BAYESIAN WAVELET APPROACHES FOR PARAMETER ESTIMATION AND CHANGE POINT DETECTION IN LONG MEMORY PROCESSES

A Dissertation

by

KYUNGDUK KO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Statistics

BAYESIAN WAVELET APPROACHES FOR PARAMETER ESTIMATION AND CHANGE POINT DETECTION IN LONG MEMORY PROCESSES

A Dissertation

by

KYUNGDUK KO

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Marina Vannucci (Chair of Committee)	
Jeffrey D. Hart (Member)	Narasimha Reddy (Member)
Bani K. Mallick (Member)	Michael T. Longnecker (Head of Department)

August 2004

Major Subject: Statistics

ABSTRACT

Bayesian Wavelet Approaches for Parameter Estimation and Change Point Detection in Long Memory Processes. (August 2004)

Kyungduk Ko, B.A., Yonsei University;

M.A., Yonsei University

Chair of Advisory Committee: Dr. Marina Vannucci

The main goal of this research is to estimate the model parameters and to detect multiple change points in the long memory parameter of Gaussian ARFIMA(p, d, q) processes. Our approach is Bayesian and inference is done on wavelet domain. Long memory processes have been widely used in many scientific fields such as economics, finance and computer science. Wavelets have a strong connection with these processes. The ability of wavelets to simultaneously localize a process in time and scale domain results in representing many dense variance-covariance matrices of the process in a sparse form. A wavelet-based Bayesian estimation procedure for the parameters of Gaussian ARFIMA(p, d, q) process is proposed. This entails calculating the exact variance-covariance matrix of given ARFIMA(p, d, q) process and transforming them into wavelet domains using two dimensional discrete wavelet transform (DWT2). Metropolis algorithm is used for sampling the model parameters from the posterior distributions. Simulations with different values of the parameters and of the sample size are performed. A real data application to the U.S. GNP data is also reported. Detection and estimation of multiple change points in the long memory parameter is also investigated. The reversible jump MCMC is used for posterior inference. Performances are evaluated on simulated data and on the Nile River dataset.

To Se Jeong and Minji,

without your love this achievement was never possible

ACKNOWLEDGEMENTS

This dissertation was possible only with the help and support of others for which I can never take credit. I would like to express my sincere gratitude to all those who have given the most profound impact to me in making this dissertation possible.

I thank my adviser, Marina Vannucci, for her guidance and support throughout the work leading to this dissertation. Her constant encouragement and willingness to let me express my thoughts made the completion of this work possible though it is not "a completion" but "a commencement" toward my future scholarship. I thank her for all that she has taught and has shown as a scholar. My sincere thanks goes to committee members, Jeffery Hart for his insightful comment, Bani Mallick for his advice about Bayesian computation and Narasimha Reddy for his idea on real data.

I am grateful for the excellent faculty at Texas A&M University who have shown an excellence in both their research and teaching. I thank Michael Longnecker for his excellent teaching of the courses in Statistical Methods and enthusiastic willingness to be accessible to the students. I also thank the staff members of the Department of Statistics for their help.

On a personal note, I first thank my mother who has given me many qualities that have enabled me to be strong and ready for the uncertain future in my life. She has always given me her steady love, encouragement and support which have sustained me in my growing up years and even still do today. Above all, I always feel grateful to my father. I also thank my parents-in-law who have given tremendous support and encouragement to me throughout my studies and after-marriage life. Thanks to my brother, sister-in-law, sister and brother-in-law for their support and advice.

I thank my daughter, Minji, who has given me great joy in my life. Her smile helped me to recuperate from my tiredness and life at school. I am very sorry for her because my wife and I forced her to go to daycare center at such an early age so that she has given up time with us that belonged to her. However we could not help doing so because we both have been in Ph.D. program. I hope that she understands our situations. Of all people I am most thankful for my wife, Se Jeong. Without her unselfish love, enduring patience and companionship, it was never possible that I completed the research for this dissertation and received this achievement.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES x

CHAPTER

I INTRODUCTION 1

II LONG MEMORY PROCESSES 5

2.1 Introduction 5
2.2 Features and definition of long memory processes 6
2.3 Class of stationary process with long memory 7
2.4 Heuristic estimation of long memory parameter 14
2.5 Semi-parametric estimation: Geweke and Porter-Hudak's estimate 15
2.6 Parametric estimation of long memory parameter 16

III WAVELETS 18

3.1 Introduction 18
3.2 Prerequisites 18
3.3 Basics of a wavelet decomposition 19
3.4 Multiresolution analysis 21
3.5 Discrete wavelet transform 23
3.6 Variances and covariances of wavelet coefficients 27
3.7 DWT and long memory processes 29

IV BAYESIAN ESTIMATION OF ARFIMA MODELS ON WAVELET DOMAIN 34

4.1 Introduction 34

CHAPTER Page

4.2 Model in the wavelet domain 34
4.3 Bayesian modeling on wavelet domain 35
4.4 Simulation study 38
4.5 Example: U.S. GNP data 41
4.6 Supplementary study on using diagonal elements of Σ_W . 42

V CHANGE POINT ANALYSIS ON WAVELET DOMAIN . . . 47

5.1 Introduction 47
5.2 Model and likelihood 47
5.3 Application of reversible jump MCMC 50
5.4 Simulation studies 55
5.5 Application to Nile River data 60

VI CONCLUSION 63

REFERENCES 65

APPENDIX A 69

APPENDIX B 70

VITA 74

LIST OF FIGURES

FIGURE		Page
1	Yearly minimum water levels of the Nile River from A.D. 622 - A.D. 1281	6
2	The covariance matrix of an ARFIMA$(0, 0.3, 0)$ process with $n = 256$	28
3	The covariance matrix of the wavelet coefficients of the series from Figure 2 using Vannucci-Corradi algorithm with Daubechies' 6 vanishing moments	29
4	The covariance matrix of the wavelet coefficients of the series from Figure 2 using Vannucci-Corradi algorithm with Daubechies' 8 vanishing moments	30
5	First differences of GNP data, in plot a), and kernel estimates of the posterior density of d for b) ARFIMA$(1, d, 0)$, c) ARFIMA$(0, d, 1)$ and d) ARFIMA$(1, d, 1)$	43
6	Nile River data: original data, plot a), and kernel estimates of the posterior density of d before A.D. 720, plot b) and after A.D. 720,plot c). Plot d) shows the MCMC trace of the number of change point	62

LIST OF TABLES

TABLE Page

1 ARFIMA$(1, d, 0)$: Estimates of d and ϕ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors. 38

2 ARFIMA$(0, d, 1)$: Estimates of d and θ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors. 39

3 ARFIMA$(1, d, 1)$: Estimates of d, ϕ and θ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors. 40

4 Wavelet-based Bayes estimates for U.S. GNP data 41

5 Simulation result when diagonal elements and full matrix of Σ_W used in the estimation of long memory parameter of ARFIMA$(0, d, 0)$ model with sample size $n = 128$. The numbers in parenthesis denote biases. 45

6 Simulation result when diagonal elements and full matrix of Σ_W used in the estimation of long memory parameter of ARFIMA$(0, d, 0)$ model with sample size $n = 512$. The numbers in parenthesis denote biases. 46

7 Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 1)$ model with ϕ=0.1, d_1=0.2, d_2=0.4, θ=0.5 and one change point c=256 56

8 Parameter estimates of ARFIMA$(1, d, 1)$ model in the case of $\hat{k} = 1$. . 56

9 Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 0)$ model with ϕ=0.3, d_1=0.2, d_2=0.3 and one change point c=256. 57

TABLE Page

10 Parameter estimates of ARFIMA$(1, d, 0)$ model from simulated data in the case of $\hat{k} = 1$. 57

11 Posterior probability of the number of change points k in the simulation of ARFIMA$(0, d, 1)$ model with d_1=0.2, d_2=0.3, $\theta = 0.3$ and one change point c=256. 58

12 Parameter estimates of ARFIMA$(0, d, 1)$ model from simulated data in the case of $\hat{k} = 1$. 58

13 Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 1)$ model with $\phi = 0.1$, $\theta = 0.4$, $d_1 = 0.05$, $d_2 = 0.35$, $d_3 = 0.45$ and two change points at $c_1 = 128$ and $c_2 = 256$. 59

14 Parameter estimates of ARFIMA$(1, d, 1)$ model from simulated data in the case of $\hat{k} = 2$.. 59

15 Posterior probability of the number of change points k in the Nile River data using ARFIMA$(0, d, 0)$ model. 61

16 Parameter estimates of ARFIMA$(0, d, 0)$ model in the Nile River data when $\hat{k} = 1$. 61

CHAPTER I

INTRODUCTION

Long memory processes have been widely used in many fields, such as economics, finance and telecommunications to model characteristic phenomena. Most remarkable characteristics of long memory process compared to short memory is that dependencies between distant observations are not negligible. Common models for long memory behavior are the fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). Commonly used long memory processes in time series are the ARFIMA(p, d, q), first introduced by Granger and Joyeaux (1980) and Hosking (1984). For these models the value of the spectral density function goes to infinity as the frequency goes to zero and classical time series methods for estimation and testing cannot be applied. Also, the structure of the variance-covariance matrix makes inferential methods computationally expensive and causes inaccurate estimates.

In early stage of the parameter estimation of Gaussian ARFIMA(p, d, q) models, approximate maximum likelihood methods were used by Li and McLeod (1986) and Fox and Taqqu (1986), but these methods showed inaccuracy for finite samples. Sowell (1992a) calculated the exact form of the variance-covariance function to compute the likelihood function under the assumption that the roots of an autoregressive polynomial are simple, that is, the roots do not have repeated roots. This exact likelihood method achieved accuracy, but is computationally exhaustive. Beran (1994) investigated asymptotic sampling theory properties of exact and approximate max-

---

The format and style follow that of *Journal of the American Statistical Association*.

imum likelihood methods. As for Bayesian approaches, Pai and Ravishanker (1996; 1998) adopt the Metropolis algorithm to estimate the model parameters and Koop et al. (1997) used importance sampling with the exact form of the variance-covariance matrix of Sowell (1992a).

While estimation of model parameters has been widely investigated, little work has been done in designing methods for change point analysis of the long memory parameter in ARFIMA(p, d, q) models. Change points in a given series result from unexpected changes in the physical mechanism or environmental condition that generate data. Routine estimation techniques for model parameters may be inaccurate when change points are not properly located. Beran and Terrin (1996) proposed a test for detecting a single change in ARFIMA(p, d, q) model. Whicher, Guttorp and Percival (2000) used Iterative Cumulative Sums of Squares (ICSS) algorithm to test and locate multiple change points in variances of fractional difference process. They took discrete wavelet transform (DWT) of Gaussian I(d) processes to use ICSS which needs independently and identically distributed normal random variables for data. However, their method can only estimate the number of multiple change points and locate them without estimating the values of long memory parameter d_t corresponding to the detected change points. Ray and Tsay (2002) studied multiple change points analysis of the mean level and the long memory parameter. They considered ARFIMA$(0, d, 0)$ models only and used a time-dependent Kalman filter approach with a truncated MA approximation to evaluate the likelihood. Their method allows accurate estimation only if the change points occur at the ends of pre-specified data blocks. Moreover, they used a griddy Gibbs sampler algorithm to estimate the long memory parameter, a procedure that can lead to inaccurate estimates. Also, they used Bayes factor to assess proper number of change points. An alternative Bayesian method for multiple change point analysis was proposed by Liu and Kao (1999).

They adopted griddy Gibbs sampler algorithm but allowed the number of change points to be unknown and used the reversible jump MCMC of Green (1995). These authors considered the special case of ARFIMA$(1, d, 0)$ models with conditionally heteroscedastic innovations.

Wavelets have been known as a powerful tool for the analysis and synthesis of data from long memory processes. The ability of the wavelets to simultaneously localize a process in time and scale domain results in representing many dense matrices in a sparse form. When transforming measurements from a long memory process, wavelet coefficients are approximately uncorrelated, in contrast with the dense long memory covariance structure of the data, see Tewfik and Kim (1992). This property of wavelet transform enables us to simplify dense covariance structure of data following ARFIMA(p, d, q) model into sparse form which we may regard as approximately uncorrelated and thus may use relatively simple likelihood function. McCoy and Walden (1996) used this wavelet property to estimate ARFIMA$(0, d, 0)$ or I(d) models with approximate wavelet coefficients-based maximum likelihood iterative procedure. Jensen (1999) proposed wavelet-based ordinary least squares estimate of the long memory parameter in I(d) process and compared his results to those by Geweke and Porter-Hudak (1983) and McCoy and Walden (1996). Also, Jensen (2000) proposed an alternative maximum likelihood estimator of ARFIMA$(0, d, 0)$ models using compactly supported wavelets.

We here propose an estimation procedure of the model parameters in Gaussian ARFIMA(p, d, q) models and a change point analysis of the long memory parameter in Gaussian ARFIMA(p, d, q) models with unknown multiple change points. In order to handle the dense covariance matrix of ARFIMA(p, d, q) models, we transform the original data into wavelet coefficients using discrete wavelet transform (DWT). A Metropolis algorithm is implemented in the wavelet domain for Bayesian estimation

of the model parameters. For the change point analysis, the reversible jump Markov chain Monte Carlo proposed by Green (1995) is adopted to estimate the unknown number of multiple change points and their locations together with the estimation of other model parameters. For both methods we use Vannucci and Corradi's (1999) algorithm to transform the variance-covariance matrix in the time domain into the corresponding matrix in the wavelet domain.

In Chapter II we introduce the concept and class of long memory process and estimation procedures of long memory processes on literature. Basics of wavelets and relationship between long memory process and wavelet transforms are described in Chapter III. As a preliminary step for change point analysis, we develop a Bayesian estimation method of the parameters of ARFIMA(p, d, q) models on wavelet domain in Chapter IV. In Chapter V we propose a wavelet-based Bayesian change point analysis of the long memory parameter of ARFIMA(p, d, q) processes.

CHAPTER II

LONG MEMORY PROCESSES

2.1 Introduction

Standard time series analysis is concerned with series having the property that two observations separated by a long time interval are completely or nearly independent. However, there are many cases were we should not ignore the dependence between distant observations,even if it is small. It is said that a series having characteristic has "long memory (or long range dependence)". We can find this phenomenon in many fields such as economics, finance, hydrology and engineering. If we apply standard statistics to data having long memory the results may misguide us to wrong decisions. The reason is that variance of the sample mean of such series is not equal to σ^2/n and thus routine estimation, such as interval estimation and testing about population mean, cannot be used.

In general, we can classify the memory type of a given time series in three ways: no memory, short memory and long memory. No memory series has no pattern over time and knowing the past of the series provides no information about its behavior in the future except for mean and variance. A white noise is a typical case of no memory. On the contrary, time series possessing short memory has exponentially decaying autocorrelation function. ARMA(p, q) processes are short memory processes. Long memory process can be non-stationary but typically it is predictable. In this case we need to identify some transformations to reduce the process to short memory type-process and then model the process by a whitening filter to have no memory residuals.

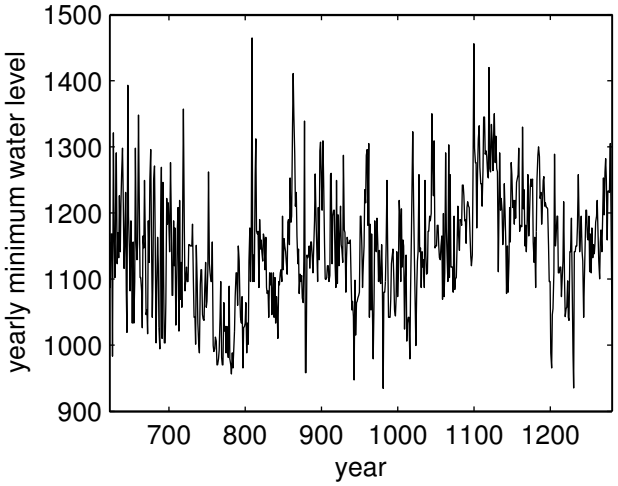

Figure 1: Yearly minimum water levels of the Nile River from A.D. 622 - A.D. 1281

2.2 Features and definition of long memory processes

The qualitative behavior of a long memory process, for which the first two moments exist, can be described as follows. First, the path of series look like stationary. Second, there is no lasting cycle or trend through the whole series although we could find cycles or local trends on short time periods of the path. A typical example showing these features is the Nile River data (A.D. 622-A.D. 1281). If we have a closer look at a short time period of Figure 1 then we could find cycles or trends but we cannot find those patterns in the whole series. On the other hand, quantitative properties of a long memory process are (i) the variance of the sample mean of this process decays to zero at a slower rate than n^{-1} and (ii) the sample correlations decay hyperbolically to zero. These are main differences between short and long memory processes.

There are some mathematical definitions of a long memory process with stationarity in terms of autocovariance and spectral density.

Definition 2.2.1 *Suppose that X_t is a process with autocovariance function $\gamma(\tau) \sim C(\tau)\tau^{2d-1}$ as $\tau \to \infty$, $C(\tau) \neq 0$. Then we call X_t a process with long memory.*

Moreover, if $0 < d < 0.5$ *then the process is stationary.*

Here $C(\tau)$ does not depend on d and is a slowly varying function as $|\tau| \to \infty$ (Beran and Terrin, 1996). The d is called long memory(or long range) parameter. From the definition 2.2.1, we know that the correlations are not summable. Some authors give further distinction according to the range of d, that is, if $d < 0$ and hence $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty$ then the process is "intermediate" and the process is "long memory" if $0 < d < 0.5$ and hence $\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| = \infty$. The following definition describes the property of long memory in terms of the spectral density.

Definition 2.2.2 *Suppose that a process* X_t *has the spectral density such that* $f(\lambda) \sim k(\lambda)\lambda^{-2d}$ *as* $\lambda \to 0$. *Then we call* X_t *a stationary process with long memory if* $0 < d < 0.5$.

From the definition 2.2.2 the spectral density of a process with long memory has a pole at zero.

The hydrologist Hurst (1951) proposed the so-called "Hurst" (or self-similar) parameter, H, which has a simple relationship with the long memory parameter, that is $H = d + \frac{1}{2}$. He noticed that the rescaled adjusted range, R/S statistic, has asymptotically a log linear relationship with the sample size and a slope H larger than $\frac{1}{2}$.

2.3 Class of stationary process with long memory

A stationary process with long memory can be classified in two ways, continuous and discrete time. In the continuous time domain, we have three processes with long memory: self-similar process, fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). A typical discrete long memory process is the fractional ARIMA. Self-similar and ARFIMA$(0, d, 0)$ processes are the basic processes of continuous and discrete long memory processes, respectively.

2.3.1 Continuous long memory process

Fractional Brownian motion and fractional Gaussian noise belong to continuous long memory process. They can be formulated from self-similar process.

Definition 2.3.1 *A stochastic process, X_t with continuous time parameter such that $c^{-H}X_{ct} =^d X_t$ is called self-similar with self-similarity parameter H.*

Here "$=^d$" means "equal in distribution". A real-valued process $\{X_t, t \in T\}$ has stationary increments if $\{X_{t+h} - X_h, t \in T\} =^d \{X_t - X_0, t \in T\}$ for all $h \in T$. Then a Gaussian self-similar process $\{B_H(t)\}$ with stationary increments is called fractional Brownian motion when its self-similarity parameter, H is in $(0, 1)$. When $var(B_H(1)) = \sigma^2$, the autocovariance function is

$$\gamma(s,t) = \text{cov}(B_H(s), B_H(t)) = .5\sigma^2[t^{2H} - (t-s)^{2H} + s^{2H}], \quad s < t.$$

Thus fBm has stationary increments but is not stationary process. Let's define a increment process of $B_H(\cdot)$ as $Y_t = B_H(t+1) - B_H(t)$ which is a stationary sequence. We call Y_t fractional Gaussian noise. The covariance function of the increment process is

$$\gamma(\tau) = \text{cov}(Y_t, Y_{t+\tau}) = .5\sigma_0^2[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}], \tag{2.1}$$

where $EY_t^2 = EB_H(1)^2 = \sigma_0^2$. Also we know that fGn is a stationary process because (2.1) does not depend on t. The correlation function of Y_t is calculated by

$$\rho(\tau) = \gamma(\tau)/\gamma(0) = .5[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}]$$

As $\tau \to \infty$, using Taylor's expansion,

$$\rho(\tau) \sim H(2H-1)\tau^{2H-2}. \tag{2.2}$$

Thus fGn is long memory process. The correlations are not summable for $0.5 < H < 1$ $(0 < d < 0.5)$ in which Y_t has long memory, summable for $0 < H < 0.5$ $(-0.5 < d < 0)$ and Y_t's are uncorrelated for $H = 0.5$ $(d = 0)$.

2.3.2 Discrete long memory process

The simplest of ARFIMA(p, d, q) models is ARFIMA$(0, d, 0)$ model which is also called integrated process I(d). A process $\{X_t\}_{t\in\mathbb{Z}}$ is called ARFIMA$(0, d, 0)$ model if it has relationship $(1-B)^d X_t = \varepsilon_t$ for any real $d > -0.5$, where ε_t is zero mean white noise with innovation variance σ^2. Fractional difference operator $\Delta^d = (1-B)^d$ where B is backshift operator can be defined as power series in B for $d \neq 0$. By binomial expansion, for any real number d,

$$(1-B)^d \equiv \sum_{j=0}^{\infty} \pi_j B^j = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j, \tag{2.3}$$

where $\binom{d}{j} = d!/j!(d-j)! = \Gamma(d+1)/\Gamma(j+1)\Gamma(d-j+1)$. This comes from the fact that gamma function is defined for all real numbers and the binomial coefficient can be extended to all real number d. Since $\pi_j = (-1)^j \binom{d}{j}$ can be approximated as

$$\begin{aligned}
(-1)^j \binom{d}{j} &= (-1)^j \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)} \\
&= (-1)^j \frac{(-1)^j(-d)(-d+1)\cdots(-d-1+j)}{\Gamma(j+1)} \\
&= (-1)^j \frac{(-1)^j \Gamma(-d+j)}{\Gamma(j+1)\Gamma(-d)} \\
&= \frac{\Gamma(-d+j)}{\Gamma(j+1)\Gamma(-d)} \\
&\sim j^{-d-1}/\Gamma(-d) \quad as\ j \to \infty,
\end{aligned}$$

the I(d) process can be expressed as an AR(∞) of the form

$$(1-B)^d X_t = \varepsilon_t \Rightarrow \sum_{j=0}^{\infty} j^{-d-1}/\Gamma(-d) B^j X_t = \varepsilon_t \quad as\ j \to \infty \tag{2.4}$$

under the condition such that $\sum_{j=1}^{\infty} \pi^2 < \infty$ in (2.4). Since $\pi_j^2 \sim j^{-2d-2}/\Gamma^2(-d)$, for the convergence of the summation, d should satisfy

$$\begin{aligned}
-2d-2 < -1 &\Rightarrow -2d < 1 \\
&\Rightarrow d > -0.5.
\end{aligned}$$

Thus I(d) process or ARFIMA$(0, d, 0)$ model is invertible if $d > -0.5$. For the condition such that $\{X_t\}_{t\in\mathbb{Z}}$ is stationary process, assume that $\Delta^d X_t = \varepsilon_t$ be inverted into $X_t = \Delta^{-d}\varepsilon_t$. Since $\Delta^{-d} = (1-B)^{-d} \equiv \sum_{j=0}^{\infty}\psi_j B^j = \sum_{j=0}^{\infty}\binom{-d}{j}(-1)^j B^j$ and

$$\begin{aligned}
(-1)^j\binom{-d}{j} &= (-1)^j\frac{\Gamma(-d+1)}{\Gamma(j+1)\Gamma(-d-j+1)} \\
&= (-1)^j\frac{(-1)^j(d)(d+1)\cdots(d+j-1)}{\Gamma(j+1)} \\
&= (-1)^j\frac{(-1)^j\Gamma(d+j)}{\Gamma(j+1)\Gamma(d)} \\
&= \frac{\Gamma(d+j)}{\Gamma(j+1)\Gamma(d)} \\
&\sim j^{d-1}/\Gamma(d) \quad as\ j\to\infty,
\end{aligned}$$

the inverted process can be expressed as an MA(∞) of the form

$$\begin{aligned}
(1-B)^d X_t = \varepsilon_t \quad &\Rightarrow \quad X_t = \Delta^{-d}\varepsilon_t \\
&\Rightarrow \quad X_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \qquad (2.5)
\end{aligned}$$

under the condition such that $\sum_{j=1}^{\infty}\psi_j^2 < \infty$ in (2.5). Since $\psi_j^2 \sim j^{2d-2}/\Gamma^2(d)$, for the convergence of the summation, we also need a condition for d such that

$$\begin{aligned}
2d-2 < -1 \quad &\Rightarrow \quad 2d < 1 \\
&\Rightarrow \quad d < 0.5.
\end{aligned}$$

So ARFIMA$(0, d, 0)$ model is stationary if $d < 0.5$. Therefore, ARFIMA$(0, d, 0)$ model is invertible and stationary if $d \in (-0.5, 0.5)$. In the case of $d \in (-0.5, 0.5)$, $\{X_t\}_{t\in\mathbb{Z}}$ has the spectral measure $dZ_X(\lambda) = |B(\lambda)|dZ_\varepsilon(\lambda) = (1-e^{-i\lambda})^{-d}dZ_\varepsilon(\lambda)$ and

the spectral density

$$\begin{aligned} f_X(\lambda) &= B(\lambda)^2 dZ_\varepsilon(\lambda) \\ &= |1-e^{-i\lambda}|^{-2d}\frac{\sigma^2}{2\pi} \\ &= |2\sin(\lambda/2)|^{-2d}\frac{\sigma^2}{2\pi} \\ &\sim \frac{\sigma^2}{2\pi}|\lambda|^{-2d}, \quad \lambda \to 0, \end{aligned} \tag{2.6}$$

where σ^2 is the white noise variance. Thus from definition (2.2.2) ARFIMA$(0,d,0)$ process has long memory.

A time series, $\{x_t\}_{t\in\mathbb{Z}}$, is said to be fractionally differenced autoregressive moving average model if the series is identified as an ARMA(p,q) model after applying the fractional difference operator (2.3). The general fractionally differenced ARMA(p,q) model can be written as

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \tag{2.7}$$

where

$$\Phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p$$

and

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

are polynomials in the backshift operator B and ε_t is a white noise process with zero mean and variance σ^2. We denote the above as ARFIMA(p,d,q) model. The model (2.7) can be rewritten by

$$\Phi(B)X_t = \Theta(B)U_t \qquad \text{and} \qquad U_t = (1-B)^{-d}\varepsilon_t. \tag{2.8}$$

Thus we can regard ARFIMA(p,d,q) process as ARMA(p,q) process driven by I(d) process, $U_t = (1-B)^{-d}\varepsilon_t$. Since ARMA$(p,q)$ model is typical short memory process

in time series and I(d) process has long memory, the long term behavior is determined by d and the short term behavior is determined by $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ in ARFIMA(p, d, q) models. That is, d and ($\boldsymbol{\phi}$, $\boldsymbol{\theta}$) describe the high-lag and low-lag correlation structures of ARFIMA(p, d, q) models. Thus the long-term behavior of ARFIMA(p, d, q) models may be similar to the one of ARIMA$(0, d, 0)$ because the influence of the ϕ's and θ's is negligible for distant observations. From this fact, the ARIMA(p, d, q) process has long memory if $d \in (-0.5, 0.5)$ and in addition, is weak stationary and invertible if and only if the roots of $\Phi(z)$ and $\Theta(z)$ are outside the unit circle. Geweke and Poter-Hudak (1983) used two-step procedures with the two models in (2.8) to estimate the model parameters of ARFIMA(p, d, q) models. Hosking (1984) proposed the algorithm for simulating an ARFIMA(p, d, q) process with the same notion, which is (i) to generate U_t from ARFIMA$(0, d, 0)$ and (ii) to generate X_t using U_t from ARMA(p, q) process.

Since the spectral measure of $\{X_t\}_{t\in\mathbb{Z}}$ in (2.7) is $dZ_X(\lambda) = \theta(e^{-i\lambda})\phi(e^{-i\lambda})^{-1}(1 - e^{-i\lambda})^{-d}dZ_\varepsilon(\lambda)$ where $\theta(\lambda) = \Theta(e^{-i\lambda})$ and $\phi(\lambda) = \Phi(e^{-i\lambda})$ are transfer functions related to AR and MA terms, its spectral density and asymptotic autocovariance function are

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi}\frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}|1 - e^{-i\lambda}|^{-2d} \\ &\sim \frac{\sigma^2}{2\pi}\frac{|\theta(1)|^2}{|\phi(1)|^2}|\lambda|^{-2d} \quad , as\ \lambda \to 0 \end{aligned} \tag{2.9}$$

and

$$\gamma(\tau) \sim C\tau^{2d-1}\ as\ \tau \to \infty, \tag{2.10}$$

where $C \neq 0$ and does not depend on τ. On the other hand, under the additional assumption that the roots of $\Phi(z)$ are simple, Sowell (1992a) showed that the auto-

covariance function of this process is

$$\gamma(\tau) = \sigma^2 \sum_{l=-q}^{q} \sum_{j=1}^{p} \psi(l)\zeta_j C(d, p-\tau+l, \rho_j) \tag{2.11}$$

where

$$\psi(l) = \sum_{\tau=max[0,l]}^{min[q,q-l]} \theta_\tau \theta_{\tau-l},$$

$$\zeta_j = [\rho_j \prod_{i=1}^{p}(1-\rho_i\rho_j) \prod_{m\neq j}(\rho_j - \rho_m)]^{-1},$$

and

$$\begin{aligned} C(d,h,\rho) &= \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(1-d+h)\Gamma(1-d)\Gamma(d)} \\ &\quad \times[\rho^{2p}F(d+h,1;1-d+h;\rho) + F(d-h,1;1-d-h;\rho) - 1]. \end{aligned}$$

Here $F(a,1;c;\rho)$ is the hypergeometric function and has the recursive relationship

$$F(a,1;c;\rho) = \frac{c-1}{\rho(a-1)}[F(a-1,1;c-1;\rho) - 1].$$

For ARFIMA$(0,d,q)$ model i.e. $p=0$, the autocovariance function has the form

$$\gamma(\tau) = \sigma^2 \sum_{l=-q}^{q} \psi(l) \frac{\Gamma(1-2d)\Gamma(d+l-\tau)}{\Gamma(1-d+l-\tau)\Gamma(1-d)\Gamma(d)}. \tag{2.12}$$

Moreover, for p=0 and q=0 in ARFIMA(p,d,q) model, it becomes

$$\gamma(\tau) = \sigma^2 \frac{\Gamma(1-2d)\Gamma(d+\tau)}{\Gamma(1-d+\tau)\Gamma(1-d)\Gamma(d)}.$$

These processes are stationary if $d < 0.5$ and possess invertibility or AR(∞) representation if $d > -0.5$. Also, they have long memory for $0 < d < 0.5$, short memory for $d = 0$ and intermediate for $-0.5 < d < 0$. Sowell (1992a) use the Levinson algorithm to reduce the order of calculating the likelihood function by decomposing the covariance.

2.4 Heuristic estimation of long memory parameter

Given a series we need to know if it has long memory. For this purpose we can use some heuristic methods based on the properties and definitions of long memory mentioned in previous section. The most basic checking method for long memory is to investigate whether the ACF plot decays hyperbolically to zero. Other heuristic methods are based on plots of basic statistics in log scale and least square fits of them.

2.4.1 the R/S statistic versus k

Let X_i denote a value of given process at time i and define $Y_j = \sum_{i=1}^{j} X_i$. The modified range $R(t,k)$ and its scaling factor $S(t,k)$ are defined as follows.

$$R(t,k) = \max_{0\le i\le k}[Y_t + i - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)] - \min_{0\le i\le k}[Y_t + i - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)]$$

and

$$S(t,k) = [k^{-1}\sum_{i=t+1}^{t+k}(X_i - \bar{X}_{t,k})^2]^{1/2}$$

In the plot of $\log[R(t,k)/S(t,k)]$ versus $\log k$, the points are scattered around a straight line of slope 0.5 for *i.i.d.* series and are scattered around a straight line of slope greater than 0.5 for long memory processes. That is, in the least square line of $\log E(R/S) \approx c + H\log k$, if $H \approx 0.5$ then the process is *i.i.d.* and if $H > 0.5$ then the process has long memory.

2.4.2 Sample variance versus its sample size

The variance of the sample mean of long memory process decays to zero at a slower rate than n^{-1}. So plotting sample variance versus sample size in log scale and fitting a least square line can be used for the estimation of long memory parameter. If the slope of the line is far from -1 then we conclude this process has long memory.

Actually the slope -1 is a theoretical value for summable correlations. Also if correlations are summable the periodogram ordinates near the origin should be scattered randomly around a constant.

2.4.3 Correlogram

From definition 2.2.1 we might draw a plot of $\rho(\tau) = \gamma(\tau)/\gamma(0)$ versus τ in log scale and fit a least square line. If the slope is between -1 and 0 we conclude that the series exhibits long range dependence.

2.5 Semi-parametric estimation: Geweke and Porter-Hudak's estimate

Definition 2.2.2 needs to be carefully understood in that the spectral density is proportional to λ^{-2d} only in a neighborhood of zero. Geweke and Porter-Hudak (1983) used this notion to get a least square estimate of long memory parameter d at low frequencies in the case of integrated models I(d). They used the periodogram $I(\lambda)$ and the Fourier frequencies $\lambda_{i,n}$ as the sample estimates of the spectral density and frequencies, respectively in the equation taking logarithm of the asymptotic relation of $f(\lambda)$. This leads to

$$\log I(\lambda_{i,n}) \approx \log k - 2d\log\lambda_{i,n} + \log\varepsilon_i \tag{2.13}$$

If the least square estimate of the slope in the equation (2.13) is $\hat{\beta}$ then the estimate of long memory parameter is

$$\hat{d} = -\frac{\hat{\beta}}{2}$$

GPH estimator has been widely used with desirable precision in estimation of long memory parameter by virtue of its simplicity and computational speed. For fractional ARFIMA(p, d, q) models Geweke and Porter-Hudak also proposed two-step estimation procedures where the d is estimated under the model I(d), the data is transformed

into $Y_t = (1-B)^{\hat{d}}$ and standard ARMA models are applied to the transformed U_t in order to estimate AR and MA parameters and identify the orders.

2.6 Parametric estimation of long memory parameter

Suppose that an ARFIMA(p,d,q) model is fitted to a Gaussian process $\{X_t\}$ with size n. Then the likelihood function is

$$L(\Psi) = (2\pi)^{-n/2}|\Sigma(\Psi)|^{-1/2}\exp\left\{-\frac{1}{2}X'\Sigma(\Psi)^{-1}X\right\}, \tag{2.14}$$

where $\Psi = (\sigma_\varepsilon^2, d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$, σ_ε^2 is the innovation variance in (2.7) and $\Sigma(\Psi)$ is the $n \times n$ covariance matrix of X. The MLE, $\hat{\Psi}$, is the value of maximizing $L(\Psi)$ or minimizing $l(\Psi) = -2\log L(\Psi)$. If we define the $(p+q+2)$-dimensional vector

$$l'(\Psi) = \frac{\partial}{\partial\Psi}l'(\Psi), \tag{2.15}$$

then the MLE, $\hat{\Psi}_E$ under mild regularity condition is the solution of the system of $(p+q+2)$-equations

$$l'(\hat{\Psi}) = 0. \tag{2.16}$$

Yajima (1985) proves the limit theorem of $\hat{\Psi}$ in the case of ARFIMA$(0,d,0)$ process and Dahlhaus (1989) shows the result for ARFIMA(p,d,q) model. If the dimension of Ψ is large or the data size n is big enough then the calculation of the exact MLE is computationally intensive because the likelihood function (2.14) is the implicit function of Ψ through the covariance matrix $\Sigma(\Psi)$ and thus (2.16) should be evaluated for many trial values of Ψ. The other problem is to get the inverse of the covariance matrix, $\Sigma(\Psi)^{-1}$. Note that a large number of data is needed to estimate the long memory parameter with desirable precision because it describes the long-term persistence between distant observations. The inversion of a large size of the covariance may be numerically unstable.

An alternative method to avoid those problems is to use an approximation to the likelihood function. All $n \times n$ symmetric Toeplitz matrices have complex orthonormal eigenvectors and the corresponding eigenvalues which can be well approximated by

$$V_j = \sqrt{n^{-1}}\{\exp(-i\lambda_j t)\}_{t=1,\ldots,n-1} \text{ and } \Gamma_j = \{2\pi f_\Psi(\lambda_j)\}, j = 1, \ldots, n-1,$$

where V_j is jth column vector of complex eigenvector matrix $\boldsymbol{V}$, Γ_j is jth diagonal element of the diagonal matrix $\boldsymbol{\Gamma}$ and $f_\Psi(\lambda)$ is the spectral density of the process. Thus

$$\Sigma(\Psi) \approx \boldsymbol{V\Gamma V^c}, \tag{2.17}$$

where $\boldsymbol{V^c}$ is the conjugate transpose of $\boldsymbol{V}$. Thus

$$\begin{aligned} l(\Psi) &= n\log(2\pi) + \log|\Sigma(\Psi)| + X'\Sigma(\Psi)^{-1}X \\ &\approx n\log(2\pi) + \log|\boldsymbol{V\Gamma V^c}| + X'\boldsymbol{V\Gamma^{-1}V^c}X' \qquad (2.18) \\ &= n\log(2\pi) + \sum_{j=1}^{n-1}[\log(2\pi f_\Psi(\lambda_j)) + n|J(\lambda_j)|^2/(2\pi f_\Psi(\lambda_j))] \\ &= 2n\log(2\pi) + \sum_{j=1}^{n-1}[\log f_\Psi(\lambda_j) + I(\lambda_j)/f_\Psi(\lambda_j)], \qquad (2.19) \end{aligned}$$

where $I(\lambda_j)$ is the periodogram of X_t and $J(\lambda_j) = n^{-1}\sum_{t=0}^{n-1} X_t \exp(-i\lambda_j t)$ which is the discrete Fourier transform of X_t. Note that $X'\boldsymbol{V} = \sqrt{n}(J(\lambda_0), \ldots, J(\lambda_{n-1}))$ because

$$\begin{aligned} X'V_j &= \sqrt{n^{-1}}\sum_{t=0}^{n-1} X_t \exp(-i\lambda_j t) \\ &= \sqrt{n}J(\lambda_j). \end{aligned}$$

The estimate, $\hat{\psi}_W$ which minimize (2.18) is called Whittle's approximate MLE. Fox and Taqqu (1986) prove the limit theorem of $\hat{\psi}_W$ for a Gaussian ARFIMA process.

CHAPTER III

WAVELETS

3.1 Introduction

Wavelets are relatively new mathematical tools for time series analysis and image analysis. Wavelet means "small wave". On the contrary, examples of "big waves" are *sine* and *cosine* function usually used for Fourier analysis. Wavelets are the building blocks of wavelet transformations in the same way as the functions $e^{inx} = \cos(nx) + i\sin(nx)$ are the building blocks of the Fourier transformation.

Wavelet methods are used in statistics for smoothing of noisy data, nonparametric density and function estimation, and stochastic process representation. A popular application in nonparametric statistics is wavelet shrinkage, which can be described as 3 step procedure: (i) the original series is transformed into a set of wavelet coefficients, (ii) a shrinkage of those coefficients is applied, and (iii) the resulting wavelet coefficients are transformed back to the original data domain. Nowadays the application of wavelets in statistics is rapidly growing and expanding to other areas like economics, finance and so on.

Wavelet theory is similar to Fourier analysis but it has a critical advantage. Wavelet transforms well localize original series in both time and frequency (scale) domains, while Fourier transforms do so only in the frequency domain. Thus one loses time information through Fourier transform but not via wavelet transform.

3.2 Prerequisites

Here some useful mathematical definitions related to wavelet theory are described. In wavelet theory we deal with functions in $\mathbb{L}_2(\mathbb{R})$ and $\ell_2(\mathbb{R})$.

Definition 3.2.1 *The space of all square integrable functions is called* $\mathbb{L}_2(\mathbb{R})$. *That is,* $f(x) \in \mathbb{L}_2(\mathbb{R})$ *if* $\int |f(x)|^2 < \infty$.

The inner product of two functions and norm of a function in $\mathbb{L}_2(\mathbb{R})$ are defined as $< f, g >= \int fg$ and $\|f\| = \sqrt{\int f^2}$, respectively.

Definition 3.2.2 *The space of all square summable sequences is called* $\ell_2(\mathbb{R})$. *That is,* $x_1, \ldots, x_n \in \ell_2(\mathbb{R})$ *if* $\sum_{i=1}^{n} x_i^2 < \infty$.

The inner product of two sequences and norm of a sequence on $\ell_2(\mathbb{R})$ are defined as $< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{n} x_i y_i$ and $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} {x_i}^2}$, respectively. The following definition plays an important role in wavelet decomposition and synthesis.

Definition 3.2.3 *Suppose that* V *is an inner product space and* W *a finite dimensional subspace of* V. *For* $v \in V$, *the orthogonal projection of* v *onto* W *is the unique vector* $v_0 \in W$ *such that*

$$\|v - v_0\| = min_{w \in W} \|v - w\|.$$

Here the orthogonal complement of W in V, $W^{\perp}$ is defined as $W^{\perp} = \{v \in V | < v, w >= 0\}$ for all $w \in W$. We can represent V as the orthogonal sum of W and $W^{\perp}$, $V = W \oplus W^{\perp}$.

3.3 Basics of a wavelet decomposition

There are two basic functions in wavelet analysis, the scaling function $\phi(x)$ and the (mother) wavelet function $\psi(x)$. Sometimes $\phi(x)$ is called "father wavelet". The building blocks to approximate a given function are constructed by the translations and dilations of the scaling function. Note that the translation $\phi(x-k)$ has the same shape as $\phi(x)$ except translated by k units and the dilation $\phi(2^j x)$ has the same shape as $\phi(x)$ but its width is as 2^{-j} times as the one of $\phi(x)$. The simplest scale function

is the Haar scaling function defined as

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & otherwise \end{cases} \tag{3.1}$$

Let V_j be the space of all functions of the form $\sum_{k\in\mathbb{Z}} a_k \phi(2^j x - k), \quad j > 0, \ a_k \in \mathbb{R}$ and k is a finite set of integers. In other words, V_j is defined to be the space spanned by the set

$$\{\ldots, \phi(2^j x + 2), \phi(2^j x + 1), \phi(2^j x), \phi(2^j x + 1), \phi(2^j x + 2), \ldots\}.$$

Thus V_j is the space of piecewise constant functions with finite support whose discontinuities are in the set $\{\ldots, -2/2^j, -1/2^j, 0, 1/2^j, 2/2^j, \ldots\}$. Any function in V_0 is contained in V_1 and likewise $V_1 \subset V_2$. The following hierarchical relation is established.

$$\ldots V_0 \subset V_1 \subset \ldots \subset V_{j-1} \subset V_j \subset V_{j+1} \ldots \tag{3.2}$$

V_j contains all relevant informations up to a resolution scale 2^{-j}. The larger gets j, the finer is resolution. We know from the containment that information is not lost as the resolution gets finer. The collection of functions $\phi_{j,k}(x) = \{2^{j/2}\phi(2^j x - k); j, k \in \mathbb{Z}\}$ is an orthonormal basis of V_j.

The mother wavelet function $\psi(x)$ enters into wavelet decomposition of an original function f to catch "spikes" which belong to V_j and do not belong to V_{j-1}. So V_j is decomposed as an orthogonal sum of V_{j-1} and its orthogonal complement W_{j-1}, which is denoted by $V_j = V_{j-1} \oplus W_{j-1}$ where $< v_{j-1}, w_{j-1} >= 0, \quad v_{j-1} \in V_j$ and $w_{j-1} \in W_{j-1}$. The orthogonal complement of V_j, W_{j-1} is generated by the translations and dilations of $\psi(x)$ as if V_j is done by the translations and dilations of $\phi(x)$. As in the case of $\phi_{j,k}(x)$, the collection of functions $\psi_{j,k}(x) = \{2^{j/2}\psi(2^j x - k); j, k \in \mathbb{Z}\}$ is an orthonormal basis of W_j. The two properties are needed to construct mother wavelet

function $\psi_{j,k}(x)$, which are (i) $\int_{-\infty}^{\infty} \psi_{j,k}(x)dx = 0$ and (ii) $\int_{-\infty}^{\infty} \psi_{j,k}(x)^2 dx = 1, \ \ \forall j,k \in \mathbb{Z}$. The simplest mother wavelet function is the Harr wavelet which is defined by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1/2 \\ -1 & \text{if } 1/2 \leq x < 1 \\ 0 & otherwise \end{cases} \tag{3.3}$$

Note that the Haar wavelet function (3.3) can be denoted by a linear combination of the Haar scaling functions (3.1) as $\psi(x) = \phi(2x) - \phi(2x-1)$.

The main idea of wavelet decomposition is to orthogonally project a given function $f \in \mathbb{L}_2$ onto the space of scaling functions at a resolution j, say V_j, which is $f \simeq f_{V_j} = \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x)$. Then f_{V_j} is again decomposed into the space of the next coarser scale V_{j-1} through the relation (3.2). This is the main stem of multiresolution analysis (MRA) by Mallat (1989).

3.4 Multiresolution analysis

A multiresolution analysis (Mallat, 1989) is a framework for creating general ϕ and ψ. MRA is a decomposition of a function in $\mathbb{L}_2$ into scaling basis $\phi_{j,k}(x)$ and wavelet basis $\psi_{j,k}(x)$. It is an embedded grid of approximation by (3.2). Approximation of $f \in \mathbb{L}_2$ at a resolution j is an orthogonal projection f_j of f on V_j, which means that $\|f - f_j\|$ is minimized (Definition 3.2.3).

Definition 3.4.1 *A sequence $\{V_j\}_{j\in\mathbb{Z}}$ of subspaces of functions in $\mathbb{L}_2$ is called a multiresolution analysis with scaling function ϕ if the following properties holds.*

(1) $V_j \subset V_{j+1}, \quad \forall j \in \mathbb{Z}$

(2) $\overline{\cup_{j\in\mathbb{Z}} V_j} = \mathbb{L}_2$ *and* $\cap_{j\in\mathbb{Z}} V_j = \phi$

(3) $f(x) \in V_j$ *if and only if* $f(2x) \in V_{j+1}, \quad \forall j \in \mathbb{Z}$

(4) $f(x) \in V_0$ *if and only if* $f(x+k) \in V_0, \quad \forall k \in \mathbb{Z}$

(5) There exists a $\phi(x) \in V_0$ *such that* $\{\phi(x-k);\ k \in \mathbb{Z}\}$ *is an orthonormal basis*

for V_0.

The $\{V_j\}_{j\in\mathbb{Z}}$'s are also called *approximation spaces*. The only condition for choosing a ϕ is that the set of its translations, $\{\phi(x-k); k \in \mathbb{Z}\}$ is a basis. Generally the most useful property of scaling functions is to have *compact support* which means that a function has identically zero outside of a finite interval and *continuity* because a scaling function having these two properties has faster computing time and better performance for decomposition or reconstruction.

Suppose that $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis with scaling function ϕ. Then $\phi_{j,k}(x) = \{2^{j/2}\phi(2^j x - k); k \in \mathbb{Z}\}$ is an orthonormal basis for V_j. Also the following two central equations in multiresolution analysis hold.

- **(Two-scale equation)** There exists $\{h_k\}_{k\in\mathbb{Z}}$ such that $\phi(x) = \sum_{k\in\mathbb{Z}} h_k 2^{1/2}\phi(2x - k) = \sum_{k\in\mathbb{Z}} h_k \phi_{1,k}$, where $h_k =< \phi(x), \phi_{1,k} >\in \ell_2(\mathbb{R})$. $\{h_k\}$ is called low-pass filter(or scaling filter).

- **(Wavelet equation)** There exists $\{g_k\}_{k\in\mathbb{Z}}$ such that $\psi(x) = \sum_{k\in\mathbb{Z}} g_k 2^{1/2}\phi(2x - k) = \sum_{k\in\mathbb{Z}} g_k \phi_{1,k}$, where $g_k =< \psi(x), \phi_{1,k} >\in \ell_2(\mathbb{R})$. $\{g_k\}$ is called high-pass filter(or wavelet filter).

The above two equations are forms of convolution with filter coefficients $\{h_k\}$ and $\{g_k\}$, respectively. The inner products are performed according to the Definition 3.2.1 or 3.2.2 depending on the attribute of x. From the two central equations, it can be shown that $\psi_{j,k}(x) = \{2^{j/2}\psi(2^j x - k); k \in \mathbb{Z}\}$ is orthonormal basis for W_j which is

the orthogonal complement of V_j in V_{j+1}. By successive orthogonal decompositions,

$$\begin{aligned} V_{j+1} &= W_j \oplus V_j \\ &= W_j \oplus W_{j-1} \oplus V_{j-1} \\ &= \dots \\ &= W_j \oplus W_{j-1} \oplus \cdots \oplus W_1 \oplus W_0 \oplus V_0 \end{aligned} \tag{3.4}$$

If $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis with scaling function ϕ and W_j is the orthogonal complement of V_j in V_{j+1}, then

$$\mathbb{L}_2(\mathbb{R}) = \cdots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus \cdots . \tag{3.5}$$

Therefore $f \in \mathbb{L}_2(\mathbb{R})$ can be denoted by a unique sum of $w_k \in W_k,\ k \in (-\infty, \infty)$ and so $\psi_{j,k}(x)$ is an orthonormal basis for $\mathbb{L}_2(\mathbb{R})$.

3.5 Discrete wavelet transform

For decomposition $f \in \mathbb{L}_2$ is orthogonally projected on V_j, that is, $f \approx f_j \in V_j$. First decomposition is started with f_j into a coarser approximation part, $f_{j-1} \in V_{j-1}$ and wavelet(detail) part $w_{j-1} \in W_{j-1}$. From the equation (3.4), $f_j = f_{j-1} + w_{j-1}$. Then the decomposition is repeated with f_{j-1} and so on. Reconstruction algorithm is the reverse of decomposition.

There are two kinds of wavelet decomposition, continuous wavelet transform (CWT) and discrete wavelet transform (DWT). CWT is designed to decompose series defined over the entire real axis and DWT is used to transform series over a range of integers. Discrete wavelet transform is the basic tool for time series analysis and has an analogous feature to the discrete Fourier transform in spectral analysis. DWT has a property to effectively decorrelate highly correlated time series under certain condition and with this reason it is often used for time series analysis. Pyramid

algorithm (or Cascade algorithm) by Mallat (1989) is mainly used for discrete wavelet transform.

- **(Pyramid algorithm)** Given approximation coefficients at level j, wavelet and approximation coefficients of all coarser levels can be computed by iterative equations,

$$\begin{aligned} c_{j,k} &= \sum_{l\in\mathbb{Z}} h_{l-2k} c_{j+1,l} \\ d_{j,k} &= \sum_{l\in\mathbb{Z}} g_{l-2k} c_{j+1,l} \end{aligned} \tag{3.6}$$

  Filter coefficients $\{g_m\}$ and $\{h_m\}$ is called high-pass filter and low-pass filter, respectively. The relation between high-pass filter and low-pass filter is

$$g_m = (-1)^m h_{1-m}. \tag{3.7}$$

  The equations (3.6) is simply a rephrase of the two-scale equation and wavelet equation.

The pyramid algorithm transforms original series X with length $N = 2^J$ into the $N/2$'s wavelet coefficients $\mathcal{W}_1 = \{d_{J-1,k}; k = 1, \ldots, N/2\}$ and $N/2$'s scaling coefficients $\mathcal{V}_1 = \{c_{J-1,k}; k = 1, \ldots, N/2\}$. Then $\mathcal{V}_1$ is decomposed into $\mathcal{W}_2$ and $\mathcal{V}_2$ in the same ways as before. Repeated this way, J's wavelet coefficient vectors $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_J$ and one scaling coefficient vector $\mathcal{V}_J$ are generated. Both $\mathcal{V}_j$ and $\mathcal{W}_j,\ j = 1, \ldots, J$ have the coefficients of $N/2^j$. Note that $\mathcal{W}_1$ is the vector of wavelet coefficients with the highest resolution(or scale) and $\mathcal{W}_J$ is the one with the lowest resolution(or scale). On the other hand, the subscripts of c_j and d_j preserve the order of resolution in the equation (3.4).

Each stage can be described in terms of filtering, that is, the elements of $\mathcal{V}_j$ are filtered through the filter coefficients $\{h_m\}$ and $\{g_m\}$ into $\mathcal{V}_{j+1}$ and $\mathcal{W}_{j+1}$. This

algorithm uses *downsampling(or decimation)* method on each stage which is denoted by h_{l-2k} and g_{l-2k} instead of h_{l-k} and g_{l-k} in equations (3.6). Downsampling is a mapping from $\ell(\mathbb{Z})$ to $\ell(2\mathbb{Z})$ which means that every other coefficients in $\mathcal{V}_j$ are used for filtering to generate $\mathcal{V}_{j+1}$ and $\mathcal{W}_{j+1}$. It retains only the values of even indices in sequences. Also it uses *circular convolution* for filtering of finite sequences. This circular convolution causes the coefficients affected by boundary condition.

3.5.1 Wavelet families and filters

There are several wavelet families such as Haar's, Daubechies's, Shannon's and so on. Two important families are as follows.

- **(Haar's wavelets)** This is the simplest wavelet family. It's scaling function is (3.1). This can be rewritten in the form of

$$\begin{aligned} \phi(x) &= \phi(2x) + \phi(2x-1) \\ &= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x-1). \end{aligned} \tag{3.8}$$

  From the two-scale equation, the scaling filters for the Haar's family are $h_0 = h_1 = \frac{1}{\sqrt{2}}$. Since Haar's wavelet function is

$$\begin{aligned} \psi(x) &= \phi(2x) - \phi(2x-1) \\ &= \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x) - \frac{1}{\sqrt{2}}\sqrt{2}\phi(2x-1), \end{aligned} \tag{3.9}$$

  the wavelet filters are $g_0 = -\frac{1}{\sqrt{2}}$ and $g_1 = \frac{1}{\sqrt{2}}$ from the wavelet equation or the equation (3.7). The Haar wavelets has compact support and are well localized in the time domain but discontinuous and so not effective in approximating smooth functions. The Haar wavelets satisfies three characteristics of wavelet basis functions which are compact support, orthogonality and symmetry.

- **(Daubechies' wavelets)** Daubechies (1992) constructed the hierarchy of the compactly supported and orthogonal wavelets with a desired degree of smoothness (regularity). This smoothness can be set by the number of vanishing moments which is defined that $\psi(x)$ has N (≥ 2) vanishing moments if $< \psi(x), x^n > = 0,\ \forall n = 0, 1, \ldots, N-1$. The simplest of Daubechies' wavelets is the Haar wavelet which has zero vanishing moments. It is the only discontinuous wavelet of Daubechies' ones. As the number of vanishing moments increases, the smoothness also does. There exists $2N$'s nonzero, real scaling filters when the number of vanishing moments is N. Two kinds of Daubechies' wavelets are the least asymmetric wavelets (symmlets) and the minimum phase wavelets.

3.5.2 Weakness of DWT

There exist some practical issues to be considered before using discrete wavelet transforms. Two major things are choice of width of the wavelet filter and handling boundary conditions.

- **(Width of the wavelet filters)** Width of wavelet filters is defined as the number of nonzero wavelet filters. For example, the width of Haar's wavelet filters is 2 and the one of Daubechies's least asymmetric wavelets with vanishing moments of 3 is 6. Large width of wavelet filters can be (1) increasing coefficients affected by boundary conditions and computation time, (2) decreasing the degree of localization of the coefficients. So it is desirable to use the smallest width giving reasonable results.

- **(Boundary condition)** This is resulted from the circular filtering of the algorithm. Suppose that the width of given wavelet filters is L. Then the number of coefficients affected by boundary condition in N_j dimensional vector $\mathcal{W}_j$ can be

calculated by $\min\{(L-2)(1-1/2^j)\}, N_j\}$. These are placed near the beginning or end of the coefficients in each resolution. It is wise to proceed in statistical inferences after eliminating those affected by boundary conditions.

3.6 Variances and covariances of wavelet coefficients

For the matrix notation of the DWT suppose that a data vector $\mathbf{X} = (X_1, X_2, \ldots, X_N$ where $N = 2^J$ is a realization of a random process $\{X_t : t = 1, 2, \ldots\}$ with the autocovariance function $\gamma(t) = \Sigma_X(i, j)$ where $|i - j| = t$. Also let's $\{W_n : n = 0, 1, \ldots, N - 1\}$ consist of the DWT coefficients of original data vector $\mathbf{X}$. Then we write $\mathbf{W} = \mathcal{W}\mathbf{X}$ where $\mathbf{W}$ is a column vector of length $N = 2^J$ whose nth element is the nth DWT coefficients W_n and $\mathcal{W}$ is a $N \times N$ orthogonal real-valued matrix defining the discrete wavelet transform. The covariance matrix of the random process $\mathbf{X}$ can be easily calculated from Σ_X as

$$\Sigma_W = \mathcal{W}\Sigma_X\mathcal{W}', \tag{3.10}$$

where Σ_W is the variance-covariance matrix of the wavelet coefficients $\mathbf{W}$ and *vice versa*. The covariance matrix, Σ_W of a wavelet transform has the form of Toeplitz matrices at different scales, while the covariance matrix is of Toeplitz form on time domain.

Vannucci and Corradi (1999) proposed an efficient algorithm to calculate Σ_W which uses the recursive filters in (3.6). This algorithm has a link to the two-dimensional discrete wavelet transform (DWT2). The DWT2 is first applied to Σ_X. Then the diagonal blocks of the resulting matrix give the variance-covariance matrices of wavelet coefficients that belong to the same scale which is called "within-scale" variance-covariance matrices. Next applying the one-dimensional DWT to the rows of the non diagonal blocks of the resulting matrix leads us to the "across-scale"

variance-covariance matrices which belong to different scales.

For example, Figure 2 shows the covariance matrix of a series with $n = 256$ from Gaussian ARFIMA$(0, 0.3, 0)$ process using the Matlab function *imagesc* that displays matrices as images. The highest gray-scale values of the images correspond to the largest entries of the matrices. In the plot we can see that almost all entries of the covariance matrix of ARFIMA$(0, 0.3, 0)$ process are away from zero, which means that the covariance matrix is dense. On the other hand, Figures 3 and 4 correspond to the covariance matrices of wavelet coefficients of the ARFIMA process with Daubechies' 6 and 8 vanishing moments, respectively. We know that there is essentially no correlation among coefficients at the same scale but some correlation between scales, as shown by the extra-diagonal gray lines, but it decreases when using wavelets with higher numbers of vanishing moments.

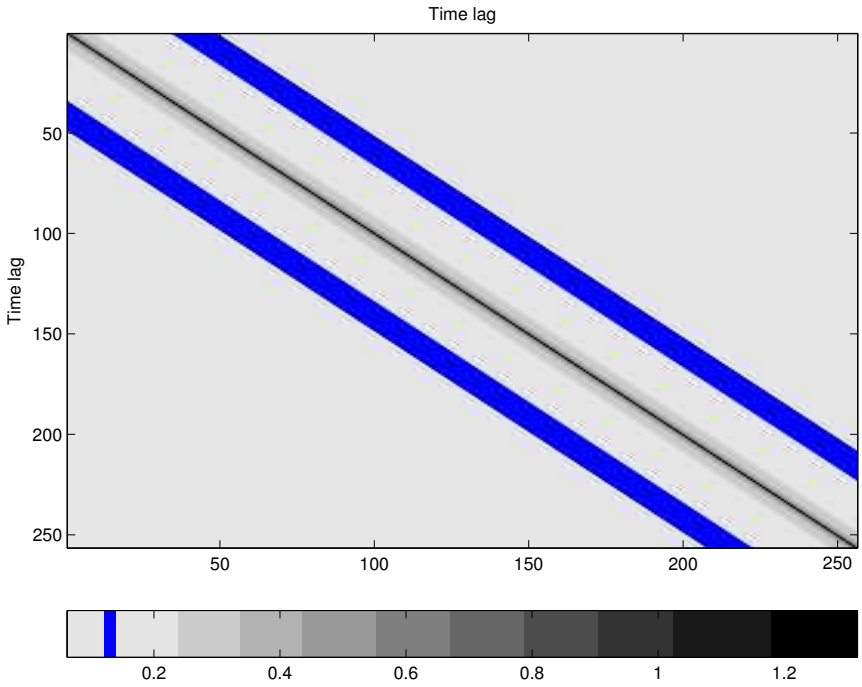

Figure 2: The covariance matrix of an ARFIMA$(0, 0.3, 0)$ process with $n = 256$

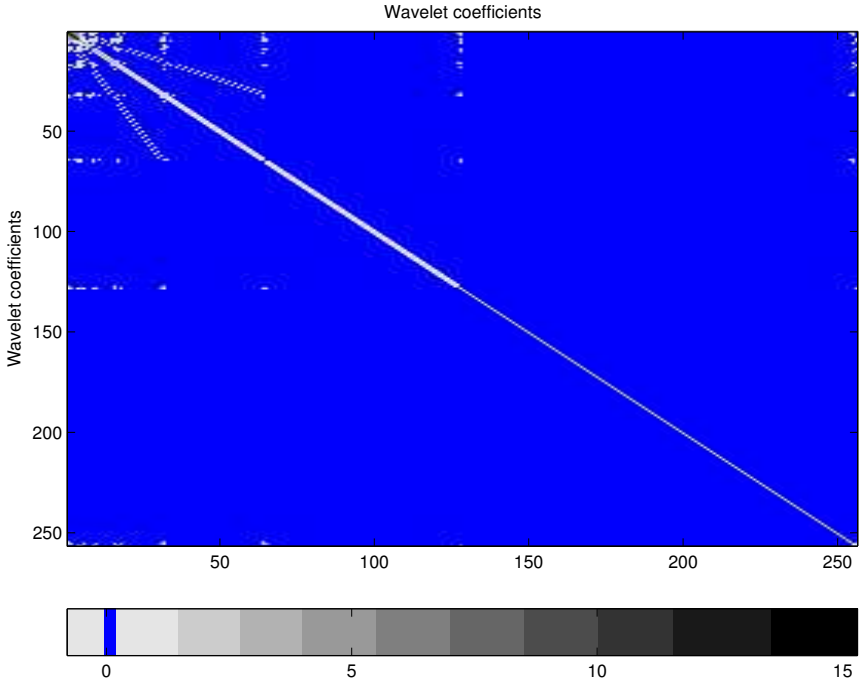

Figure 3: The covariance matrix of the wavelet coefficients of the series from Figure 2 using Vannucci-Corradi algorithm with Daubechies' 6 vanishing moments

3.7 DWT and long memory processes

In recent years discrete wavelet transforms of long memory processes have been popular because of their decorrelation properties and the simplification of the corresponding likelihood on wavelet domain. Tewfik and Kim (1992) showed that the correlations between discrete wavelet coefficients from the models of fractional Brownian motion (fBm) decrease across scales and time and thus induce sparse forms of covariance matrices compared to the dense ones on original data domain. Moreover they indicated that wavelets with a larger number of vanishing moments result in greater decorrelation of the wavelet coefficients though they suffer from the boundary condition. Wang (1996) consider the fractional Gaussian noise model,

$$Y(dx) = f(x)dx + \varepsilon^{2-2H} B_H(dx),$$

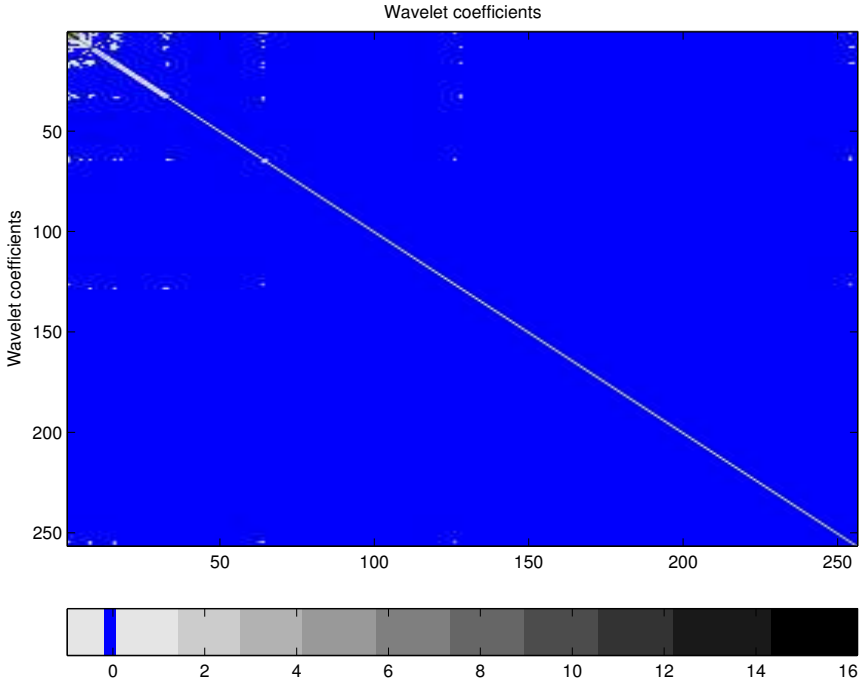

Figure 4: The covariance matrix of the wavelet coefficients of the series from Figure 2 using Vannucci-Corradi algorithm with Daubechies' 8 vanishing moments

where f is an unknown function, ϵ is the noise level which is small, and $B_H(dx)$ is a fractional Gaussian noise which is the derivative of a standard fractional Brownian motion as an approximation of the nonparametric regression model with long range dependence,

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $x_i = i/n \in [0, 1]$. Under the setting that empirical wavelet coefficients of the data Y are $y_\lambda = \int \Psi_\lambda(x) Y(dx)$ and wavelet estimates of f are defined as $\hat{f}(t_j) = \sum_\lambda \delta_{t_j}(y_\lambda)\Psi(\lambda)$, he established asymptotic results for minimax-wavelet threshold risk

$$R_W(\varepsilon; \mathcal{F}) = \inf_{t_j} \sup_{f \in \mathcal{F}} E\|f(\hat{t}_j) - f\|^2$$

and proposed wavelet shrinkage estimates with resolution level-dependent threshold tuned to achieve minimax rates.

McCoy and Walden (1996) showed the decorrelation properties of discrete wavelet transforms in the fractionally differenced Gaussian white noise processes (fdGn). They claim that there is no correlation within scales but there exist some correlation between scales. To get estimates of d and σ_ε^2, they set the off-diagonal elements of Σ_W in (3.10) to zero in order to get approximate covariance matrix, $\widetilde{\Sigma}_W = diag(S_{j_0+1}, S_{j_0}, S_{j_0-1}, \ldots, S_{j_0-1}, \ldots, S_1, \ldots, S_1)$ where $S_j = \text{var}\{d_{j,k}\}$ and $S_{j_0+1} = \text{var}\{c_{j_0,1}\}$ for $j = 1, \ldots, j_0; k = 1, \ldots, 2^{j_0-j}$ from the original one of the wavelet coefficients. Then, for zero mean Gaussian data, the wavelet coefficients and scaling coefficient will follow

$$\begin{aligned} d_{j,k} &\sim N[0, S_j(d, \sigma_\varepsilon^2)] \\ c_{j_0,1} &\sim N[0, S_{j_0+1}(d, \sigma_\varepsilon^2)], \end{aligned}$$

where $j = 1, \ldots, j_0; k = 1, \ldots, 2^{j_0-j}$. They form the approximate log-likelihood with constants ignored as

$$\begin{aligned} l(d, \sigma_\varepsilon^2) &= -N\log(\sigma_\varepsilon^2) - \log[S_{j_0+1}] - \sum_{j=1}^{j_0} \sum_{k=1}^{j_0-j} \log[S_j(d)] \\ &\quad -\frac{1}{\sigma_\varepsilon^2}\left[\frac{c_{j_0,1}^2}{S_{j_0+1}(d)} + \sum_{j=1}^{j_0} \sum_{k=1}^{j_0-j} \frac{d_{j,k}^2}{S_j(d)}\right]. \end{aligned} \tag{3.11}$$

Then the approximate MLE of $\hat{\sigma}_\varepsilon^2$ which depends on d is

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N}\left[\frac{c_{j_0,1}^2}{S_{j_0+1}(d)} + \sum_{j=1}^{j_0} \sum_{k=1}^{j_0-j} \frac{d_{j,k}^2}{S_j(d)}\right].$$

The concentrated log-likelihood (Haslett and Raftery, 1989) is formed by replacing σ_ε^2 of (3.11) by $\hat{\sigma}_\varepsilon^2$ and is numerically maximized with respect to d.

In the case of ARFIMA(p, d, q) models, the inference procedures are not easy due to the dense covariance matrices generated by the covariance function (2.11) and thus the handling of the likelihood with this kind of dense covariance matrix has

a problem of inversion. Since large size of data is needed to properly estimate the long memory parameter describing a long term persistent phenomenon of data over time, the inversion of the dense covariance matrix is crucial. Jensen (2000) showed the decorrelation properties of the DWT in ARFIMA(p, d, q) and gave an alternative(approximate) maximum likelihood estimates of the model parameters based on DWT in ARFIMA$(0, d, 0)$ and ARFIMA$(0, d, 1)$ models. He first showed that the normalized wavelet coefficients associated with a ARFIMA(p, d, q) process is self-similar for any time scale and also stationary time sequence and stationary scale sequence. In the paper, the approximate covariance matrix of wavelet coefficients is defined by $\Sigma_{<X,\Psi>} = E < X, \Psi >< X, \Psi >'$ using the approximate covariance function of the Definition 2.2.1. The approximate covariance matrix is a sparse matrix whose elements decay exponentially as one moves away from the diagonal elements of within and between scale covariance matrices. This decay creates finger-like bands. Also the banded approximate covariance matrix, $\Sigma^B_{<X,\Psi>}$ is defined as its finger-like diagonal elements are set to zero according to the condition B. For example, $\Sigma^1_{<X,\Psi>}$ consists of main diagonal elements of $\Sigma_{<X,\Psi>}$. Then approximate likelihood function is formed by

$$\begin{aligned} L_N(d| < X, \Psi >) &= (2\pi)^{-N/2}|\Sigma_{<x,\Psi>}(d)|^{-0.5} \\ &\quad \exp[-0.5 < x, \Psi >' \Sigma^{-1}_{<x,\Psi>}(d) < x, \Psi >]. \end{aligned}$$

The banded likelihood function L^B_N is equal to L_N except with $\Sigma_{<X,\Psi>}$ replaced by $\Sigma^B_{<X,\Psi>}$. Then $L_N(d| < X, \Psi >)$ and $L^B_N(d| < X, \Psi >)$ are numerically maximized.

Whitcher *et al.* (2000) applied iterated cumulative sums of squares (ICSS) algorithm to the detection and location of multiple variance change points in a stationary Gaussian fractionally differenced (FD) process via discrete wavelet transform and maximal overlap discrete wavelet transform (MODWT) which is so called "undec-

imated discrete wavelet transform". For detection of a variance change point, the normalized cumulative sums of squares test statistics $D \equiv \max(D^+, D^-)$ which has been investigated by Brown et al. (1975), Hsu (1977) and Inclán and Tiao (1994) is calculated for the wavelet coefficients through DWT, where

$$\mathcal{P}_k \equiv \frac{\sum_{j=0}^{k} X_j^2}{\sum_{j=0}^{N-1} X_j^2},$$

$$\mathcal{D}^+ \equiv \max_{0 \le k \le N-2} \left(\frac{k+1}{N-1} - \mathcal{P}_k \right),$$

and

$$\mathcal{D}^- \equiv \max_{0 \le k \le N-2} \left(\mathcal{P}_k - \frac{k+1}{N-1} \right),$$

under the null hypothesis $H_0 : \sigma_0^2 = \sigma_1^2 = \ldots = \sigma_{N-1}^2$. Since this testing procedure is originally designed for a sequence of independent Gaussian white noise with zero means and variances $\sigma_0^2, \sigma_1^2, \ldots, \sigma_{N-1}^2$, they transform a fractionally difference Gaussian white noise process, $\{X_t\}_{t=0}^{N-1}$ into nearly uncorrelated process through DWT device. After assuring a change point through the testing procedure, the location of it is searched using the undecimated wavelet coefficients through MODWT. Furthermore, based on this scheme to detect single change point, 'binary segmentation' procedure proposed by Vostrikova (1981) is used to keep searching multiple change points.

CHAPTER IV

BAYESIAN ESTIMATION OF ARFIMA MODELS ON WAVELET DOMAIN

4.1 Introduction

A Bayesian estimation procedure for the model parameters of Gaussian ARFIMA (p, d, q) processes is presented. The original data on time domain is transformed to wavelet coefficients on wavelet domain using discrete wavelet transform (DWT). Then Markov chain Monte Carlo (MCMC) is applied to those wavelet coefficients for Bayesian estimation. Vannucci and Corradi's (1999) algorithm is used to transform the variance-covariance matrix on time domain by parameter values proposed on each MCMC iteration to the one on wavelet domain. Simulations and application to the U.S. GNP data are presented.

4.2 Model in the wavelet domain

The decorrelation property of the wavelet transform is a great advantage when using data from a long memory process. Long memory data have a dense covariance structure that makes the exact likelihood of the data difficult to handle. A wavelet transform can represent the long-memory structure in a sparse form becoming a useful tool to simplify the likelihood. Let $\Psi = (\boldsymbol{\phi}, \boldsymbol{\theta}, d, \sigma^2)$ and $\Psi_0 = (\boldsymbol{\phi}, \boldsymbol{\theta}, d)$, where $\boldsymbol{\phi}$=$(\phi_1, \phi_2, \ldots, \phi_p)$ and $\boldsymbol{\theta}$=$(\theta_1, \theta_2, \ldots, \theta_q)$.

Suppose that $X = (x_1, x_2, \ldots, x_n)$ come from Gaussian ARFIMA(p, d, q) process. Then,

$$[X|\Psi] \sim N(0, \Sigma_X(\Psi)). \tag{4.1}$$

The model in the wavelet domain can be written as

$$[Z_i|\Psi] \sim N(0, \Sigma_{Z_i}(\Psi)) \tag{4.2}$$

independently for $i = 1, 2, \ldots, n$. Recall that the DWT is a linear orthogonal transformation and so wavelet coefficients inherit some of the features of the data, specifically zero mean Gaussian.

By using the Vannucci and Corradi (1999) algorithm, the variances $\Sigma_{Z_i}(\Psi)$ can be easily calculated from the autocovariance function (2.11) as previously described. Notice that the computations are considerably simple because only the variances are used to construct the approximate likelihood on wavelet domain. We simply apply the DWT2 to (2.11) and the diagonal elements of the resulting matrix, $\Sigma_{Z_i}(\Psi)$ are used as the variance $\sigma^2_{Z_i}(\Psi)$ of the corresponding wavelet coefficients Z_i. Given the form of the autocovariance function, we can write $\Sigma^2_{Z_i}(\Psi)$ in (4.2) as

$$\Sigma_{Z_i}(\Psi) = \sigma^2 \cdot \sigma^2_{Z_i}(\Psi_0), \tag{4.3}$$

where $\sigma^2_{Z_i}(\Psi_0)$ depend only on $(d, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)$. Thus (4.2) can be written as

$$[Z_i|\Psi] \sim N(0, \sigma^2 \cdot \sigma^2_{Z_i}(\Psi_0)), \tag{4.4}$$

where $\sigma^2 \cdot \sigma^2_{Z_i}$ is the ith diagonal element of Σ_Z.

4.3 Bayesian modeling on wavelet domain

4.3.1 Prior distributions

Priors for the unknowns, i.e. the long memory parameter, d, autoregressive coefficients, $\phi_1, \phi_2, \ldots, \phi_p$, moving average coefficients, $\theta_1, \theta_2, \ldots, \theta_q$, and the nuisance scale parameter, σ^2 are needed to be specified in order to do inference. For the prior specification, independent priors are assumed, which means that $\pi(\Psi) = \pi(\phi)\pi(\theta)\pi(d)\pi(\sigma)$.

As for the long memory parameter d, a sensible choice would be a mixture of two Beta distributions on -1/2< d <0 and 0< d <1/2 with a point mass at d=0. However, an uninformative prior, i.e. Uniform in (-1/2,1/2) is given for d, which is the range to satisfy the stationarity and inevitability of the model (2.7). For the priors of ϕ's and θ's, uniform priors defined on the ranges that satisfies the causality and invertibility of the ARMA process. Also, improper priors are used for μ on $\mathbb{R}^1$ and σ^2 on $\mathbb{R}^+$. Thus, $\pi(\Psi) \propto 1/\sigma^2$ over the parameter space and 0 elsewhere. Also the likelihood form of (4.4) is

$$\pi(\Psi|z_n) = (1/\sigma^2)^{n/2} \left(\frac{1}{\prod_{i=1}^n \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \exp\left[-\frac{\sum_{i=1}^n (z_i^2/\sigma_{z_i}^2(\Psi_0))}{2\sigma^2} \right] \tag{4.5}$$

4.3.2 Posterior distribution

The joint posterior distribution for Ψ may be written as

$$\pi(\Psi|z_n) \propto (1/\sigma^2)^{(n/2)+1} \left(\frac{1}{\prod_{i=1}^n \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \exp\left[-\frac{\sum_{i=1}^n (z_i^2/\sigma_{z_i}^2(\Psi_0))}{2\sigma^2} \right] \tag{4.6}$$

where σ^2 is the variance of white noise. Integrating out σ^2 in equation (4.6) leads to the marginal posterior distribution of Ψ_0,

$$\pi(\Psi_0|z_n) \propto \left(\frac{1}{\prod_{i=1}^n \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \left[\sum_{i=1}^n \left(\frac{z_i^2}{\sigma_{z_i}^2(\Psi_0)} \right) \right]^{-n/2}, \tag{4.7}$$

which is shown in Appendix A. In equation (4.6), since σ^{-2} follows a gamma distribution, the full conditional distribution of σ^2 does a inverse gamma distribution:

$$\pi(\sigma^2|\Psi_0, z_n) \sim IG\left(\frac{n}{2}, \frac{2}{\sum_{i=1}^n (z_i^2/\sigma_{z_i}^2(\Psi_0))} \right). \tag{4.8}$$

The marginal posterior distribution of σ^2, given data z_n is:

$$\pi(\sigma^2|z_n) = \int_{\Psi_0} \pi(\sigma^2|\Psi_0, z_n)\pi(\Psi_0|z_n)d\Psi_0. \tag{4.9}$$

Samples may be drawn from equation (4.7), using the Metropolis algorithm. Based on these samples $\Psi_0^j = (\phi_j^*, \theta_j^*, d_j), j = 1, \ldots, N$, the samples for σ^2 can be obtained from the Rao-Blackwellization steps shown above, i.e. for each Ψ_0^j and their corresponding samples $\sigma_j^2|\Psi_0^j, z_n$ from (4.8), we may sample $\sigma_j^2|z_n$ from equation (4.9).

There exists a technical problem in sampling $\Psi_0 = (\boldsymbol{\phi}, d, \boldsymbol{\theta})$ from the posterior distribution (4.7), using Gaussian proposal centered at the maximum likelihood estimates of the parameters and with covariance matrix given by the inverse of the observed Fisher's information matrix in the Metropolis algorithm. It comes from the restrictions on the ranges of $\Psi_0 = (\boldsymbol{\phi}, \boldsymbol{\theta}, d)$. The problems arise because if we use a Gaussian proposal for Metropolis algorithm, the ranges of the parameters in the distribution are from $-\infty$ to $+\infty$ for each parameter, d, θ and ϕ. A transformation method can be used to handle this. For example, in case of $p = 1$ and $q = 1$, that is $-1/2 < d < 1/2$, $-1 < \theta < 1$ and $-1 < \phi < 1$, the following transformations for each parameter are used.

$$d = \frac{-1 + e^{d^\dagger}}{2(1 + e^{d^\dagger})},\ \theta = \frac{-1 + e^{\theta^\dagger}}{1 + e^{\theta^\dagger}} \text{ and } \phi = \frac{-1 + e^{\phi^\dagger}}{1 + e^{\phi^\dagger}},$$

where $-\infty < \phi^\dagger, \theta^\dagger$ and $d^\dagger < \infty$ are from Gaussian proposal distribution. Then, the transformed posterior density for the model ARFIMA(p, d, q) where $p = 1$ and $q = 1$ is

$$\pi(\Psi_0|Z)|J| = \pi(\Psi_0^\dagger|Z),$$

where $\Psi_0^* = (\phi^\dagger, \theta^\dagger, d^\dagger)$ and

$$\begin{aligned} |J| &= \left|\frac{\partial \Psi_0}{\partial \Psi_0^\dagger}\right| \\ &= \frac{\exp(d^\dagger)\exp(\phi_1^\dagger)\exp(\theta_1^\dagger)}{g(d^*, \phi_1^\dagger, \theta_1^\dagger)}, \end{aligned}$$

where $g(d^\dagger, \phi_1^\dagger, \theta_1^\dagger) = (1 + \exp(d^\dagger))^2(1 + \exp(\phi_1^\dagger))^2(1 + \exp(\theta_1^\dagger))^2$.

Table 1: ARFIMA$(1, d, 0)$: Estimates of d and ϕ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors.

(ϕ, d)		ϕ=.5	d=.2	ϕ=-.8	d=.4
2^7	KV	.5474(.0702)	.1502(.0274)	-.6669(.0601)	.3801(.0281)
	MLE	.6466	.0015	-.6373	.3290
	GPH		.0550		.4993
2^9	KV	.5599(.0458)	.1585(.0247)	-.7245(.0332)	.3729(.0256)
	MLE	.6015	.1252	-.7198	.3671
	GPH		.2529		.5018

4.4 Simulation study

There are a number of ways to generate a time series that exhibits long-memory properties. A computationally simple one was proposed by McLeod and Hipel (1978) and involves the Cholesky decomposition of the correlation matrix $R_{X(i,j)} = [\rho(|i-j|)]$. Given $R_X = MM'$ with $M = [m_{i,j}]$ a lower triangular matrix, if ϵ_t, $t = 1, \ldots, n$ is a Gaussian white noise series with zero mean and unit variance, then the series

$$X_t = \gamma_0^{1/2} \sum_{i=1}^{t} m_{t,i}\epsilon$$

will have the autocorrelation $\rho(\tau)$. The data for Simulation are here generated by using the McLeod and Hipel's method with $\sigma^2 = 1$. Simulations to estimate the autoregressive, moving average parameters and the long memory parameter are performed for ARFIMA$(1, d, 0)$, ARFIMA$(0, d, 1)$ and ARFIMA$(1, d, 1)$ models. For checking robustness of the estimates the simulations are carried out according to different values of parameters ($\phi = .5, -.8$, $\theta = .5, -.8$ and $d = .2, .4$) and of the sample size $n = 2^7, 2^9$. The wavelet family which are used in all simulations is Daubechies (1992) minimum phase wavelets MP(7).

The similar settings to those of Pai and Ravishanker (1996) are used. For each combination of the parameters the estimates based on ten parallel MCMC chains

Table 2: ARFIMA$(0, d, 1)$: Estimates of d and θ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors.

(d, θ)		d=.2	θ=.5	d=.4	θ=-.8
2^7	KV	.2059(.0659)	.3484(.2051)	.4104(.0170)	-.6968(.0816)
	MLE	.0699	.4671	.1367	-.4029
	GPH	.4237		.2913	
2^9	KV	.1948(.0253)	.4029(.0797)	.3192(.0147)	-.7155(.0313)
	MLE	.1559	.5547	.3024	-.7189
	GPH	.2316		.3580	

are reported. The maximum likelihood estimates are used as initial values for each MCMC sampler and then perturbed to obtain over-dispersed values to be used to initialize the ten independent parallel MCMC chains. All chains were run for 10,000 iterations after a burn-in time of 10,000. Estimates were computed as posterior means together with posterior standard errors, obtained from the pooled MCMC sample. Tables 1, 2 and 3 summarize the numerical results. For comparison the values of the MLE and the classical estimator of Geweke and Porter-Hudak (1983) based on a regression on periodogram ordinate are reported together.

In the result of ARFIMA$(1, d, 0)$ process wavelet-based Bayes estimates which are here called "KV" are always better than MLE and GPH estimates for both small and large sample sizes. MLE as expected improve considerably large sample sizes. For ARFIMA$(0, d, 1)$ model wavelet-based Bayes estimates are almost better except the case of $\theta = .5$, for which case KV estimates do not improve on MLE. In case of ARFIMA$(1, d, 1)$ process wavelet-based Bayes estimates for d improve on the other methods, while MLE do a better job at estimating ϕ, although they are very close to the MLE values. These results seem to indicate a better performance of the proposed method based on wavelet domain with respect to the other two methods.

Table 3: ARFIMA$(1, d, 1)$: Estimates of d, ϕ and θ from wavelet-based Bayesian method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively. Numbers in parentheses are standard errors.

(ϕ, d, θ)		ϕ=.1 d=.4 θ=.5			ϕ=-.1 d=.4 θ=-.5		
2^7	KV	.1703(.1579)	.2617(.0042)	.6650(.2324)	-.2568(.1709)	.3755(.0226)	-.2995(.1558)
	MLE	.0847	.2588	.7020	-.2147	.2447	-.2344
	GPH		.4777			.3145	
2^9	KV	.0459(.1346)	.3915(.0075)	.5577(.2136)	-.0529(.0982)	.2834(.0261)	-.4183(.0921)
	MLE	.0698	.3859	.5646	-.0386	.2783	-.4374
	GPH		.5134			.1962	

4.5 Example: U.S. GNP data

We examined post-war quarterly U.S. GNP data from 1947(1) to 1991(1) with $n = 177$. The original series shows strong trend and that the variability increases along time. The logarithms of the original data and their first differences are taken to treat the variability and trend, respectively. That is, when y_t is the original U.S. GNP series on given period, we use $\ln(y_t) - \ln(y_{t-1})$ as the analysis data. The (a) in Figure 5 shows the first differences of the logarithms of the original series.

This series has been analyzed by several authors. A slightly shorter time series was fitted by Sowell (1992b) with both ARMA and ARFIMA models. The AIC criterion indicated the ARFIMA$(3, d, 2)$ as the best model. Sowell also reported MLE estimates for the long memory and the autoregressive and moving average parameters of all models. As for Bayesian approaches, Pai and Ravishanker (1996) showed evidence for the ARFIMA$(0, d, 0)$ model without mean as the best fit, while Koop *et al.* (1997) reported ARFIMA$(1, d, 0)$ as best model. For some models there appears to be some discrepancy between the parameter estimates they report and those of Sowell.

We fitted ARFIMA(p, d, q) models with $p, q = 0, 1$. We used a circulant filter by padding the series with replicas and truncating the wavelet transform. Estimates and standard deviations for the parameters of the different models are reported in Table 4.

Table 4: Wavelet-based Bayes estimates for U.S. GNP data

Model	d	ϕ	θ
(1,d,0)	-.4499(.0027)	-.6910(.0376)	
(0,d,1)	.1925(.0303)		.1705(.0521)
(1,d,1)	-.3628(.0124)	-.6991(.0203)	.0456(.0361)

In the case of ARFIMA$(1, d, 0)$, ARFIMA$(0, d, 1)$ and ARFIMA$(1, d, 1)$, estimates are based on ten independent parallel MCMC chains with 1,000 iterations each and burn-in times of 1,000 iterations. With respect to previous works, these estimates appear to be in better agreement with those reported by Sowell (1992b). For example, Koop *et al.* (1997) reported the estimates of the long memory parameter $d = -.29, .23, -.16$, while MLE estimates were $d = -.45, .16, -.38$, for ARFIMA$(1, d, 0)$, ARFIMA$(0, d, 1)$ and ARFIMA$(1, d, 1)$, respectively. The (b), (c) and (d) in Figure 5 show kernel density estimates of the posterior distributions of the long memory parameter for the three different ARFIMA models, ARFIMA$(1, d, 0)$, ARFIMA$(0, d, 1)$ and ARFIMA$(1, d, 1)$.

4.6 Supplementary study on using diagonal elements of Σ_W

Although Tewfik and kim (1992), McCoy and Walden (1996) and Jensen (2000) showed a decorrelation property of discrete wavelet transform on long memory series, we here report the empirical evidence about the property on ARFIMA(p, d, q) process through a simulation study with ARFIMA$(0, d, 0)$ process.

Tables 5 and 6 show numerical results for the estimation of the long memory parameter, d of ARFIMA$(0, d, 0)$ models. The simulated data are generated with $d = 0.05, 0.20, 0.45$ and $\sigma^2 = 0.05$. For checking the robustness of the results, we use different sample sizes $n = 128$ (Table 5), $n = 512$ (Table 6) and different vanishing moments 2 and 7 with Daubechies's minimum phase wavelets.

In Tables 5 and 6, there are no remarkable differences between the estimated values of the long memory parameter in the aspect of precision when we used the full matrix and the diagonal elements of the covariance matrix Σ_W of the wavelet coefficients from ARFIMA$(0, d, 0)$ models. In some cases, the biases when diagonal terms are used is less than the case when th full matrix is used. When $n = 512$, high

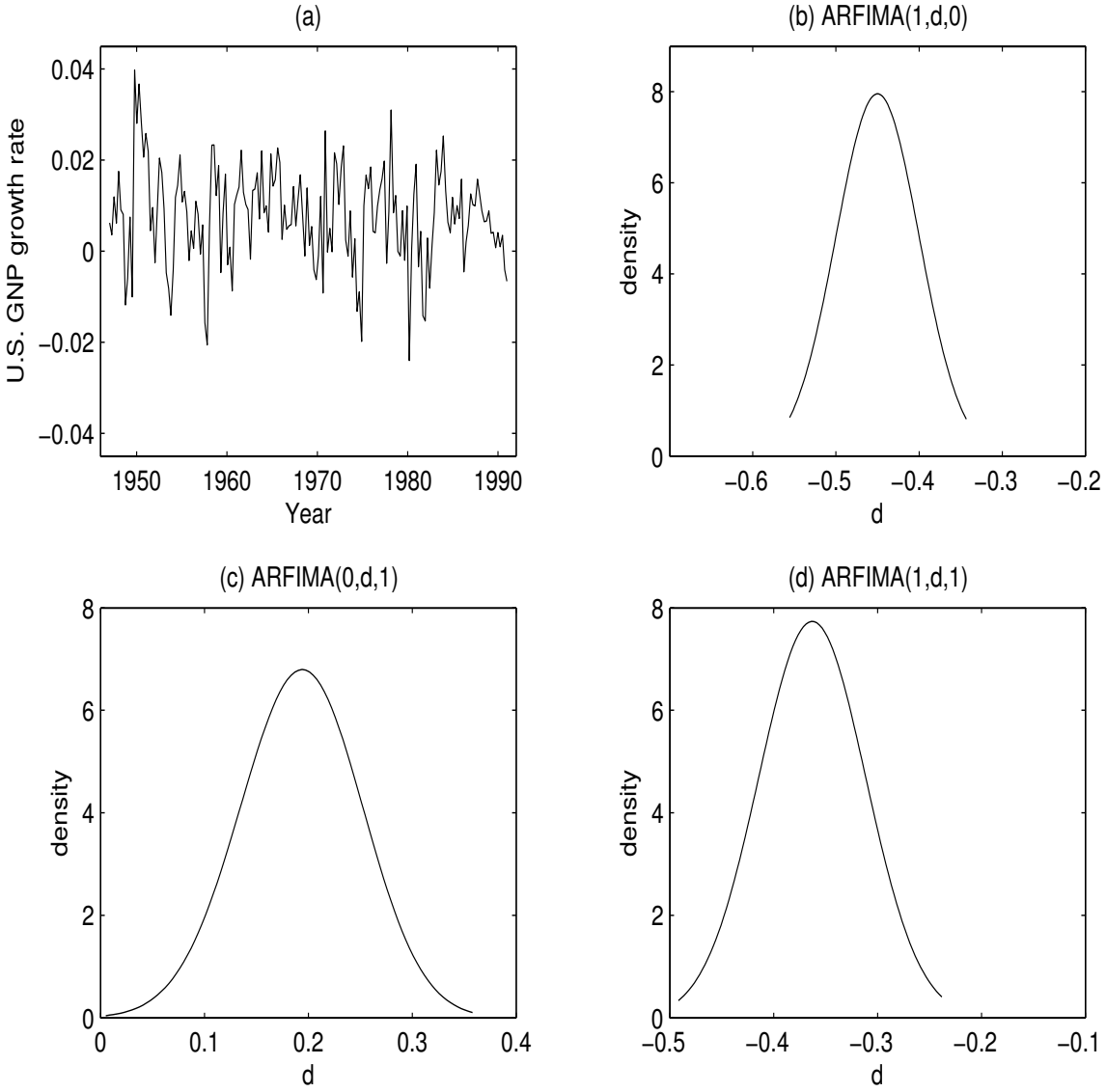

Figure 5: First differences of GNP data, in plot a), and kernel estimates of the posterior density of d for b) ARFIMA$(1, d, 0)$, c) ARFIMA$(0, d, 1)$ and d) ARFIMA$(1, d, 1)$

number of vanishing moments give less biases than low number of vanishing moments.

From this empirical study with theoretical works of Jensen (2000), we have an evidence to assume uncorrelated wavelet coefficients and use only diagonal terms, $\sigma^2 \cdot \sigma^2_{Z_i}(\Psi_0)$ of the covariance matrix Σ_W in (4.3) to evaluate the likelihood function the estimation procedure in (4.5) .

Table 5: Simulation result when diagonal elements and full matrix of Σ_W used in the estimation of long memory parameter of ARFIMA$(0, d, 0)$ model with sample size $n = 128$. The numbers in parenthesis denote biases.

	Diagonal Matrix		Full Matrix	
No. of Vanishing moments	2	7	2	7
$d = 0.05$	0.0605(0.0105)	0.0411(-0.0089)	0.0478(-0.0022)	0.0460(-0.0040)
$\sigma^2 = 0.90$	0.6859(-0.2141)	0.6831(-0.2169)	0.6824(-0.2176)	0.6839(-0.2161)
$d = 0.20$	0.1379(-0.0621)	0.1971(-0.0029)	0.1209(-0.0791)	0.1532(-0.0468)
$\sigma^2 = 0.90$	0.7500(-0.1500)	0.7583(-0.1417)	0.7462(-0.1538)	0.7611(-0.1389)
$d = 0.45$	0.4557(0.0057)	0.4410(-0.0090)	0.4276(-0.0224)	0.4282(-0.0218)
$\sigma^2 = 0.90$	1.2132(0.3132)	1.2333(0.3333)	1.2258(0.3258)	1.2208(0.3208)

Table 6: Simulation result when diagonal elements and full matrix of Σ_W used in the estimation of long memory parameter of ARFIMA$(0, d, 0)$ model with sample size $n = 512$. The numbers in parenthesis denote biases.

	Diagonal Matrix		Full Matrix	
No. of Vanishing moments	2	7	2	7
$d = 0.05$	0.0496(-0.0004)	0.0487(-0.0013)	0.0512(0.0012)	0.0542(0.0042)
$\sigma^2 = 0.90$	0.8075(-0.0925)	0.8110(-0.089)	0.8072(-0.0928)	0.8091(-0.0909)
$d = 0.20$	0.2529(0.0529)	0.2286(0.0286)	0.1895(-0.0105)	0.2113 (0.0113)
$\sigma^2 = 0.90$	0.8548(-0.0452)	0.8900(-0.0100)	0.8612(-0.0388)	0.8673 (-0.0327)
$d = 0.45$	0.4783(0.0283)	0.4659(0.0159)	0.4408(-0.0092)	0.4451(-0.0049)
$\sigma^2 = 0.90$	0.7763(-0.1237)	0.7952(-0.1048)	0.7985(-0.1015)	0.8010(-0.0990)

CHAPTER V

CHANGE POINT ANALYSIS ON WAVELET DOMAIN

5.1 Introduction

The number of change points is assumed unknown. Instead we specify the maximum number of change points *a priori*. Original data is transformed to wavelet coefficients through discrete wavelet transform as in Chapter IV. Reversible jump Markov chain Monte Carlo is applied to the wavelet coefficients to estimate the number of change points, locate the corresponding change points and estimate the other model parameters of the Gaussian ARFIMA(p, d, q) process.

5.2 Model and likelihood

In this section, we consider the case that multiple changes exists in the long memory parameter. The model is as follows.

$$\Phi(B)(1-B)^{d_t}X_t = \Theta(B)\varepsilon_t,$$

where

$$d_t = \begin{cases} d_1, & \text{if } 1 \le t < c_1, \\ d_2, & \text{if } c_1 \le t < c_2, \\ \vdots & \vdots \\ d_{k+1}, & \text{if } c_k \le t \le n. \end{cases}$$

Integer-valued k satisfies $0 \le k \le k_{max}$ where k_{max} is pre-specified. This means that we set up the maximum number of changes for long memory parameter d_t a priori. $\boldsymbol{C_k} = (c_1, \ldots, c_k)$ is a vector of the positions that changes in long memory parameter occur. $\boldsymbol{d_k} = (d_1, d_2 \ldots, d_{k+1})$ is a vector of long memory parameters cor-

responding to each locations such that $d_i \in (0, .5)$. Again, let $\Psi_0 = (\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{d}_k)$, where $\boldsymbol{\phi}$=$(\phi_1, \phi_2, \ldots, \phi_p)$ and $\boldsymbol{\theta}$=$(\theta_1, \theta_2, \ldots, \theta_q)$.

Here we assume that the original ARFIMA(p, d, q) model has constant mean μ. We model wavelet coefficients rather than the original data. The DWT is a linear and orthogonal transformation and wavelet coefficients therefore inherit the distribution of the data, specifically they are zero mean Gaussian. After taking DWT, we get wavelet coefficients $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$ with mean zero. The likelihood function is

$$f(\mathbf{z}|\boldsymbol{\phi}, \mathbf{d}_k, \boldsymbol{\theta}, \sigma^2) = (\sqrt{2\pi})^{-n} \prod_{i=1} \left(\frac{1}{\sigma^2 \sigma^2_{z_i}(\Psi_0)} \right)^{1/2} \exp \left[-\frac{1}{2} \sum_{i=1}^{n} \frac{z_i^2}{\sigma^2 \sigma^2_{z_i}(\Psi_0)} \right].$$

5.2.1 Prior Distribution

We specify priors for the unknowns, i.e. the long memory parameter, d_t, autoregressive coefficients, $\boldsymbol{\phi}$, moving average coefficients, $\boldsymbol{\theta}$, the number of change points k and the variance of white noise σ^2. Let $\xi_k = (k, \omega_k) = (k, \boldsymbol{C_k}, \boldsymbol{d_k}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$, where $\omega_k = (\boldsymbol{C_k}, \boldsymbol{d_k}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$. For the prior specification we assume

$$\begin{aligned} \pi(k, \omega_k) &= \pi(k)\pi(\omega_k|k) \\ &= \pi(k)\pi(\boldsymbol{C_k}|k)\pi(\boldsymbol{d_k}|k, \mathbf{C_k})\pi(\boldsymbol{\phi})\pi(\boldsymbol{\theta})\pi(\sigma^2) \end{aligned}$$

For the long memory parameter $\boldsymbol{d_k}$ where $k = 0, 1, 2, 3, \ldots, k_{max}$, we use an uninformative prior, i.e. uniform in $(0, 0.5)$ which is the range of the model to have long memory. For the priors of ϕ's and θ's, we give uniform priors with the ranges to satisfy the causality and invertibility of the ARMA process. As prior distributions of σ^2 we use an improper prior with $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$. For k, we set

$$k \sim U\{0, 1, 2, \ldots, k_{max}\}.$$

For a prior distribution of $\boldsymbol{C_k}$, suppose that k^* is generated on an iteration from discrete uniform distribution with a range $\{0, 1, 2, \ldots, k_{max}\}$. Then the possible ways we

can pick k^*'s change points out of $n-2$ candidates are $\binom{n-2}{k^*}$. Using this information, a reasonable choice for the prior distribution of $\mathbf{C}_k$ will be

$$\boldsymbol{C_k}|k \sim \binom{n-2}{k}^{-1}. \tag{5.1}$$

Also one can use an ordered prior distribution as the prior distribution for $\boldsymbol{C_k}$. It is

$$\boldsymbol{C_k}|k \sim \frac{k!}{(n-2)^k}. \tag{5.2}$$

When we use the unordered prior distribution (5.1) the calculations of the acceptance rates in birth step and death step for reversible jump MCMC in Section 5.3 become simpler.

5.2.2 Posterior and full conditional distributions

The joint posterior distribution of the model parameters including $\boldsymbol{C_k}$ and k which is derived in Appendix B.1 is

$$\begin{aligned} f(k, \boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}, \boldsymbol{C_k}, \sigma^2|\boldsymbol{z}) \;\; \propto \;\; & \left(\frac{1}{\prod_{i=1}^n \sigma^2 \sigma^2_{z_i}(\Psi_0)}\right)^{1/2} \exp\left[-\frac{1}{2}\sum_{i=1}^n \frac{z_i^2}{\sigma^2 \sigma^2_{z_i}(\Psi_0)}\right] \\ & \times \binom{n-2}{k}^{-1} \sigma^{-2}. \end{aligned} \tag{5.3}$$

We treat σ^2 as a nuisance parameter and integrate it out to get a proper marginal posterior distribution of $\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}$ and $\boldsymbol{C_k}$ as follows:

$$\begin{aligned} f(k, \boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}, \boldsymbol{C_k}|\boldsymbol{z}) \;\; \propto \;\; & \left(\frac{1}{\prod_{i=1}^n \sigma^2_{z_i}(\Psi_0)}\right)^{1/2} \left[\sum_{i=1}^n \frac{z_i^2}{\sigma^2_{z_i}(\Psi_0)}\right]^{-n/2} \\ & \times \binom{n-2}{k}^{-1}, \end{aligned}$$

which is a proper distribution because all the prior distributions except σ^2 and likelihood function are proper.

In order to implement our MCMC estimation procedure we need full conditional distributions of a change point and model parameters $\Psi_0 = (\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta})$. From the joint

posterior distribution, we get the conditional posterior distributions for c_m as

$$\begin{aligned}
&P(c_m|\boldsymbol{z}, k, c_1, c_2, \ldots, c_{m-1}, c_{m+1}, \ldots, c_k, \Psi) \\
&\propto \left(\frac{1}{\prod_{i=c_{m-1}+1}^{c_m-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_{m-1}, \boldsymbol{\theta})} \right)^{1/2} \left(\frac{1}{\prod_{i=c_m}^{c_{m+1}-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_m, \boldsymbol{\theta})} \right)^{1/2} \\
&\times \left[\sum_{i=c_{m-1}+1}^{c_m-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_{m-1}, \boldsymbol{\theta})} \right]^{-(c_m-c_{m-1})/2} \\
&\times \left[\sum_{i=c_m}^{c_{m+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_m, \boldsymbol{\theta})} \right]^{-(c_{m+1}-c_{m-1})/2}, \qquad (5.4)
\end{aligned}$$

where $c_m = c_{m-1}+1, c_{m-1}+2, \ldots, c_{m+1}-1$, $c_0 = 1$ and $m = 1, 2, \ldots, k$.

Also, the full conditional distribution of $\Psi_0 = (\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta})$ is

$$\begin{aligned}
&P(\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}|\boldsymbol{z}, k, \mathbf{C}_k) \\
&\propto \left(\frac{1}{\prod_{i=1}^{c_1-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_1, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=1}^{c_1-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_1, \boldsymbol{\theta})} \right]^{-(c_1-1)/2} \\
&\times \left(\frac{1}{\prod_{i=c_1}^{c_2-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_2, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_1}^{c_2-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_2, \boldsymbol{\theta})} \right]^{-(c_2-c_1)/2} \\
&\qquad \vdots \qquad\qquad\qquad \vdots \\
&\times \left(\frac{1}{\prod_{i=c_k}^{n} \sigma_{z_i}^2(\boldsymbol{\phi}, d_{k+1}, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_k}^{n} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_{k+1}, \boldsymbol{\theta})} \right]^{-(n-c_k+1)/2} \qquad (5.5)
\end{aligned}$$

5.3 Application of reversible jump MCMC

We adopt the reversible jump Markov chain Monte Carlo method introduced by Green (1995) to make inferences about ARFIMA(p, d, q) model with an unknown number of changes but known number of maximum changes, k_{max}. Our methodology using a reversible jump MCMC is similar to Liu and Kao's (1999) one except:

- We deal with general types of ARFIMA(p, d, q) models on wavelet domain while they had inferences of ARFIMA$(1, d, 0)$-ARCH(1) models on original data domain.

- we use metropolis algorithm to sample parameter values in (2) while they used griddy Gibbs sampler.

Now we define three types of moving-patterns with (i) Birth step of a change-point, (ii) Death step of a change-point and (iii) Transition step of model parameters $\omega_k{}' = (\boldsymbol{C_k}, \boldsymbol{d_k}, \boldsymbol{\phi}, \boldsymbol{\theta})$, which handle the varying dimension of parameter space along on each iteration. A reversible jump MCMC algorithm randomly evolves with one of the above steps at each iteration with the following probabilities, respectively,

- $P(\text{birth of a change-point}) = \alpha_k$,
- $P(\text{death of a change-point}) = \beta_k$ and
- $P(\text{transition of } \omega_k{}') = \gamma_k$,

where $\alpha_k + \beta_k + \gamma_k = 1$. Green (1995) proposed that these probabilities could be chosen in order to ensure $\alpha_k \pi(k) = \beta_{k+1}\pi(k+1)$ which guarantees certain acceptance in the corresponding Hastings sampler only for the number of steps so that

$$\alpha_k = c \cdot \min\{1, \pi(k+1)/\pi(k)\}, \quad \beta_{k+1} = c \cdot \min\{1, \pi(k)/\pi(k+1)\},$$

where the constant c is as large as possible subject to $\alpha_k + \beta_k \leq 0.9$ for all $k = 0, 1, \ldots, k_{\max}$. Since $\pi(k) = \pi(k+1) = 1/(k_{\max}+1)$ in our setting of prior distribution for k, $\alpha_k = \beta_{k+1} = c \leq 0.9$ and so α_k and β_{k+1} can be set to 0.45. Therefore, we can set up each probability as

$$\begin{cases} \alpha_k = 0.0, & \beta_k = 0.9, & \text{and} & \gamma_k = 0.1 & \text{if } k = k_{max}, \\ \alpha_k = 0.9, & \beta_k = 0.0, & \text{and} & \gamma_k = 0.1 & \text{if } k = 0, \\ \alpha_k = 0.45, & \beta_k = 0.45 & \text{and} & \gamma_k = 0.1 & \text{if } k = 1, 2, 3, \ldots, k_{max} - 1. \end{cases}$$

5.3.1 Birth step of a change-point

The birth step of a change-point add a new change-point to the current state of parameters, $\zeta_k = (k, \boldsymbol{C_k}, \boldsymbol{d_k}, \boldsymbol{\omega_k}')$. That is, the number of change-points get changed from k to $k+1$. Two uniform random variables, u_1 and u_2 are generated to match the dimension of parameters such that $u_1 \sim U\{2, 3, \ldots, n-1\} \cap \{c_1, c_2, \ldots, c_k\}^c$, where $\{c_1, \ldots, c_k\}$ is a set of existing change points and $u_2 \sim U(-R_{u_1}, R_{u_1})$, where $R_{u_1} = \min\{d_{m^*-1},\ 0.5 - d_{m^*-1}\}$ and $m^* = \min\{i | u_1 < c_i\}$.

Let the suggested locations of $(k+1)$ change points be $\boldsymbol{C_{k+1}}' = (c_1', c_2', \ldots, c_{k+1}')$ and the corresponding long memory parameters $\boldsymbol{d_{k+2}}' = (d_1', d_2', \ldots, d_{k+2}')$, where

$$c_i' = \begin{cases} c_i, & \text{if } 1 \le i \le m^* - 1, \\ u_1, & \text{if } i = m^*, \\ c_{i-1}, & \text{if } m^* + 1 \le i \le k+1, \end{cases}$$

and

$$d_i' = \begin{cases} d_i, & \text{if } 1 \le i \le m^* - 1, \\ d_{m^*} - u_2, & \text{if } i = m^*, \\ d_{m^*} + u_2, & \text{if } i = m^* + 1, \\ d_{i-1}, & \text{if } m^* + 2 \le i \le k+2. \end{cases}$$

Then the proposed parameter space becomes $\zeta_{k+1}' = (k+1, \boldsymbol{C_{k+1}}', \boldsymbol{d_{k+1}}', \boldsymbol{\omega_k}')$. Note that $\boldsymbol{\omega_k}'$ is not changed. We accept a proposed transition from ζ_{k+1} to ζ_{k+1}' with probability

$$\begin{aligned} &\alpha_{\text{birth}}(\zeta_k, \zeta_{k+1}') \\ &\quad = \min(1, \textit{posterior ratio} \times \textit{proposal ratio} \times |J|) \\ &\quad = \min\left(1, \frac{Q(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})}{Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \times \frac{4R_{u_1}\beta_{k+1}}{\alpha_k}\right), \end{aligned} \tag{5.6}$$

where

$$Q(\boldsymbol{\phi}, d_i', \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_i'}^{c_{i+2}'-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i'}^{c_{i+2}'-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right]^{-(c_{i+2}'-c_i')/2}$$

and

$$Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_i}^{c_{i+1}-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i}^{c_{i+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right]^{-(c_{i+1}-c_i)/2},$$

which is derived in Appendix B.2.

5.3.2 Death step of a change-point

The death step of a change-point drops one among the existing change-points. That is, the number of change points is changed from k to $k-1$. A uniform random variable u is generated to match the dimension of parameters with a range over $\{c_1, c_2, \ldots, c_k\}$. Let the proposed locations with $(k-1)$ change points be $\boldsymbol{C_{k-1}}^* = (c_1^*, c_2^*, \ldots, c_{k-1}^*)$ and the corresponding long memory parameters $\boldsymbol{d_{k-1}}^* = (d_1^*, d_2^*, \ldots, d_k^*)$, where

$$c_i^* = \begin{cases} c_i, & \text{if } 1 \le i \le u-1, \\ c_{i+1}, & \text{if } u \le i \le k-1, \end{cases}$$

and

$$d_i^* = \begin{cases} d_i, & \text{if } 1 \le i \le u-1, \\ (d_u + d_{u+1})/2, & \text{if } i = u, \\ d_{i+1}, & \text{if } u+1 \le i \le k. \end{cases}$$

The resulting parameter space is $\zeta_{k-1}^* = (k-1, \boldsymbol{C_{k-1}}^*, \boldsymbol{d_{k-1}}^*, \boldsymbol{\omega_k}')$. Note again that $\boldsymbol{\omega_k}'$ still remains same. We accept a transition from ζ_k to ζ_{k-1}^* with probability

$$\begin{aligned} \alpha_{\text{death}}(\zeta_k, \zeta_{k-1}^*) &= \min(1, \textit{posterior ratio} \times \textit{proposal ratio} \times |J|) \\ &= \min\left(1, \frac{Q(\boldsymbol{\phi}, d_i^*, \boldsymbol{\theta})}{Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \times \frac{\alpha_{k-1}}{4R_u\beta_k}\right), \end{aligned} \tag{5.7}$$

where

$$Q(\boldsymbol{\phi}, d_i^*, \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_{i-1}^*}^{c_i^*-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i^*, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_{i-1}^*}^{c_i^*-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i^*, \boldsymbol{\theta})} \right]^{-(c_i^*-c_{i-1}^*)/2}$$

and

$$Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_{i-1}}^{c_{i+1}-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_{i-1}}^{c_{i+1}-1} \frac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right]^{-(c_{i+1}-c_{i-1})/2},$$

which is derived in Appendix B.3.

5.3.3 Transition step of model parameters, $\omega_{\mathbf{k}}{}'$

We update the model parameters by sampling $\boldsymbol{C_k}$ from the full conditional distribution (5.4) and $(\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta})$ from the full conditional distribution (5.5). For the latter we use a Metropolis algorithm with a multivariate normal distribution as proposal distribution, centered at the maximum likelihood estimates of the parameters and with covariance matrix given by observed Fisher information.

Unlike the estimation procedure of the model parameters of ARFIMA(p, d, q) model in Chapter IV, we cannot use the transformation technique in this Metropolis step, which transforms the proposal values from multivariate normal distribution into the values over the restricted ranges which are needed for the invertibility and stationary conditions of AR and MA parameters and the range of long memory parameter d_t. The reason is that d_t in birth and death step of reversible jump algorithm are updated from one-to-one function using uniform distributions which have the ranges satisfying those restrictions and so the acceptance rates in birth and death steps do not use the same transformation technique as in Metropolis step of Chapter IV. For this reason, we use rejection sampling method instead of transformation technique in the Metropolis step. This method is computationally expensive but, more reasonable than transformation technique in our situation.

5.4 Simulation studies

We carried out simulation studies with data sets of sample sizes $n = 512$ which are generated by McLeod and Hipel's (1978) method. For the single change point case, we simulated datasets with a change at $t = 256$. For multiple changes, we induced changes at $t = 128$ and 256. We simulated data from ARFIMA$(1, 1, 1)$, ARFIMA$(1, 1, 0)$ and ARFIMA$(0, 1, 1)$ models with one change point and from ARFIMA$(1, 1, 1)$ with two change points. In all analyses we used discrete wavelet transforms with Daubechies' minimum phase wavelets with seven vanishing moments, Daubechies (1992). For the detection of the change points, the wavelet coefficients of the highest level are used and those affected by the boundary conditions are discarded.

The initial value for the number of change points, k, is set to zero and the maximum number of change points, k_{max} is allowed to 3. This is a reasonable choice because we usually estimate the model parameters with the assumption that there do not exist change points in a given data. For MCMC sampling, we use parametric optimal values as the initial values of ϕ, θ and d and inverse of Fisher's information matrix as the variance-covariance matrices of multivariate normal proposal distribution for the Metropolis step. When sampling c_m from (5.4), the c's changed during the reversible jump step is re-calculated and standardized so to sum up to one. Then only one point among the candidate points is sampled from a multinomial distribution with the standardized values as multinomial probabilities and passed to the next Metropolis step with (5.5). The iteration size is $N = 20,000$ with a burn in of $10,000$.

At each iteration, one of the three move type is randomly chosen and new parameter values are obtained according to the chosen move type. This MCMC algorithm produces a sample on all parameters from which the number of change points, k, can be estimated as the posterior mean, denoted by $\hat{k}$. Inference on the model parameters

is then obtained given $\hat{k}$.

Tables 7, 9 and 11 show the posterior probabilities for the number of change points, k for ARFIMA$(1, d, 1)$, ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$ models with one change point and Tables 8, 10 and 12 report the estimates of the model parameters and the location of change points, c's corresponding to $\hat{k}$ with the highest posterior probability $\pi(\hat{k}|\cdot)$. Tables 13 and 14 refer to an ARFIMA$(1, d, 1)$ model with two change points. The number of change points is correctly estimated in all cases. The method also shows good performance in the estimation of the model parameters and in the location of the change points, although there appears to be a slight overestimation of the location for ARFIMA$(0, d, 1)$ models with one change point and for the multiple change point case.

Table 7: Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 1)$ model with ϕ=0.1, d_1=0.2, d_2=0.4, θ=0.5 and one change point c=256 .

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$
0	0.0320
1	0.5381
2	0.3204
3	0.1095

Table 8: Parameter estimates of ARFIMA$(1, d, 1)$ model in the case of $\hat{k} = 1$.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$	Parameter estimates
1	0.5381	$\hat{\phi}$=0.1651 $\hat{\theta}$=0.5225 $\hat{d}_1$=0.1065 $\hat{d}_2$=0.4187 $\hat{c}_1$=252

Table 9: Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 0)$ model with ϕ=0.3, d_1=0.2, d_2=0.3 and one change point c=256.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$
0	0.3559
1	0.3932
2	0.1890
3	0.0619

Table 10: Parameter estimates of ARFIMA$(1, d, 0)$ model from simulated data in the case of $\hat{k} = 1$.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$	Parameter estimates
1	0.3931	$\hat{\phi}$=0.2425 $\hat{d}_1$=0.1362 $\hat{d}_2$=0.3282 $\hat{c}_1$=256

Table 11: Posterior probability of the number of change points k in the simulation of ARFIMA$(0, d, 1)$ model with d_1=0.2, d_2=0.3, $\theta = 0.3$ and one change point c=256.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$
0	0.1832
1	0.4982
2	0.2487
3	0.0699

Table 12: Parameter estimates of ARFIMA$(0, d, 1)$ model from simulated data in the case of $\hat{k} = 1$.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$	Parameter estimates
1	0.4982	$\hat{\theta}$=0.2902 $\hat{d}_1$=0.1123 $\hat{d}_2$=0.3739 $\hat{c}_1$=276

Table 13: Posterior probability of the number of change points k in the simulation of ARFIMA$(1, d, 1)$ model with $\phi = 0.1$, $\theta = 0.4$, $d_1 = 0.05$, $d_2 = 0.35$, $d_3 = 0.45$ and two change points at $c_1 = 128$ and $c_2 = 256$.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$
0	0.0012
1	0.3675
2	0.5447
3	0.0866

Table 14: Parameter estimates of ARFIMA$(1, d, 1)$ model from simulated data in the case of $\hat{k} = 2$..

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$	Parameter estimates
2	0.5447	$\hat{\phi}$=0.0660 $\hat{\theta}$=0.4022 $\hat{d}_1$=0.0798 $\hat{d}_2$=0.2801 $\hat{d}_3$=0.4451 $\hat{c}_1$=138 $\hat{c}_2$=284

5.5 Application to Nile River data

As an application of the method proposed in this chapter, the yearly minimum level of the Nile River from A.D. 622 to A.D. 1134 with $n = 512$ is used, which is a benchmark in the long memory literature.

Beran and Terrin (1996) fit the data with a ARFIMA$(0, d, 0)$. They reported that there is a change between the first 100 years (A.D. 722) and thereafter in the long memory parameter. They segmented the whole series into six consecutive disjoint sets of observations of length 100, and estimated the Hurst parameter H on each set. The estimated values of d were 0.0433 for the first 100 years, and 0.3531, 0.3652, 0.3281, 0.3435 and 0.4354 for the following sets. Ray and Tsay (2002) reported an estimate of $\hat{d} = 0.05$ for the first 100 years and of $\hat{d} = 0.45$.

We fitted an ARFIMA$(0, d, 0)$ model to this data. Table 15 shows the posterior estimates for the number of change points, which indicates the posterior mode at $\hat{k} = 1$ suggesting the existence of one change point in the data. Table 16 reports the estimates of the model parameters and of the location of the change point for $\hat{k} = 1$. Our method puts the change at the year A.D. 720, slightly earlier than what suggested by Beran and Terrin (1996). Our estimates of the long memory parameter before and after the estimated change point are $d_1 = 0.0668$ and $d_2 = 0.4007$, respectively. Notice that our estimate of d for $\hat{k} = 0$, i.e. $\hat{d} = 0.3704$, is very close to Beran and Terrin's estimate, 0.40, under the assumption that there is no change point. This value is also very close to the average estimate, 0.36506, for the five consecutive disjoint segments after the first 100 years found by Beran and Terrin (1996). Our inference is summarized in Figure 6.

Table 15: Posterior probability of the number of change points k in the Nile River data using ARFIMA$(0, d, 0)$ model.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$
0	0.3665
1	0.4192
2	0.1600
3	0.0543

Table 16: Parameter estimates of ARFIMA$(0, d, 0)$ model in the Nile River data when $\hat{k} = 1$.

No. of change points $\hat{k}$	Posterior probability $\pi(\hat{k}\|\cdot)$	Parameter estimates
1	0.4192	$\hat{d}_1$=0.0668 $\hat{d}_2$=0.4007 $\hat{c}_1$=98(A.D. 720)

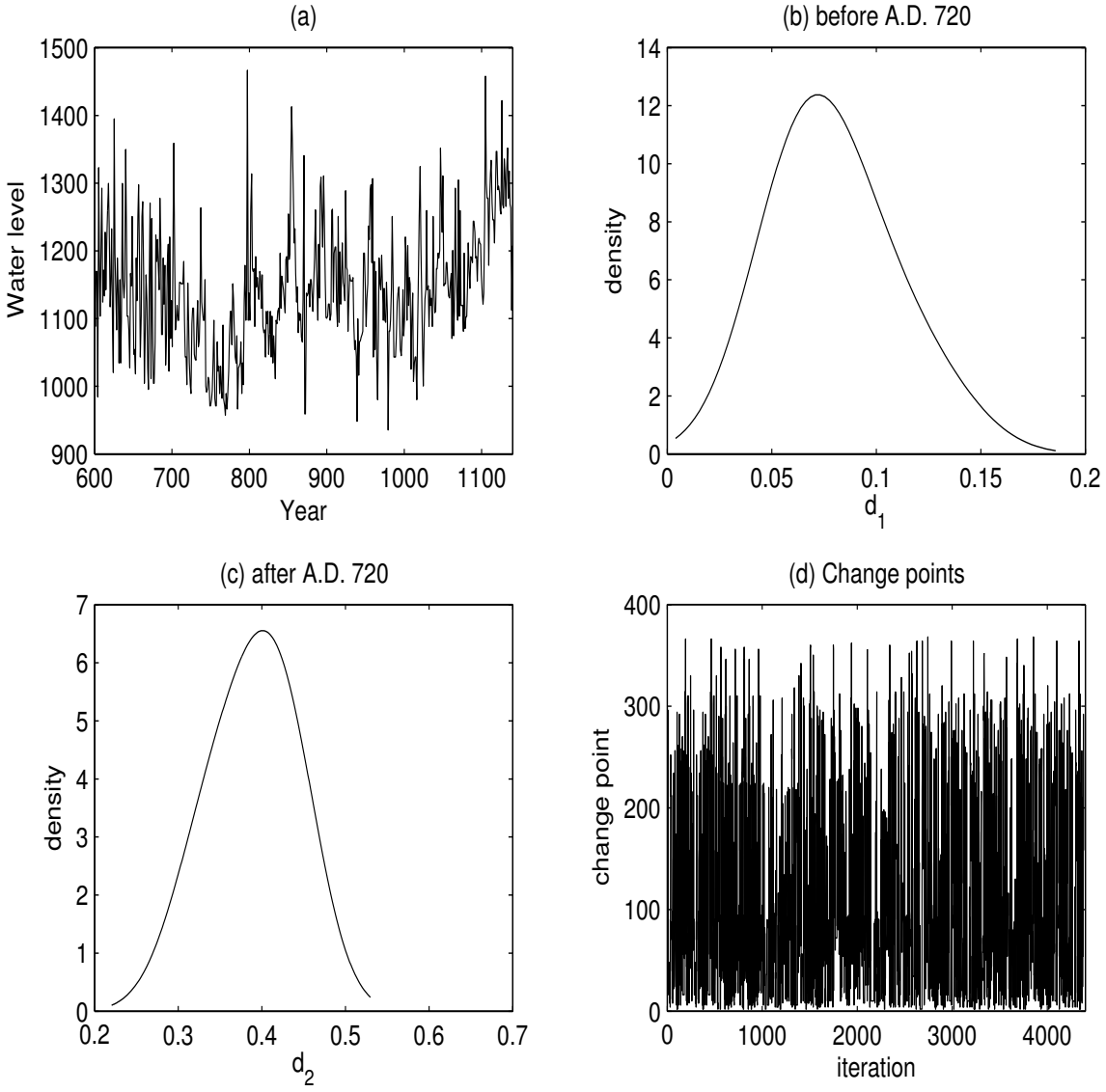

Figure 6: Nile River data: original data, plot a), and kernel estimates of the posterior density of d before A.D. 720, plot b) and after A.D. 720,plot c). Plot d) shows the MCMC trace of the number of change point

CHAPTER VI

CONCLUSION

We have proposed a wavelet-based Bayesian approach to the analysis of long memory processes, specially Gaussian ARFIMA(p, d, q), autoregressive fractionally integrated moving average models with unknown autoregressive and moving average parameters. The two main topics are Bayesian estimation of the model parameters and Bayesian detection and estimation of change points of the long memory parameter over time on wavelet domain. We have used the decorrelation nature of the discrete wavelet transform for long memory processes and carried out Bayesian posterior inference on the parameters by Markov chain Monte Carlo methods. Vannucci and Corradi's algorithm (1999) enables us to easily implement Bayesian MCMC on wavelet domain. Simulation studies and real examples have shown the usefulness of wavelet methods and Bayesian methodologies in the analysis of data from long memory processes.

For the wavelet-based Bayesian estimation of Gaussian ARFIMA(p, d, q) models, the main concern is placed on the estimation of autoregressive, moving average and long memory parameters. However the inferential procedure can be easily generalized to include other model parameter such as the white noise variance either by embedding the Metropolis sampler into a Gibbs sampler that uses the full conditional distribution of the parameter or by using a Rao-Blackwellization procedure. An open problem is to determine the order of the autoregressive and moving average polynomials together in the above estimation procedure although the simplest way is to first get the estimate of the long memory parameter from ARFIMA$(0, d, 0)$ model and then identify ARMA model using standard time series technique. The wavelet-based Bayesian estimation

procedure presented in this dissertation can be easily generalized to non-stationary ARFIMA(p, d, q) models in which the long memory parameter d is larger than or equal to 0.5.

For wavelet-based Bayesian detection and estimation of change points of the long memory parameter, the number of change points is assumed unknown. To handle the uncertainty of the number of change points reversible jump MCMC is used. Though Bayesian model averaging methods can be adopted for solving this kind of problem, reversible jump MCMC gives us an additional information which is the posterior probabilities of the number of change points and we can use them for posterior inference.

REFERENCES

Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Beran, J. and Terrin, N. (1996). "Testing for a Change of the Long-Memory Parameter." *Biometrika*, 83, 627–638.

Brown, R., Durbin, J., and Evans, J. (1975). "Techniques for Testing the Constancy of Regression Relationships over Time." *Journal of the Royal Statistical Society, Series B*, 37, 149–163.

Dahlhaus, R. (1989). "Efficient Parameter Estimation for Self-Similar Processes." *Annals of Statistics*, 17, 1749–1766.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.

Fox, R. and Taqqu, M. (1986). "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series." *Annals of Statistics*, 14, 517–532.

Geweke, J. and Porter-Hudak, S. (1983). "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis*, 4, 221–238.

Granger, C. and Joyeux, R. (1980). "An Introduction to Long Memory Time Series Models and Fractional Differencing." *Journal of Time Series Analysis*, 1, 15–29.

Green, P. (1995). "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination." *Biometrika*, 82, 711–732.

Haslett, J. and Raftery, A. (1989). "Space-Time Modeling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource." *Applied Statistics*, 38, 1–50.

Hosking, J. (1984). "Modeling Persistence in Hydrological Time Series using Fractional Differencing." *Water Resources Research*, 20, 1898–1908.

Hsu, D.-A. (1977). "Test for Variance Shift at an Unknown Time Point." *Applied Statistics*, 26, 279–284.

Hurst, H. (1951). "Long-Term Storage Capacity of Reservoirs." *Transactions of the American Society of Civil Engineers*, 116, 770–799.

Inclán, C. and Tiao, G. (1994). "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance." *Journal of the American Statistical Association*, 89, 913–923.

Jensen, M. (1999). "Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long-Memory Parameter." *Journal of Forecasting*, 18, 17–32.

— (2000). "An Alternative Maximum Likelihood Estimator of Long-Memory Processes using Compactly Supported Wavelets." *Journal of Economic Dynamics and Control*, 24(3), 361–386.

Koop, G., Ley, E., Osiewalski, J., and Steel, M. (1997). "Bayesian Analysis of Long Memory and Persistence using ARFIMA Models." *Journal of Econometrics*, 76, 149–169.

Li, W. and McLeod, A. (1986). "Fractional Time Series Modeling." *Biometrika*, 73, 217–221.

Liu, S. and Kao, M. (1999). "Bayesian Analysis for Multiple Changes of the Long Memory Parameter." Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan.

Mallat, S. (1989). "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693.

McCoy, E. and Walden, A. (1996). "Wavelet Analysis and Synthesis of Stationary Long-Memory Processes." *Journal of Computational and Graphical Statistics*, 5(1), 26–56.

McLeod, A. and Hipel, K. (1978). "Preservation of the Rescaled Adjusted Range, Parts 1, 2 and 3." *Water Resources Research*, 14, 491–512.

Pai, J. and Ravishanker, N. (1996). "Bayesian Modeling of ARFIMA Processes by Markov Chain Monte Carlo Methods." *Journal of Forecasting*, 15, 63–82.

— (1998). "Bayesian Analysis of Autoregressive Fractionally Integrated Moving-Average Processes." *Journal of Time Series Analysis*, 19(1), 99–112.

Ray, B. and Tsay, R. (2002). "Bayesian Methods for Change-Point Detection in Long-Range Dependent Processes." *Journal of Time Series Analysis*, 23, 687–705.

Sowell, F. (1992a). "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models." *Journal of Econometrics*, 53, 165–188.

— (1992b). "Modeling Long-Run Behavior with the Fractional ARIMA Model." *Journal of Monetary Economics*, 29, 277–302.

Tewfik, A. and Kim, M. (1992). "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion." *IEEE Transactions on Information Theory*, 38(2), 904–909.

Vannucci, M. and Corradi, F. (1999). "Covariance Structure of Wavelet Coefficients: Theory and Models in a Bayesian Perspective." *Journal of the Royal Statistical Society, Series B*, 61(4), 971–986.

Vostrikova, L. (1981). "Detecting 'Disorder' in Multidimensional Random Processes." *Soviet Mathematics Doklady*, 24, 55–59.

Wang, Y. (1996). "Function Estimation via Wavelet Shrinkage for Long Memory Data." *Annals of Statistics*, 24(2), 466–484.

Whitcher, B., Guttorp, P., and Percival, D. (2000). "Multiscale Detection and Location of Multiple Variance Changes in the Presence of Long Memory." *Journal of Statistical Computation and Simulation*, 68, 65–88.

Yajima, Y. (1985). "On Estimation of Long-Memory Time Series Models." *Australian Journal of Statistics*, 27, 303–320.

APPENDIX A

IN BAYESIAN ESTIMATION OF ARFIMA MODELS ON WAVELET DOMAIN

In this appendix, the posterior distribution 4.7 of $\Psi_0 = (\boldsymbol{\phi}, d, \boldsymbol{\theta})$ in Chapter IV is derived from the joint posterior distribution of Ψ, (4.6) through integrating out σ^2 which is the the variance of white noise ε. If we integrate out σ^2 in equation (4.6),

$$\begin{aligned}
\pi(\Psi_0|z_n) &\propto \left(\frac{1}{\prod_{i=1}^n \sigma^2_{z_i}(\Psi_0)}\right)^{1/2} \\
&\quad \cdot \int (1/\sigma^2)^{(n/2)+1} \exp\left[-\frac{\sum_{i=1}^n (z_i^2/\sigma^2_{z_i}(\Psi_0))}{2\sigma^2}\right] d\sigma^2 \\
&= 2^{n/2}\left(\frac{1}{\prod_{i=1}^n \sigma^2_{z_i}(\Psi_0)}\right)^{1/2}\left[\sum_{i=1}^n \left(\frac{z_i^2}{\sigma^2_{z_i}(\Psi_0)}\right)\right]^{-n/2} \\
&\quad \cdot \int \Gamma(n/2)^{-1} 2^{-n/2}\left[\sum_{i=1}^n \left(\frac{z_i^2}{\sigma^2_{z_i}(\Psi_0)}\right)\right]^{n/2} (\sigma^{-2})^{n/2-1} \\
&\quad \times \exp\left[-\frac{\sigma^{-2}}{2/\sum_{i=1}^n (z_i^2/\sigma^2_{z_i}(\Psi_0))}\right] d\sigma^{-2}.
\end{aligned}$$

Since the integrand is the form of a Gamma p.d.f., the marginal posterior distribution of Ψ_0 becomes

$$\pi(\Psi_0|z_n) \propto \left(\frac{1}{\prod_{i=1}^n \sigma^2_{z_i}(\Psi_0)}\right)^{1/2}\left[\sum_{i=1}^n \left(\frac{z_i^2}{\sigma^2_{z_i}(\Psi_0)}\right)\right]^{-n/2}.$$

APPENDIX B

IN CHANGE POINT ANALYSIS ON WAVELET DOMAIN

This appendix provides some details of the reversible jump methods described in Chapter V. The joint posterior distribution of Ψ is calculated in Section B.1. The acceptance probability α_{birth} of a proposed change point for addition in the birth step is derived in Section B.2. Finally the acceptance probability α_{death} of an existing change point for deletion in the death step is derived in Section B.3.

B.1 Joint posterior distribution of Ψ in (5.3)

Since the posterior distribution is proportional to the product of likelihood function and prior distributions,

$$\begin{aligned} f(k, \boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}, \boldsymbol{C_k}, \sigma^2 | \boldsymbol{z}) &\propto f(\mathbf{z}|\boldsymbol{\phi}, \boldsymbol{d_k}, \boldsymbol{\theta}) \cdot \pi(k, \omega_k) \\ &= \left(\frac{1}{\prod_{i=1}^{n} \sigma^2 \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \frac{z_i^2}{\sigma^2 \sigma_{z_i}^2(\Psi_0)} \right] \\ &\quad \cdot (1 + k_{max})^{-1} \cdot \binom{n-2}{k}^{-1} \sigma^{-2} \cdot 1 \cdot 1 \cdot \frac{1}{2} \\ &\propto \left(\frac{1}{\prod_{i=1}^{n} \sigma^2 \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \frac{z_i^2}{\sigma^2 \sigma_{z_i}^2(\Psi_0)} \right] \\ &\quad \cdot \binom{n-2}{k}^{-1} \sigma^{-2}. \end{aligned}$$

B.2 Derivation of the acceptance probability α_{birth} for birth step in (5.6)

To get the acceptance probability for birth step, we need three components which are Jacobian for dimension changing, proposal ratio and posterior ratio. Then the acceptance probability is calculated by $\alpha_{\text{birth}}(\zeta_k, \zeta'_{k+1}) = \min(1, \textit{posterior ratio} \times$

proposal ratio $\times |J|$). Let $T(d_{m^*}, u_2) = (d_{m^*} - u_2, d_{m^*} + u_2)$. The Jacobian for the acceptance probability for this birth stage is

$$\begin{aligned} |J| &= \left| \frac{\partial T(d_{m^*}, u_2)}{\partial (d_{m^*}, u_2)} \right| \\ &= \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} \\ &= 2. \end{aligned}$$

The posterior ratio is

$$\begin{aligned} &\textit{posterior ratio}_{\{k \to (k+1)\}} \\ &\propto \frac{\left(\dfrac{1}{\prod_{i=c_i'}^{c_{i+2}'-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i'}^{c_{i+2}'-1} \dfrac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right]^{-(c_{i+2}'-c_i')/2} \cdot \binom{n-2}{k+1}^{-1}}{\left(\dfrac{1}{\prod_{i=c_i}^{c_{i+1}-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i}^{c_{i+1}-1} \dfrac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right]^{-(c_{i+1}-c_i)/2} \cdot \binom{n-2}{k}^{-1}} \\ &\propto \frac{\left(\dfrac{1}{\prod_{i=c_i'}^{c_{i+2}'-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i'}^{c_{i+2}'-1} \dfrac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i', \boldsymbol{\theta})} \right]^{-(c_{i+2}'-c_i')/2} \cdot (k+1)}{\left(\dfrac{1}{\prod_{i=c_i}^{c_{i+1}-1} \sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right)^{1/2} \left[\sum_{i=c_i}^{c_{i+1}-1} \dfrac{z_i^2}{\sigma_{z_i}^2(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \right]^{-(c_{i+1}-c_i)/2} \cdot (n-k-2)}. \end{aligned}$$

Also, the proposal ratio can be written as

$$\begin{aligned} \textit{proposal ratio}_{\{k \to (k+1)\}} &= \frac{\beta_{k+1} \cdot (k+1)^{-1}}{\alpha_k \{2R_{u_1}(n-k-2)\}^{-1}} \\ &= \frac{2R_{u_1}\beta_{k+1}(n-k-2)}{\alpha_k(k+1)} \end{aligned}$$

The denominator of the above is from the fact that when the transition is made from ζ_k to ζ_{k+1}, three probabilistic components are needed, which one of them is a birth probability α_k since k's change points already exist and the other is a product of probability that we choose one candidate point for a birth among the $(n-k-2)$

points and a probability to split one long memory parameter into two. The numerator consists of a death probability β_{k+1} and a probability for dropping one point among the existing points, $(k+1)$. Then the acceptance probability of a proposed transition from ζ_{k+1} to ζ'_{k+1} is

$$\begin{aligned}&\alpha_{\text{birth}}(\zeta_k, \zeta'_{k+1})\\ &\quad = \min\left(1, \frac{Q(\boldsymbol{\phi}, d'_i, \boldsymbol{\theta})}{Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})} \times \frac{4R_{u_1}\beta_{k+1}}{\alpha_k}\right),\end{aligned}$$

where

$$Q(\boldsymbol{\phi}, d'_i, \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c'_i}^{c'_{i+2}-1} \sigma^2_{z_i}(\boldsymbol{\phi}, d'_i, \boldsymbol{\theta})}\right)^{1/2} \left[\sum_{i=c'_i}^{c'_{i+2}-1} \frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi}, d'_i, \boldsymbol{\theta})}\right]^{-(c'_{i+2}-c'_i)/2}$$

and

$$Q(\boldsymbol{\phi}, d_i, \boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_i}^{c_{i+1}-1} \sigma^2_{z_i}(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})}\right)^{1/2} \left[\sum_{i=c_i}^{c_{i+1}-1} \frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi}, d_i, \boldsymbol{\theta})}\right]^{-(c_{i+1}-c_i)/2}.$$

B.3 Derivation of the acceptance probability α_{death} for death step in (5.7)

Similar to the calculation of the acceptance probability in birth step, we also need three components for the acceptance probability α_{death} which is $\min(1, \textit{posterior ratio} \times \textit{proposal ratio} \times |J|)$. Let $T^*(d_u, d_{u+1}) = ((d_u + d_{u+1})/2, d_u)$. The Jacobian for dimension matching is

$$\begin{aligned}|J| &= \left|\frac{\partial T^*(d_u, d_{u+1})}{\partial(d_u, d_{u+1})}\right| \\ &= \begin{vmatrix} 1/2 & 0 \\ 1/2 & 1 \end{vmatrix} \\ &= 1/2.\end{aligned}$$

The posterior ratio is

$$
\begin{aligned}
&\textit{posterior ratio}_{\{k\to(k-1)\}} \\
&= \frac{f(k-1,\boldsymbol{\phi},\boldsymbol{d^*_{k-1}},\boldsymbol{\theta},\boldsymbol{C^*_{k-1}}|\boldsymbol{z})}{f(k,\boldsymbol{\phi},\boldsymbol{d_k},\boldsymbol{\theta},\boldsymbol{C_k}|\boldsymbol{z})} \\
&\propto \frac{\left(\frac{1}{\prod_{i=c^*_{i-1}}^{c^*_i-1}\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c^*_{i-1}}^{c^*_i-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right]^{-(c^*_i-c^*_{i-1})/2}\cdot\binom{n-2}{k-1}^{-1}}{\left(\frac{1}{\prod_{i=c_{i-1}}^{c_{i+1}-1}\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c_{i-1}}^{c_{i+1}-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right]^{-(c_{i+1}-c_{i-1})/2}\cdot\binom{n-2}{k}^{-1}} \\
&\propto \frac{\left(\frac{1}{\prod_{i=c^*_{i-1}}^{c^*_i-1}\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c^*_{i-1}}^{c^*_i-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right]^{-(c^*_i-c^*_{i-1})/2}\cdot(n-k-1)}{\left(\frac{1}{\prod_{i=c_{i-1}}^{c_{i+1}-1}\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c_{i-1}}^{c_{i+1}-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right]^{-(c_{i+1}-c_{i-1})/2}\cdot k}.
\end{aligned}
$$

Finally the proposal ratio can be written as

$$
\begin{aligned}
\textit{proposal ratio}_{\{k\to(k-1)\}} &= \frac{\alpha_{k-1}\{2R_u(n-k-1)\}^{-1}}{\beta_k k^{-1}} \\
&= \frac{\alpha_{k-1}k}{2R_u\beta_k(n-k-1)},
\end{aligned}
$$

where $R_u = \min\{d^*_u, 0.5-d^*_u\}$. We accept a transition from ζ_k to ζ^*_{k-1} with probability

$$
\alpha_{\text{death}}(\zeta_k,\zeta^*_{k-1}) = \min\left(1, \frac{Q(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}{Q(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\times\frac{\alpha_{k-1}}{4R_u\beta_k}\right),
$$

where

$$
Q(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c^*_{i-1}}^{c^*_i-1}\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c^*_{i-1}}^{c^*_i-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d^*_i,\boldsymbol{\theta})}\right]^{-(c^*_i-c^*_{i-1})/2}
$$

and

$$
Q(\boldsymbol{\phi},d_i,\boldsymbol{\theta}) = \left(\frac{1}{\prod_{i=c_{i-1}}^{c_{i+1}-1}\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right)^{1/2}\left[\sum_{i=c_{i-1}}^{c_{i+1}-1}\frac{z_i^2}{\sigma^2_{z_i}(\boldsymbol{\phi},d_i,\boldsymbol{\theta})}\right]^{-(c_{i+1}-c_{i-1})/2}.
$$

VITA

Kyungduk Ko, son of Nam-Soon Ko and Kyung-Ja Chang, was born on February 15, 1968 in Seoul, Korea. He received a Bachelor of Art degree in applied statistics from Yonsei University in Seoul, Korea in 1991. He received a Master of Arts degree in applied statistics from the same university under the direction of Dr. Chul-Eung Kim in 1993. After working at the Economic Institute and Management Research Center in Yonsei University, Seoul, Korea from September 1993 to July 1999, he continued his studies in statistics under the direction of Dr. Marina Vannucci and received a Doctor of Philosophy degree in statistics from Texas A&M University in College Station, Texas, in August 2004.

Kyungduk is married to Se Jeong Kim. They have one daughter, Minji. His permanent address is Hanyang Villa #201, 1677-2 Bongcheon 6-Dong, Kwanak-Gu, Seoul 151-056, Korea.