정말 행복해. 우리 앞으로도 함께 웃으며 아름다운 여성이 되자꾸나. 먼저 사회인이 되었다는 이유로 학생인 저에게 물심양면으로 맛있는 식사를 제공했던 미소. 이제는 내 차례야. 앞으로는 내가 더 많이 응원하고, 대접할게. 우리 곧 보자.

그 외에도 저에게 많은 힘이 되어 주신 모든 분들께 진심으로 감사의 인사를 전합니다.

2013년 1월

박 수 진