

TIME SERIES EXPONENTIAL MODELS: THEORY AND METHODS

A Dissertation

by

SCOTT HAROLD HOLAN

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

May 2004

Major Subject: Statistics

TIME SERIES EXPONENTIAL MODELS: THEORY AND METHODS

A Dissertation

by

SCOTT HAROLD HOLAN

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:

Emanuel Parzen
(Chair of Committee)

H. Joseph Newton
(Member)

Michael Sherman
(Member)

Joseph Ward
(Member)

Michael Longnecker
(Head of Department)

May 2004

Major Subject: Statistics

ABSTRACT

Time Series Exponential Models: Theory and Methods. (May 2004)

Scott Harold Holan, B.S.; M.S., University of Illinois at Chicago

Chair of Advisory Committee: Dr. Emanuel Parzen

The exponential model of Bloomfield (1973) is becoming increasingly important due to its recent applications to long memory time series. However, this model has received little consideration in the context of short memory time series. Furthermore, there has been very little attempt at using the EXP model as a model to analyze observed time series data.

This dissertation research is largely focused on developing new methods to improve the utility and robustness of the EXP model. Specifically, a new nonparametric method of parameter estimation is developed using wavelets. The advantage of this method is that, for many spectra, the resulting parameter estimates are less susceptible to biases associated with methods of parameter estimation based directly on the raw periodogram. Additionally, several methods are developed for the validation of spectral models. These methods test the hypothesis that the estimated model provides a whitening transformation of the spectrum; this is equivalent to the time domain notion of producing a model whose residuals behave like the residuals of white noise. The results of simulation and real data analysis are presented to illustrate these methods.

To my Mom, Dad, and Christie

ACKNOWLEDGEMENTS

I am indebted to many people for helping me attain my Ph.D. degree. I attribute a great deal of my success to the positive interactions I have had with my family, friends, and various mentors. I feel very fortunate that I have been able to surround myself with so many people that have had a significant impact on my life.

First, I am greatly indebted to my adviser Dr. Emanuel Parzen. I am honored to have had the opportunity to be his student; it has been a privilege to get a glimpse of the field of statistics through his eyes. Dr. Parzen has taught me to be a self-sufficient researcher; I will always be grateful to him.

Additionally, I would like to thank my parents. My mother's emotional support and encouragement over the years has been immeasurable. Additionally, her willingness to make sacrifices on my behalf has always inspired me to try to achieve more than I thought was possible. My father and his wife, Lila, gave me encouragement and financial support, without which my educational goals could not have been realized. Additionally, I feel that my drive, hard work and focus are a result of seeing these same characteristics in my father.

I would also like to thank the members of my committee: Dr. H. Joseph Newton, Dr. Michael Sherman, and Dr. Joseph Ward. I appreciate all of the time they took out of their busy schedules to attend my preliminary examination and dissertation defense. Additionally, I would like to thank Dr. Newton for always making time to answer my questions despite his busy schedule as Dean of Science. Furthermore, I would like to thank Dr. Sherman for all he has taught me about consulting and collaborating with non-statisticians.

Another person I am indebted to is Dr. Charles Phillips. I have been extremely

fortunate to have been able to work as a graduate research assistant for him and for Dr. Catherine Hawes. The consulting and collaborating experiences that I have obtained through my assistantship are priceless. I have grown as a person and a statistician as a result of having worked for him and Dr. Hawes.

Many other people have been instrumental in helping me fulfill my educational goals. For example, Judith Baxter, Dr. Shmuel Friedland, and Dr. P. Fred Dahm each played an important role in helping me obtain either my undergraduate or graduate education. My sister, Jodi has always shown me that she wants only what is best for me. My friends, Gaspar Porta and Joshua Rogers, have given me constant support and encouragement. I am grateful for everything that these individuals, and many others, have done to help me succeed.

Lastly, I want to express my warmest thanks to my wife Christine Spinka. Without her friendship, emotional support, and technical assistance none of this may have been possible. I am unbelievably fortunate to have someone so inspirational in my life. I am extremely thankful to her for making my graduate school experience infinitely more pleasant.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER	
I INTRODUCTION	1
1.1 The Problem	1
1.2 Literature Review	2
II BLOOMFIELD'S EXPONENTIAL MODEL FOR THE SPEC- TRUM OF A SCALAR TIME SERIES	4
2.1 Bloomfield's Exponential Model	4
2.2 Bloomfield's Maximum Likelihood Parameter Estimation .	5
2.3 Alternative Definition for the EXP Model	7
III CEPSTRAL CORRELATIONS, $MA(\infty)$, AND POURAH- MADI'S FORMULAS	10
3.1 Cepstral Correlations	10
3.2 Pourahmadi's Formulas	10
3.3 Empirical Examples	13
IV MEMORY TYPE OF A TIME SERIES	15
4.1 Definitions	15
4.2 Identifying Time Series Memory Type	22

CHAPTER		Page
V	SIMULATING DATA FROM EXPONENTIAL MODELS	23
	5.1 Introduction	23
	5.2 Davies-Hart Method	23
	5.3 Exact Time Domain Method	25
	5.4 Indirect Method	26
	5.5 Example of Indirect Simulation Method	28
VI	METHODS FOR SHORT MEMORY EXP PARAMETER ESTIMATION	33
	6.1 Introduction	33
	6.2 Parameter Estimation by Minimum Renyi Information Divergence	33
	6.3 Parameter Estimation by Log Periodogram Regression	35
	6.4 Multitaper Spectral Density Estimation	38
	6.5 EXP(m) via Multitaper Wavelet Spectrum Estimation	43
	6.6 Simulation Results	47
VII	ORDER SELECTION AND MODEL VALIDATION	58
	7.1 Introduction and Philosophy	58
	7.2 Linhart and Volkers Criterion	60
	7.3 Order Selection	65
	7.4 Distribution of the Sample Whitening Correlations	68
	7.5 Information Diagnostic	70
	7.6 Whitening Spectral Distribution	82
	7.7 EXP Model Validation by AIC	84
	7.8 Frequency Domain Data-Driven Portmanteau Test	87
	7.9 Data Analysis	90
VIII	AREXP MODELS	98
	8.1 Overview and Algorithm	98
	8.2 Wolf Sunspot Data	101
IX	CONCLUSION	106
	9.1 Concluding Remarks	106
	9.2 Problems for Further Study	107

	Page
REFERENCES	108
VITA	112

LIST OF TABLES

TABLE		Page
1	MA(k), $k = 12$, representation of the EXP(4) model and the associated $SS_{sim}(k)$	29
2	Model 1 with N=1000: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	49
3	Model 1 with N=512: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	50
4	Model 1 with N=256: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	51
5	Model 1 with N=100: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	52
6	Model 2 with N=1000: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	53
7	Model 2 with N=512: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	54
8	Model 2 with N=256: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	55
9	Model 2 with N=100: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation.	56
10	Linhart-Volkers criterion: Frequency of order chosen for model 1 using both maximum likelihood and wavelet parameter estimation.	63

TABLE	Page
11 Linhart-Volkers criterion: Frequency of order chosen for model 2 using both maximum likelihood and wavelet parameter estimation.	64
12 This table displays the parameter estimates and $ISE_m^{(d)}$ for the simulated model given in expression (7.25).	68
13 Descriptive statistics for $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$, and Shapiro-Wilk p-value.	70
14 Critical values for the test statistic given in expression (7.64).	77
15 Percentiles of the limiting distribution of $Q_w(\hat{m})$	89
16 This table displays the parameter estimates and $ISE_m^{(d)}$ for the estimated model.	91
17 Values of EXP model validation criteria using AIC.	96
18 The values of $\hat{\phi}(\tau)$ and $\text{Err}(\tau)$ for values of $\tau = 1, 2, \dots, 15$	102
19 The values of the EXP(12) model parameters for the AREXP model fit to the Wolf Sunspot data 1700-1988.	102

LIST OF FIGURES

FIGURE		Page
1	Plot of $SE(f_{EXP}, f_{MA})$	30
2	An example of the log EXP spectral density and its corresponding truncated log MA spectral density.	31
3	Plot of $\log f(\lambda)$ and $\log f^{\sim}(\lambda) + \gamma$	32
4	An example plot of $\log f(\lambda)$ and $\log f^{\sim}(\lambda) + .57721$	57
5	Plot of $\log f^{\sim}(\lambda) + \gamma$ superimposed with $g_{(mthr)}^{\sim}$ and plot of EXP(3) with $g_{(mthr)}^{\sim}$ superimposed.	67
6	Time plot of the logarithm of tree ring indices/100, for <i>Widdingtonia cedarbergensis</i>	92
7	Plot of log periodogram +.57721 with fitted EXP(3) model.	94
8	Plot of the sample whitening spectral distribution function with asymptotic 95% confidence bands from fitting the EXP(3) model.	95
9	Plot of log periodogram +.57721 with fitted log AREXP model for the Wolf Sunspot data 1700-1988.	104
10	Plot of the sample whitening spectral distribution function with asymptotic 95% confidence bands from fitting the AREXP model.	105

CHAPTER I

INTRODUCTION

1.1 The Problem

The exponential model (EXP) of Bloomfield (1973) is becoming increasingly important due to its recent applications to modelling long memory time series. However, this model has received little consideration in the context of short memory time series. Furthermore, there has been very little attempt at using the EXP model to analyze observed time series data.

This work is comprised of three main goals; the first is to highlight the work of Bloomfield (1973) and to provide some remarks and observations not found in the literature. The second goal is to discuss important contributions that have been made in this area. The third goal is to extend the scope of the EXP model, increasing its utility and robustness in short memory time series modelling. In particular, we develop a strategy for parameter estimation and model fitting of EXP models.

Additionally, in this work EXP models are considered algorithmic models rather than data models. Algorithmic models are characterized by the lack of any assumption that the observed time series is generated from a stochastic model of a stationary time series with a hypothesized spectral density. Conversely, data models make this assumption. Thus, the overall problem to be solved is to produce a unified framework for applying EXP models as algorithmic models and to empirically investigate the properties of the developed methods. This framework will include new methods for simulation, parameter estimation, order selection, and model validation.

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

1.2 Literature Review

Very few sources exist containing information addressing the problem of modelling short memory time series using the EXP(m) model. In Bloomfield (1973) the exponential model, for the spectrum of a scalar time series, is introduced. The model suggested by Bloomfield is a finite parameter model that is explicitly formulated in terms of the spectral density. The model is motivated by the observation that if the log of an estimated spectral density is a fairly well behaved function then it can be well approximated by a truncated Fourier series. The form Bloomfield chose for this model was

$$g(\lambda) = \frac{\tau^2}{2\pi} \exp \left\{ 2 \sum_{r=1}^p \theta_r \cos(r\lambda) \right\}.$$

Furthermore, Bloomfield presents an iterative procedure for implementing maximum likelihood estimation of the model parameters.

The work of Pourahmadi (1983) focuses on obtaining recursive formulas for the AR(∞) and MA(∞) coefficients using cepstral correlations, see Chapter III. Furthermore, Pourhamadi points out that “Bloomfield (1973) proposes a parametric method for density estimation where the parameters are cepstral correlations.” Thus, in this work the author provides a means of connecting Bloomfield’s frequency domain model with existing time domain representations.

Linhart and Volkers (1985) develop a general order selection criterion for parametric models of stationary Gaussian time series. The criterion is based on the (one step ahead) mean squared error of prediction. Additionally, the order selection criterion is explicitly formulated for the case of Bloomfield’s EXP(m) models.

Fitting models to spectra using standard software regression packages is examined in Cameron and Turner (1987). One aspect discussed in their paper is the use of regression software in fitting Bloomfield’s EXP model. More specifically, the au-

thors describe how the frequency domain maximum likelihood estimation algorithm of Bloomfield (1973) can be restated as an iteratively reweighted least squares regression and hence calculated using standard regression software.

More recently, Parzen (1993) examines the use of Renyi information divergence as a means of estimating finite parameter spectral densities, as well as goodness of fit by components and exponential models. In this context Parzen discusses exponential models of the form

$$\log f_{\theta,m}(\lambda) = \theta_0 + \theta_1 J_1(\lambda) + \cdots + \theta_m J_m(\lambda)$$

where $\{1, J_1(\lambda), J_2(\lambda), \dots\}$ form a complete orthonormal set of functions on $L_2[0, 1]$. The case where $J_k(\lambda) \equiv 2 \cos(2\pi k\lambda)$, $k = 1, 2, \dots$ and $\theta_0 \equiv \int_0^1 \log f(\lambda) d\lambda = \log \sigma_\infty^2$ reduces to Bloomfield's EXP model.

None of the above papers takes the approach considered in this work. The approach considered in this work is to both fit EXP models algorithmically and to validate goodness of fit on the criterion that the estimated model provides a whitening transformation of the spectrum. This criterion is equivalent to testing the hypothesis that the residual (whitening) correlations, the correlations associated with the whitening spectral density, behave like the residuals of white noise.

CHAPTER II

BLOOMFIELD'S EXPONENTIAL MODEL FOR THE SPECTRUM OF A
SCALAR TIME SERIES**2.1 Bloomfield's Exponential Model**

Let $\{X_t\}$ be an observed time series and $\{\epsilon_t\}$ be an unobserved white noise series. One class of models frequently used when fitting parametric models to observed time series is the autoregressive moving average (ARMA) model. In the original notation of Bloomfield (1973) this model can be written as

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \epsilon_t - \sum_{r=1}^q \theta_r \epsilon_{t-r}. \quad (2.1)$$

The spectral density function of this series is given by

$$g(\omega) = \frac{\tau^2}{2\pi} \left| \frac{1 - \sum_{r=1}^q \theta_r e^{ir\omega}}{1 - \sum_{r=1}^p \phi_r e^{ir\omega}} \right|^2 \quad (2.2)$$

where τ^2 is the variance of ϵ_t , the innovation variance.

The exponential model, pioneered by Bloomfield (1973), is motivated by the observation that often the logarithm of an estimated spectral density is a fairly well behaved function and therefore can be well approximated by a truncated Fourier series. The form chosen, by Bloomfield, for the model is

$$\begin{aligned} g(\omega) &= \frac{\tau^2}{2\pi} \exp \left(2 \sum_{r=1}^p \theta_r \cos(r\omega) \right) \\ &= \frac{\tau^2}{2\pi} h(\omega; \theta). \end{aligned} \quad (2.3)$$

Note that the evenness of g restricts the Fourier series to be a cosine series.

2.2 Bloomfield's Maximum Likelihood Parameter Estimation

The estimation of a general model of the form

$$g(\omega) = \frac{\tau^2}{2\pi} h(\omega; \theta). \quad (2.4)$$

is discussed by Walker (1964). In this work, Walker shows that if the time series is Gaussian, then the log likelihood of the parameters given the observed data x_1, \dots, x_N is approximately

$$-\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\tau^2) - \frac{N}{2\tau^2} \int_{-\pi}^{\pi} \frac{I(\omega)}{h(\omega; \theta)} d\omega \quad (2.5)$$

where

$$I(\omega) = \frac{1}{2\pi N} \left| \sum_{t=1}^N x_t e^{it\omega} \right|^2 \quad (2.6)$$

is the sample spectral density (periodogram) of x_1, \dots, x_N . Therefore, θ may be estimated by minimizing

$$\int_{-\pi}^{\pi} \frac{I(\omega)}{h(\omega; \theta)} d\omega = \int_{-\pi}^{\pi} \frac{1}{2\pi N} \left| \sum_{t=1}^N x_t e^{it\omega} \right|^2 \exp \left(-2 \sum_{r=1}^p \theta_r \cos(r\omega) \right) d\omega \quad (2.7)$$

and τ^2 may be estimated by the minimized value. Furthermore, Walker shows that $\hat{\theta}$ and $\hat{\tau}^2$ are consistent for θ and τ^2 respectively and that

$$\sqrt{N} (\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, W(\theta_0)^{-1}) \quad (2.8)$$

where

$$W(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \theta} \log(h(\omega; \theta)) \right] \left[\frac{\partial}{\partial \theta} \log(h(\omega; \theta)) \right]' d\omega. \quad (2.9)$$

Assuming the model is a true EXP(p) and not misspecified, it can be verified that $W(\theta)$ is equal to the identity matrix. This result can be used to implement direct minimization of expression (2.7), in order to obtain estimates of $\hat{\theta}$ and $\hat{\tau}^2$.

Using a Newton-Raphson minimization procedure gives rise to an iterative scheme in which

$$\hat{\theta}_{\mathbf{n}+1} = \hat{\theta}_{\mathbf{n}} - \left[V(\hat{\theta}_{\mathbf{n}}) \right]^{-1} v(\hat{\theta}_{\mathbf{n}}) \quad (2.10)$$

where $v(\hat{\theta}_{\mathbf{n}})$ and $V(\hat{\theta}_{\mathbf{n}})$ are the vector and matrix of first and second derivatives of expression (2.7) evaluated at $\hat{\theta}_{\mathbf{n}}$. Next, $V(\hat{\theta}_{\mathbf{n}})$ can be approximated by $V(\theta_0)$; while $V(\theta_0)$ can be approximated by its expectation since its variance is $O(n^{-1})$. Furthermore, $E[V(\theta_0)]$ can be approximated by $-2\tau^2 W(\theta_0) = -2\tau^2 \mathbf{I}$. Replacing τ^2 by its n^{th} iterated estimate $\hat{\tau}_n^2$, which is expression (2.7) evaluated at $\hat{\theta}_{\mathbf{n}}$, an iterative solution in which

$$\hat{\theta}_{\mathbf{n}+1} = \hat{\theta}_{\mathbf{n}} + \frac{1}{2\hat{\tau}_n^2} v(\hat{\theta}_{\mathbf{n}}) \quad (2.11)$$

where

$$\begin{aligned} \frac{1}{2\hat{\tau}_n^2} v(\hat{\theta}_{\mathbf{n}}) &= \frac{1}{2\hat{\tau}_n^2} \int_{-\pi}^{\pi} \frac{I(\omega)}{h(\omega; \theta)} \frac{\partial}{\partial \theta} \log[h(\omega; \theta)]|_{\theta_{\mathbf{n}}} d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{2\pi I(\omega)}{\hat{\tau}_n^2 h(\omega; \hat{\theta}_{\mathbf{n}})} \mathbf{c}(\omega) d\omega \\ &= \int_{-\pi}^{\pi} \frac{1}{\hat{\tau}_n^2} \frac{1}{2\pi N} \left| \sum_{t=1}^N x_t e^{it\omega} \right|^2 \exp\left(-2 \sum_{r=1}^p \hat{\theta}_{r_n} \cos(r\omega)\right) \mathbf{c}(\omega) d\omega \end{aligned} \quad (2.12)$$

and $\mathbf{c}(\omega)' = (\cos(\omega), \dots, \cos(p\omega))$ is obtained. Note, in the original work, Bloomfield updates his parameter estimates using

$$\hat{\theta}_{\mathbf{n}+1} = \hat{\theta}_{\mathbf{n}} - \frac{1}{2\hat{\tau}_n^2} v(\hat{\theta}_{\mathbf{n}}) \quad (2.13)$$

where the negative sign is presumably the result of a typo. Additionally, note that central to this procedure is that under the assumption of normality and the absence of any model misspecification, $W(\theta_0) = \mathbf{I}$. This observation makes direct minimization of expression (2.7) more tractable when implementing the modified Newton-Raphson iterative solution.

The starting values Bloomfield suggests for the iterative parameter estimation scheme are taken to be

$$\hat{\theta}_1 = \frac{2}{N} \sum_{j=1}^m \log[I(\omega_j)] \cos(r\omega_j), \quad r = 1, \dots, p \quad (2.14)$$

where $\omega_j = \frac{2\pi j}{N}$ are the Fourier frequencies and $m = \lceil (\frac{1}{2}(N-1)) \rceil$. Note, $\lceil x \rceil$ is the ceiling function, the smallest integer greater than or equal to x . Additionally, the criterion for determining convergence is based on the sum of squares of $\hat{\theta}_n - \hat{\theta}_{n-1}$ multiplied by the series length and divided by the number of parameters being estimated. The iterations are terminated when this quantity first falls below .01, this is when the root mean square modification first falls below one tenth of the common asymptotic standard deviation. In practice we choose a slightly stronger criterion in order to insure convergence of the parameter estimates. The criterion we choose for determining convergence is when the quantity first falls below .00001. This will be discussed further in Chapter VI. Note, in Bloomfield's iterative procedure all multiparameter models are fit sequentially. Additionally, goodness of fit of the estimated model is assessed by the residual variance, since this is an estimate of the one step ahead mean square error of prediction achieved using the fitted model.

2.3 Alternative Definition for the EXP Model

Even though there only exists a small body of literature discussing Bloomfield's EXP model, several definitions for this model have been suggested. In this work, the EXP model also differs from its original definition. In this section we define the EXP model and set up some of the notation that will be used.

Let $f(\lambda)$ be the true spectral density function. It is important to note that in this work the spectral density is defined on the unit interval $0 \leq \lambda < 1$, and interpreted on the interval $0 \leq \lambda < .5$. Additionally, note, $f(\lambda) = f(1 - \lambda)$. An important way to

approximate a spectral density $f(\lambda)$ is by using exponential spectral models of order m . They can be defined by

$$\log f_{\theta,m}(\lambda) = \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k\lambda) \quad (2.15)$$

where

$$\theta_0 = \int_0^1 \log f_{\theta,m}(\lambda) d\lambda = \log \sigma_\infty^2. \quad (2.16)$$

Thus, $\exp(\theta_0)$ is equal to the innovation variance. Defining the EXP model of order m as in expression (2.15) has immediate benefits in terms of interpretation. Consider the system $C = \{1, \cos(2\pi\lambda), \cos(2\pi 2\lambda), \dots\}$ of cosine functions. For any spectral density function $f(\lambda)$, such that $\log f(\lambda)$ is absolutely integrable on $(0, 1)$, define the Fourier coefficients of $\log f(\lambda)$ by

$$\theta_k = \int_0^1 \log f(\lambda) \cos(2\pi k\lambda) d\lambda, \quad k = 0, 1, \dots \quad (2.17)$$

The system C is complete for $C[0,1]$, the class of continuous functions on $[0,1]$. In other words given any spectral density function, such that $\log f(\lambda) \in C[0,1]$,

$$\log f_{\theta,m}(\lambda) = \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k\lambda), \quad 0 \leq \lambda < 1, \quad (2.18)$$

converges to $\log f(\lambda)$ in mean square as $m \rightarrow \infty$, see Hart (1997). This implies that the integrated squared error $\text{ISE}[\log f_{\theta,m}(\lambda), \log f(\lambda)]$ converges to 0 as $m \rightarrow \infty$.

Additionally, the system C forms an orthogonal basis for $C[0,1]$ which implies that any continuous function on the interval $[0,1]$ can be well approximated by a finite linear combination of the elements of C . Often, in practice, when fitting the EXP model to short memory spectra the θ_k decay to 0 quickly implying that there exists a small value of m such that

$$\log f_{\theta,m}(\lambda) \approx \log f(\lambda), \quad \forall \lambda \in [0, 1]. \quad (2.19)$$

A more thorough discussion of the parameters θ_k , also known as cepstral correlations (coefficients), will be given in Chapter III.

Another benefit of defining the EXP model by (2.15) is its similarity to the spectral representation of $R(\nu)$, where $R(\nu)$ is the autocovariance function of lag ν , for a covariance stationary time series. For example, given that $R(\nu)$ is absolutely summable and that $f(\lambda)$ is continuous

$$R(\nu) = \int_0^1 f(\lambda) \cos(2\pi\nu\lambda) d\lambda, \quad \nu \in Z \quad (2.20)$$

and

$$f(\lambda) = R(0) + 2 \sum_{\nu=1}^{\infty} R(\nu) \cos(2\pi\nu\lambda), \quad \lambda \in [0, 1]. \quad (2.21)$$

Moreover, it is well known that if the spectral density comes from an MA(m), then $R(\nu) = 0$ for all $|\nu| > m$. This implies

$$f(\lambda) = R(0) + 2 \sum_{\nu=1}^m R(\nu) \cos(2\pi\nu\lambda), \quad \lambda \in [0, 1]. \quad (2.22)$$

Thus, the order m for the moving average model is such that the autocorrelations of lag greater than m are equal to zero. Similarly, the order m for the autoregressive model and exponential model are such that the inverse autocorrelations of lag greater than m and the cepstral correlations of order greater than m are equal to zero respectively. Thus, the spectral density of a short memory process can be modelled parametrically by f , $1/f$ and $\log(f)$ where f corresponds to a MA model, $1/f$ to an AR model and $\log(f)$ to an EXP model. The order of the coefficient where the Fourier expansion of f , $1/f$ and $\log(f)$ vanishes is then the order of the MA, AR and EXP model respectively.

CHAPTER III

CEPSTRAL CORRELATIONS, $MA(\infty)$, AND POURAHMADI'S FORMULAS**3.1 Cepstral Correlations**

Traditionally computation of $MA(\infty)$ coefficients proceeds parametrically by inversion of an $AR(p^{\wedge})$ model. Another method of computing $MA(\infty)$ coefficients is through the use of recursive formulas developed by Pourahmadi (1983), which link the $MA(\infty)$ coefficients, ψ_j , to the cepstral correlations.

Definition 1. The cepstral correlations (coefficients) are defined by

$$\begin{aligned} c(h) &= \int_0^1 \log f(\lambda) \exp(2\pi i h \lambda) d\lambda \\ &= \int_0^1 \log f(\lambda) \cos(2\pi h \lambda) d\lambda. \end{aligned} \quad (3.1)$$

Note, the name ‘‘ceps’’ was originally given by Tukey as a reverse spelling of ‘‘spec’’. Moreover, in speech recognition literature the name ‘‘cepstral coefficient’’ is given to descriptive statistics similar to $c(h)$. Additionally, note,

$$c(0) = \int_0^1 \log f(\lambda) d\lambda \quad (3.2)$$

and that θ_h of expression (2.17) are equivalent to $c(h)$, for all h .

3.2 Pourahmadi's Formulas

Let $\{Y(t)\}$ be a covariance stationary time series with spectral density $f(\lambda)$, $0 \leq \lambda \leq 1$. It is well known that $\{Y(t)\}$ is purely nondeterministic if and only if

$$\int_0^1 \log f(\lambda) d\lambda > -\infty, \quad (3.3)$$

see Brockwell and Davis (1996). Next, let $\{Y^\nu(t)\}$ denote the normalized innovation process of $\{Y(t)\}$. Then $\{Y^\nu(t)\}$ is also covariance stationary and has constant spectral density function identically equal to 1 on $[0, 1]$. By Wold's decomposition, Doob (1953) (as cited in Pourahmadi, 1983), $Y(t)$ has a one sided MA(∞) representation given by

$$Y(t) = \sum_{k=0}^{\infty} \psi_k Y^\nu(t-k) \quad (3.4)$$

where $\psi_0 \equiv 1$. Additionally,

$$\sigma Y^\nu(t) = Y(t) - \widehat{Y}_t(t-1) \quad (3.5)$$

and

$$\sigma^2 = \exp\left(\int_0^1 \log f(\lambda) d\lambda\right) \quad (3.6)$$

where $\widehat{Y}_t(t-1)$ is the linear least squares predictor of $Y(t)$ based on $\{Y_k : k \leq t-1\}$.

Furthermore, Pourahmadi states it is known that

$$f = |\phi|^2 = \phi \bar{\phi} \quad (3.7)$$

where

$$\phi(\lambda) = \sum_{k=0}^{\infty} \psi_k e^{2\pi i k \lambda}, \quad (3.8)$$

$$\sum_{k=0}^{\infty} \psi_k^2 < \infty \quad (3.9)$$

and

$$\sigma = \psi_0 > 0. \quad (3.10)$$

Additionally, ϕ (the optimal factor of f) is unique and has an analytic extension to the open disc D in the complex plane which can be written

$$\phi_+(z) = \exp\left[\frac{c(0)}{2} + \sum_{k=1}^{\infty} c(k) z^k\right] \quad (3.11)$$

where

$$c(k) = \int_0^1 \log f(\lambda) \exp(-2\pi i k \lambda) d\lambda. \quad (3.12)$$

Moreover, Pourahmadi states that (3.7) and (3.12) imply the coefficients ψ_k and $c(k)$ satisfy the identity

$$\sum_{k=0}^{\infty} \psi_k z^k = \exp \left[\frac{c(0)}{2} + \sum_{k=1}^{\infty} c(k) z^k \right]. \quad (3.13)$$

Next, define the MA(∞) coefficients $\psi(h)$

$$Y(t) = Y^\nu(t) + \psi_1 Y^\nu(t-1) + \psi_2 Y^\nu(t-2) + \dots \quad (3.14)$$

and the AR(∞) coefficients $a(h)$ by

$$Y(t) + a(1)Y(t-1) + a(2)Y(t-2) + \dots = Y^\nu(t). \quad (3.15)$$

Differentiating both sides of expression (3.13) followed by some algebra, Pourahmadi (1983) obtains the following recursive formula

$$\psi_h = \frac{1}{h} \sum_{k=1}^h k c(k) \psi_{h-k}, \quad h = 1, 2, \dots \quad (3.16)$$

Similarly, since ϕ_+ has all its roots outside of D, ϕ_+^{-1} is analytic in D and thus has a power series representation given by

$$\phi_+^{-1} = \sum_{k=0}^{\infty} a(k) z^k. \quad (3.17)$$

Comparing this expression with expression (3.13) it follows that

$$\sum_{k=0}^{\infty} a(k) z^k = \exp \left[-\frac{c(0)}{2} - \sum_{k=1}^{\infty} c(k) z^k \right]. \quad (3.18)$$

Therefore, the recursive formula

$$-a(h) = \frac{1}{h} \sum_{k=1}^h k c(k) a(h-k), \quad h = 1, 2, \dots \quad (3.19)$$

is also obtained.

3.3 Empirical Examples

3.3.1 Example 1

Consider the following AR(2) model

$$Y(t) - .90Y(t-1) + .70Y(t-2) = Y^\nu(t). \quad (3.20)$$

Assuming the variance of $Y^\nu(t)$ equals 1, the spectral density of $Y(t)$ is given by

$$f(\lambda) = \frac{1}{|1 - .90 \exp(2\pi i\lambda) + .70 \exp(2\pi 2i\lambda)|^2}, \quad 0 \leq \lambda \leq 1. \quad (3.21)$$

Thus,

$$g(\lambda) = \log f(\lambda) = -\log (|1 - .90 \exp(2\pi i\lambda) + .70 \exp(2\pi 2i\lambda)|^2). \quad (3.22)$$

Next, the first four cepstral correlations are obtained numerically (only cepstral correlations greater than .14 in absolute value are retained). These are .8994, -.2961, -.3871, and -.1571. Using Pourahmadi's formula the first three autoregressive coefficients are -.90, .70, and 0, precisely the coefficients of the original AR(2) model.

3.3.2 Example 2

Similarly, consider the following MA(3) model

$$Y(t) = Y^\nu(t) - .70Y^\nu(t-1) - .10Y^\nu(t-2) + .60Y^\nu(t-3). \quad (3.23)$$

Again, assuming the variance of $Y^\nu(t)$ equals 1, the spectral density of $Y(t)$ is given by

$$f(\lambda) = |1 - .70 \exp(2\pi i\lambda) - .1 \exp(2\pi i2\lambda) + .60 \exp(2\pi i3\lambda)|^2, \quad 0 \leq \lambda \leq 1. \quad (3.24)$$

Thus,

$$g(\lambda) = \log f(\lambda) = \log (|1 - .70 \exp(2\pi i\lambda) - .1 \exp(2\pi i2\lambda) + .60 \exp(2\pi i3\lambda)|^2). \quad (3.25)$$

Next, the first five cepstral correlations are obtained numerically (only cepstral correlations greater than .14 in absolute value are retained). These are -.6997, -.3436, .4170, .3062, and .2783. Using Pourahmadi's formula the first four MA coefficients are -.70, -.10, .60 and 0, precisely the coefficients of the original MA(3) model.

CHAPTER IV

MEMORY TYPE OF A TIME SERIES

4.1 Definitions

In order to successfully model a time series a criterion is needed for determining its “memory type”. Memory type is usually divided into three broad classes, no memory, short memory, and long memory. Additionally, time series can be further classified as stationary or non-stationary. Similar to Newton (1988) memory can be defined as follows:

1. **No memory**- This type of series is completely random and shows no trends over time. That is observations at different time points are uncorrelated. This type of series is also called white noise in analogy to the physical spectrum of white light, which is constant over all frequencies. Moreover, a sequence of uncorrelated random variables is called white noise.
2. **Long memory**- This type of series is the opposite of no memory. Furthermore, this type of series can be characterized by the feature that the series is almost perfectly predictable far into the future. This is because the correlation between observations at one point in time and observations at previous time points decays slowly.
3. **Short memory**- This type of series falls between no memory and long memory. Furthermore, this type of series has the property that the dependence on past observations dies out rapidly. Therefore, knowing the past provides some information but does not allow us to forecast the future with a high degree of accuracy.

Lastly, stationarity can be thought of intuitively by the notion that realizations of the time series over two equal time intervals should possess similar statistical characteristics. The above classifications can be stated using many different precise mathematical definitions. In what follows some of these definitions are stated explicitly in order to aid in the development of diagnostics for proper memory classification and model validation.

One characterization of memory is due to long memory modelling. This characterization can be formulated using the recent application of the EXP model in this area. More specifically, the fractionally differenced exponential model (FEXP), discussed by Janacek (1982) and Beran (1993) among others. The FEXP model is defined by

$$f(\lambda) = |1 - e^{-2\pi i\lambda}|^{-2d} f^*(\lambda) \quad (4.1)$$

where $f^*(\lambda)$ is the short memory EXP model. Memory type is then classified in terms of the memory parameter d ,

$$\begin{aligned} d &\in \left(-\frac{1}{2}, \frac{1}{2}\right), \neq 0 \longrightarrow \text{stationary, long memory} \\ d &< -\frac{1}{2} \longrightarrow \text{long memory, non-invertible} \\ d &> \frac{1}{2} \longrightarrow \text{long memory, non-stationary} \\ d &= 0 \longrightarrow \text{stationary, short memory.} \end{aligned} \quad (4.2)$$

In this work, unless otherwise stated, it is assumed that $d = 0$, thus expression (4.1) reduces to the (stationary) short memory EXP model defined by expression (2.15). Although, this characterization of memory in terms of the memory parameter is mathematically precise, knowing d does not give full information about the memory type of the process. That is, given only the information that $d = 0$, provides no indication whether the process is short memory or no memory. Therefore, when

modelling processes suspected of being stationary and short memory it may be more illuminating to consider other diagnostics in order to characterize memory type, as they may also be adapted for the purposes of model validation.

Another method of defining memory type is in terms of the autocorrelation function. Let $Y(t)$ be a mean zero stationary process with autocovariance function $R(\nu) = E[Y(t)Y(t + \nu)]$. One definition of the autocorrelation function, $\rho(\nu)$, is $\rho(\nu) = R(\nu)/R(0)$. As was previously eluded to, time series memory type can be characterized by how rapidly $\rho(\nu)$ decays to zero. More specifically, in terms of $\rho(\nu)$, the definition of the three memory types is:

$$\begin{aligned} \sum_{\nu=1}^{\infty} |\rho(\nu)| &= 0 \quad \longrightarrow \text{no memory} \\ 0 < \sum_{\nu=1}^{\infty} |\rho(\nu)| &< \infty \quad \longrightarrow \text{short memory} \\ \sum_{\nu=1}^{\infty} |\rho(\nu)| &= \infty \quad \longrightarrow \text{long memory.} \end{aligned} \tag{4.3}$$

Let $f_n(\lambda) = f(\lambda)/R(0)$, $0 \leq \lambda \leq 1$, be the normalized spectral density (i.e. $\int_0^1 f_n(\lambda)d\lambda = 1$). Then the spectral density can be defined as the Fourier transform of $\rho(\nu)$ by

$$f_n(\lambda) = \sum_{\nu=-\infty}^{\infty} \rho(\nu) \exp(2\pi i \lambda \nu) \tag{4.4}$$

and memory type can also be classified in terms of the dynamic range. The dynamic range and its memory types are defined by

$$\begin{aligned} \text{dynamic range} &= \log \left(\frac{\max f_n(\lambda)}{\min f_n(\lambda)} \right) \\ &= \log [\max f_n(\lambda)] - \log [\min f_n(\lambda)] \\ &= \max \log [f_n(\lambda)] - \min \log [f_n(\lambda)] \end{aligned} \tag{4.5}$$

and

$$\begin{aligned}
\text{dynamic range} = 0 &\longrightarrow \text{no memory} \\
0 < \text{dynamic range} < \infty &\longrightarrow \text{short memory} \\
\text{dynamic range} = \infty &\longrightarrow \text{long memory,}
\end{aligned} \tag{4.6}$$

see Parzen (1981). This definition can be extended to accommodate the case of a stationary time series whose correlation function $\rho(\nu)$ is not summable, see Parzen (1981).

Another diagnostic considered in this work is the prediction variance horizon, with horizon equal to one, originally defined by Parzen (1981). In modelling a stationary time series using the EXP model this turns out to be an especially useful diagnostic of memory type. In order to define the prediction variance horizon, $\text{PVH}(h)$, probability concepts are defined using prediction theory. Consider the mean zero stationary time series $Y(t)$ and let $Y^{\mu,m}(t)$ be the finite memory one step ahead predictor

$$Y^{\mu,m}(t) = \text{E}(Y(t)|Y(t-1), \dots, Y(t-m)). \tag{4.7}$$

The infinite memory one step ahead predictor $Y^\mu(t)$ is defined analogously by

$$Y^\mu(t) = \text{E}(Y(t)|Y(t-1), Y(t-2), \dots). \tag{4.8}$$

Similarly, the infinite memory predictor with horizon h is defined by

$$Y^\mu(t+h|t) = \text{E}(Y(t+h)|Y(t), Y(t-1), \dots). \tag{4.9}$$

Their prediction errors are denoted

$$Y^{\nu,m}(t) = Y(t) - Y^{\mu,m}(t), \tag{4.10}$$

$$Y^\nu(t) = Y(t) - Y^\mu(t) \tag{4.11}$$

and

$$Y^\nu(t+h|t) = Y(t+h) - Y^\mu(t+h|t) \quad (4.12)$$

respectively. $Y^\nu(t)$ are white noise innovations and form an orthogonal basis for the Hilbert space spanned by $Y(t)$, for all t . Assuming $Y(t)$ is nondeterministic then, by Wold's decomposition, $Y(t)$ has a one sided MA(∞) representation given by

$$Y(t) = \psi_0 + \psi_1 Y(t-1) + \psi_2 Y(t-2) + \dots \quad (4.13)$$

where $\psi_0 \equiv 1$. An equivalent condition for the existence of a one sided MA(∞) representation of $Y(t)$ is

$$\int_0^1 \log f(\lambda) d\lambda > -\infty. \quad (4.14)$$

Moreover, the variances of $Y(t)$ and $Y^\nu(t)$ satisfy the following relationship,

$$E(|Y(t)|^2) = E(|Y^\nu(t)|^2) (1 + \psi_1^2 + \psi_2^2 + \dots). \quad (4.15)$$

Next, define the normalized prediction error as the innovation variance (one step ahead mean square error of prediction) divided by the variance of $Y(t)$. Parzen (2002) defines the prediction variance horizon

$$\text{PVH}(h) = \frac{E(|Y^\nu(t+h|t)|^2)}{E(|Y(t)|^2)} \quad (4.16)$$

and

$$\text{PVH}(1) = \sigma_\infty^2 = (1 + \psi_1^2 + \psi_2^2 + \dots)^{-1}. \quad (4.17)$$

It should be noted that the prediction variance horizon, originally defined by Parzen (1981), satisfies the relationship

$$\text{PVH}(h)_{2002} = 1 - \text{PVH}(h)_{1981}. \quad (4.18)$$

Computational formulas for $PVH(h)$ are given by

$$PVH(h) = PVH(1) \left(\sum_{j=0}^{h-1} \psi_j^2 \right) \quad (4.19)$$

$$= \frac{\left(\sum_{j=0}^{h-1} \psi_j^2 \right)}{\left(\sum_{j=0}^{\infty} \psi_j^2 \right)}. \quad (4.20)$$

Parzen (2002) gives an intuitive diagnostic for memory type in terms of $PVH(1)$:

$$\begin{aligned} PVH(1) > .95 & \longrightarrow \text{no memory} \\ .05 < PVH(1) < .95 & \longrightarrow \text{short memory} \\ PVH(1) < .05 & \longrightarrow \text{long memory.} \end{aligned} \quad (4.21)$$

A natural way of relating this diagnostic to the EXP model is through the cepstral correlations (coefficients). Recall, the cepstral coefficients are defined by

$$c(h) = \int_0^1 \log f(\lambda) \cos(2\pi ih\lambda) d\lambda \quad (4.22)$$

and $c(h) \equiv \theta_h$, where θ_h are the parameters of the EXP model. Assuming $f(\lambda)$ is a normalized spectral density, $PVH(1)$ can be alternatively defined using $c(0)$,

$$\begin{aligned} PVH(1) &= (1 + \psi_1^2 + \psi_2^2 + \dots)^{-1} \\ &= \sigma_{\infty}^2 \\ &= \exp \left(\int_0^1 \log f(\lambda) d\lambda \right) \\ &= \exp [c(0)]. \end{aligned} \quad (4.23)$$

In practice, memory identification proceeds by using plug in estimators of theoretical diagnostics. One empirical diagnostic of memory type is the sample spectral distribution function $F^{\sim}(\lambda)$. In order to define $F^{\sim}(\lambda)$, let $\{Y(t), t = 1, \dots, N\}$ be an observed time series with mean zero. The sample spectral density (periodogram) is

defined by

$$f^{\sim}(\lambda) = \frac{1}{N} \left| \sum_{t=1}^N Y(t) \exp(2\pi i \lambda t) \right|^2, \quad 0 \leq \lambda \leq 1. \quad (4.24)$$

Similarly the normalized sample spectral density (periodogram) is defined by

$$f_n^{\sim}(\lambda) = \frac{\left| \sum_{t=1}^N Y(t) \exp(2\pi i \lambda t) \right|^2}{\sum_{t=1}^N Y^2(t)}, \quad 0 \leq \lambda \leq 1. \quad (4.25)$$

The sample correlation function $\hat{\rho}(\nu)$, $\nu = 0, 1, \dots, N-1$ is defined by

$$\hat{\rho}(\nu) = \frac{\sum_{t=1}^{N-|\nu|} Y(t)Y(t+\nu)}{\sum_{t=1}^N Y^2(t)} \quad (4.26)$$

$$= \int_0^1 f_n^{\sim}(\lambda) \exp(2\pi i \lambda \nu) d\lambda. \quad (4.27)$$

Finally, the sample spectral distribution is defined by

$$F^{\sim}(\lambda) = 2 \int_0^{\lambda} f_n^{\sim}(\lambda') d\lambda', \quad 0 \leq \lambda \leq .5. \quad (4.28)$$

In practice, we use a definition for the sample spectral distribution function, with a more convenient computational formula,

$$\begin{aligned} F^{\sim}(\lambda_k) &= \frac{\sum_{j=1}^k f_n^{\sim}(\lambda_j)}{\sum_{j=1}^M f_n^{\sim}(\lambda_j)} \\ &= \frac{\sum_{j=1}^k f^{\sim}(\lambda_j)}{\sum_{j=1}^M f^{\sim}(\lambda_j)}, \quad k = 1, \dots, M \end{aligned} \quad (4.29)$$

where $M = [N/2] + 1$ and λ_j is the j^{th} Fourier frequency.

One method of determining memory type, using the sample distribution function, is graphically. Parzen (1981) suggests

$$\begin{array}{lll} F^{\sim} & \text{uniform} & \longrightarrow \text{no memory} \\ F^{\sim} & \text{otherwise} & \longrightarrow \text{short memory} \\ F^{\sim} & \text{has sharp jumps} & \longrightarrow \text{long memory.} \end{array} \quad (4.30)$$

4.2 Identifying Time Series Memory Type

The approach we take, in subsequent chapters, for modelling an observed time series begins by identifying the memory type. To identify memory type of a time series will involve the following steps:

1. Estimate the log spectral density, $\log f(\lambda)$, see Chapter VI.
2. Compute PVH(1), see expression (4.23).
3. Compute and plot the spectral distribution function, $F^\sim(\lambda_k)$, $k = 1, \dots, M$, as defined in expression (4.29).
4. Compute the sample dynamic range. The sample dynamic range is defined using expression (4.5) and $f_n^\sim(\lambda)$. Note, in practice, short memory processes usually have sample dynamic range $\ll 15$.

Once memory type is determined, we employ the following model fitting strategy. First, if Y is short memory we fit an appropriate EXP model (see Chapters VI and VII). If Y is long memory transform it (parsimoniously) to be barely short memory Y^s (see Parzen 1982), we then fit an appropriate EXP model. In this work the majority of methods developed apply only to stationary short memory time series. Thus, unless otherwise stated, it is assumed that Y is a stationary, short memory time series.

CHAPTER V

SIMULATING DATA FROM EXPONENTIAL MODELS

5.1 Introduction

Simulating a process from an EXP(m) model requires a method that does not make explicit use of residuals. In this chapter three applicable methods are described. The first method is an exact frequency domain method due to Davies and Harte (1987). The second is a well known exact time domain method, described by Percival (1992), and the third method is an indirect method we develop, based on Pourahmadi's formula. The benefits of our method are that it is easy to implement, extremely accurate, and requires less computing time.

5.2 Davies-Hart Method

The Davies-Hart method is based on the fast Fourier transform and simulates a stationary Gaussian time series of length N with autocovariances $R(0), R(1), \dots, R(N)$. The following is the Davies-Hart algorithm and is adapted from Beran (1994).

1. Define

$$\lambda_k = \frac{(k-1)}{2N-2} \quad (5.1)$$

for $k = 1, \dots, 2N-2$, and the finite Fourier transform g_k of the sequence $R(0), R(1), \dots, R(N-2), R(N-1), R(N-2), \dots, R(1)$,

$$g_k = \sum_{j=1}^{N-1} R(j-1)e^{2\pi i(j-1)\lambda_k} + \sum_{j=N}^{2N-2} R(2N-j-1)e^{2\pi i(j-1)\lambda_k} \quad (5.2)$$

for $k = 1, \dots, 2N-2$.

2. Check that $g_k > 0$ for all $k = 1, \dots, 2N-2$.

3. Simulate two independent series of zero mean normal random variables, say U_1, U_2, \dots, U_N and V_1, V_2, \dots, V_{N-1} such that

$$\text{var}(U_1) = \text{var}(U_N) = 2 \quad (5.3)$$

and, for $k \neq 1, N$,

$$\text{var}(U_k) = \text{var}(V_k) = 1. \quad (5.4)$$

Define $V_1 = V_N = 0$ and complex random variables Z_k by

$$Z_k = U_k + iV_k, \quad k = 1, \dots, N \quad (5.5)$$

and

$$Z_k = U_{2N-k} + iV_{2N-k} \quad (5.6)$$

for $k = N + 1, \dots, 2N - 2$.

4. For $t = 1, \dots, N$, define

$$Y_t = \frac{1}{2\sqrt{N-1}} \sum_{k=1}^{2N-2} \sqrt{g_k} e^{2\pi i(t-1)\lambda_k} Z_k. \quad (5.7)$$

Theoretically, the series Y_t has the desired distribution.

In using this method to simulate from the EXP(m) model, the autocovariances need to be estimated. The most frequently used method of estimating the autocovariances is via the spectral density using numerical integration. That is

$$\begin{aligned} R(\nu) &= \int_0^1 f(\lambda) \exp(2\pi i\nu\lambda) d\lambda \\ &= \int_0^1 \exp\left(\theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k\lambda)\right) \exp(2\pi i\nu\lambda) d\lambda. \end{aligned} \quad (5.8)$$

Thus, in practice using this method can be time consuming and error prone. Additionally, since the process being simulated is short memory, one approach taken in

practice is to only calculate the first M autocovariances, $M \ll N$ and set the remaining $N - M$ autocovariances to zero. A second, perhaps more accurate, method of estimating the autocovariance sequence is described in Hurvich (2002). Even though the accuracy of the Davies-Hart algorithm can be improved using this method of estimation, the algorithm is still more computationally intensive than the indirect method that we propose.

5.3 Exact Time Domain Method

The following method is a well known theoretically exact time domain method and can be found in Percival (1992). Let $\{R(\nu)\}$ be the autocovariance sequence out to lag $N - 1$. Then samples $\{Y_t\}$ can be generated using the following algorithm:

1. Generate Z_0, Z_1, \dots, Z_{N-1} where $Z_i \sim i.i.d. N(0, 1)$ random variables.
2. Set $Y_0 = \sigma_0 Z_0$, where $\sigma_0^2 = R(0)$.
3. Calculate

$$Y_t = \sum_{j=1}^t \phi_{j,t} Y_{t-j} + Z_t \sigma_t, \quad t = 1, \dots, N - 1 \quad (5.9)$$

where

$$\phi_{t,t} = \frac{R(t) - \sum_{j=1}^{t-1} \phi_{j,t-1} R(t-j)}{\sigma_{t-1}^2} \quad (5.10)$$

$$\phi_{j,t} = \phi_{j,t-1} - \phi_{t,t} \phi_{t-j,t-1}, \quad 1 \leq j \leq t - 1 \quad (5.11)$$

$$\sigma_t^2 = \sigma_{t-1}^2 (1 - \phi_{t,t}). \quad (5.12)$$

If $t = 1$, define the sum on the right hand side of equation (5.9) equal to zero.

As in the exact frequency domain method, in practice this algorithm can be error prone since the autocovariances, $R(\nu)$, need to be estimated using numerical integration. Even though the accuracy of the estimated autocovariance sequence can be

improved using the method described in Hurvich (2002), this simulation method uses the estimated $R(\nu)$'s to calculate $\phi_{t,t}$ which are then reused to calculate $\phi_{j,t}$. Due to the nature of the iterative scheme in this algorithm, any numerical errors incurred from estimating the $R(\nu)$'s are compounded. Thus, this method can be inaccurate and computationally intensive.

5.4 Indirect Method

The method that we propose is based on Pourahmadi's formula and makes use of the MA(∞) representation for the EXP(m) process. Let

$$\log f_{\theta,m}(\lambda) = \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(k2\pi\lambda) \quad (5.13)$$

be the log EXP(m) representation associated with the process $\{Y_t\}$ and let

$$Y(t) = \psi_0 Y^\nu(t) + \psi_1 Y^\nu(t-1) + \psi_2 Y^\nu(t-2) + \dots \quad (5.14)$$

be its associated MA(∞) representation ($\psi_0 \equiv 1$). Recall $Y^\nu(t)$, for all t , are called the innovations where $Y^\nu(t)$ are mean zero Gaussian white noise. Recall, Pourahmadi's formula enables direct computation of ψ_h from θ_k :

$$\begin{aligned} \psi_0 &\equiv 1 \\ \psi_h &= \frac{1}{h} \sum_{k=1}^h k \theta_k \psi_{h-k} \quad h = 1, 2, \dots \end{aligned} \quad (5.15)$$

In order to simulate samples $\{Y_t\}$ from the EXP(m) model we use the following algorithm:

1. Given the EXP(m) model calculate $\psi_0, \psi_1, \dots, \psi_q$ for q sufficiently large (i.e. $q \geq 1000$), using Pourahmadi's formula.
2. For $k = 0, 1, \dots, q$ form

$$SS_{sim}(k) = \frac{SS(k)}{SS(q)} = \frac{\sum_{h=0}^k \psi_h^2}{\sum_{h=0}^q \psi_h^2}. \quad (5.16)$$

3. Find the largest value of k such that $SS_{sim}(k) < 1 - \epsilon$ (i.e. $SS_{sim}(k) = .9999999$).

4. Form

$$Y(t) = \psi_0 Y^\nu(t) + \psi_1 Y^\nu(t-1) + \cdots + \psi_k Y^\nu(t-k) \quad (5.17)$$

then

$$\begin{aligned} \log f_{\theta,m}(\lambda) &= \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(k2\pi\lambda) \\ &\approx \log \left(\sigma^2 \left| \sum_{h=0}^k \psi_h \exp(2\pi i h \lambda) \right|^2 \right) \end{aligned} \quad (5.18)$$

where $\sigma^2 = \exp(\theta_0)$ and $\psi_0 \equiv 1$.

5. Simulate the process $\{Y_t\}$ from its k^{th} order truncated MA(∞) representation using $N(0, \sigma^2 = \exp(\theta_0))$ innovations.

This simulation procedure can be made virtually exact in the sense that the integrated squared error between the spectral density of the EXP(m) process and the spectral density of the process being simulated from can always be made arbitrarily small. Consider the MA(∞) representation associated with $f_{\theta,m}(\lambda)$, the spectral density of the EXP(m) process, given by expression (5.14). It follows that

$$\begin{aligned} f_{\theta,m}(\lambda) &= \exp \left(\theta_0 + 2 \sum_{k=1}^m \theta_k \cos(k2\pi\lambda) \right) \\ &= \sigma^2 \left| \sum_{h=0}^{\infty} \psi_h \exp(2\pi i h \lambda) \right|^2 \\ &= \lim_{k \rightarrow \infty} \sigma^2 \left| \sum_{h=0}^k \psi_h \exp(2\pi i h \lambda) \right|^2 \\ &= f_{MA(\infty)}(\lambda). \end{aligned} \quad (5.19)$$

Moreover, the integrated squared error between $f_{\theta,m}(\lambda)$ and $f_{MA(k)}(\lambda)$ is

$$\text{ISE}[f_{\theta,m}(\lambda), f_{MA(k)}(\lambda)] = \int_0^1 [f_{\theta,m}(\lambda) - f_{MA(k)}(\lambda)]^2 d\lambda; \quad (5.20)$$

thus

$$\lim_{k \rightarrow \infty} \text{ISE}[f_{\theta, m}(\lambda), f_{MA(k)}(\lambda)] = 0. \quad (5.21)$$

Hence, in practice this method is not error prone like the exact time and frequency domain methods previously described. Additionally, this method is easy to implement using any software capable of looping and generating normal random variables. Note, this method can also be extended to allow simulation of long memory processes through the use of fractional differencing via the FEXP model defined in expression (4.1). However, when simulating FEXP processes the method of Hurvich (2002) may be preferable and should be considered.

5.5 Example of Indirect Simulation Method

Consider the EXP(4) model given by

$$\begin{aligned} g_{EXP}(\lambda) &= \log f_{EXP}(\lambda) \\ &= -.05 + 2[-.90 \cos(2\pi\lambda) + .40 \cos(2\pi 2\lambda) + .30 \cos(2\pi 3\lambda) \\ &\quad + .15 \cos(2\pi 4\lambda)]. \end{aligned} \quad (5.22)$$

In order to simulate a time series from this model using the indirect method, the first step is to compute the coefficients of the MA(q) representation using Pourahmadi's formula ($q = 1000$). For $k = 0, \dots, q$ calculate $SS_{sim}(k)$, where $SS_{sim}(k)$ is defined in expression (5.16).

The equivalent MA(∞) representation associated with the EXP(4) can be well approximated by the MA(12) with the coefficients found in Table 1. The log spectral density for this model is

$$\begin{aligned} g_{MA}(\lambda) &= \log f_{MA}(\lambda) \\ &= \log (\sigma^2 |1 + \psi_1 \exp(2\pi i\lambda) + \dots + \psi_{12} \exp(2\pi i 12\lambda)|^2). \end{aligned} \quad (5.23)$$

Table 1. $MA(k)$, $k = 12$, representation of the $EXP(4)$ model and the associated $SS_{sim}(k)$. Note that $\psi_0 \equiv 1$.

k	ψ_k	$SS_{sim}(k)$	k	ψ_k	$SS_{sim}(k)$
1	-.900	.7183015	7	-.008	.9998606
2	.805	.9754713	8	.018	.9999835
3	-.182	.9885445	9	.004	.9999893
4	.149	.9973950	10	.005	.9999985
5	-.019	.9975386	11	.001	.9999991
6	.076	.9998332	12	.001	.9999999

One measure of the “closeness” between the two spectral densities $f_{EXP}(\lambda)$ and $f_{MA}(\lambda)$ is the squared error

$$SE(f_{EXP}, f_{MA}) = [f_{EXP}(\lambda) - f_{MA}(\lambda)]^2, \quad (5.24)$$

see Figure 1. Additionally, this measure can be quantified using the integrated squared error (ISE), where

$$ISE[f_{EXP}(\lambda), f_{MA}(\lambda)] = \int_0^1 [f_{EXP}(\lambda) - f_{MA}(\lambda)]^2 d\lambda. \quad (5.25)$$

In this case $ISE[f_{EXP}(\lambda), f_{MA}(\lambda)] \approx 5.73 \times 10^{-7}$. Moreover, the “validity” of this simulation is confirmed by a graph of the log EXP spectral density with the corresponding truncated log MA spectral density drawn on the same axis, see Figure 2. Lastly, a plot of $\log f(\lambda)$ and $\log \tilde{f}(\lambda) + \gamma$ provides another means of visual confirmation that the process is being simulated from the correct model, see Figure 3.

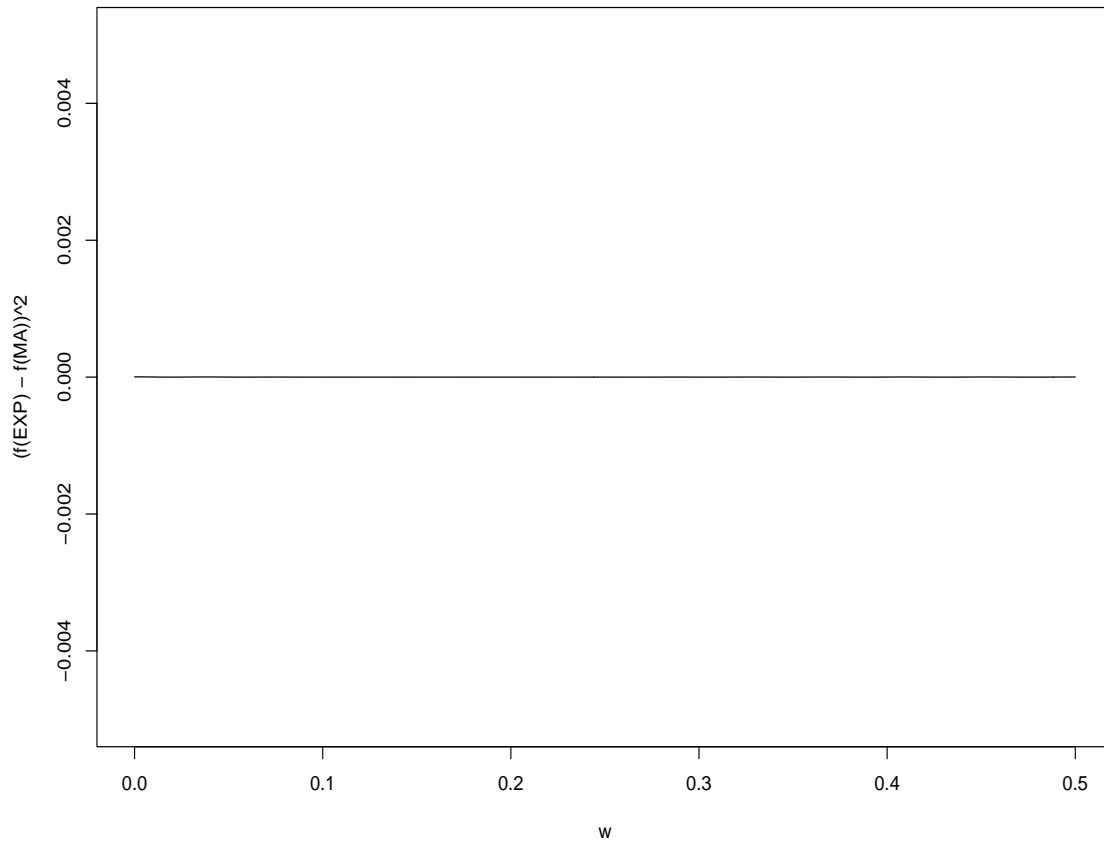


Figure 1. Plot of $SE(f_{EXP}, f_{MA})$. This plot shows the squared error between the EXP spectral density and its corresponding truncated MA spectral density. This plot is approximately equal to zero for all $\lambda \in [0, 1]$. Note only $0 \leq \lambda \leq .5$ is graphed due to the symmetry of the spectral density on $[0, 1]$.

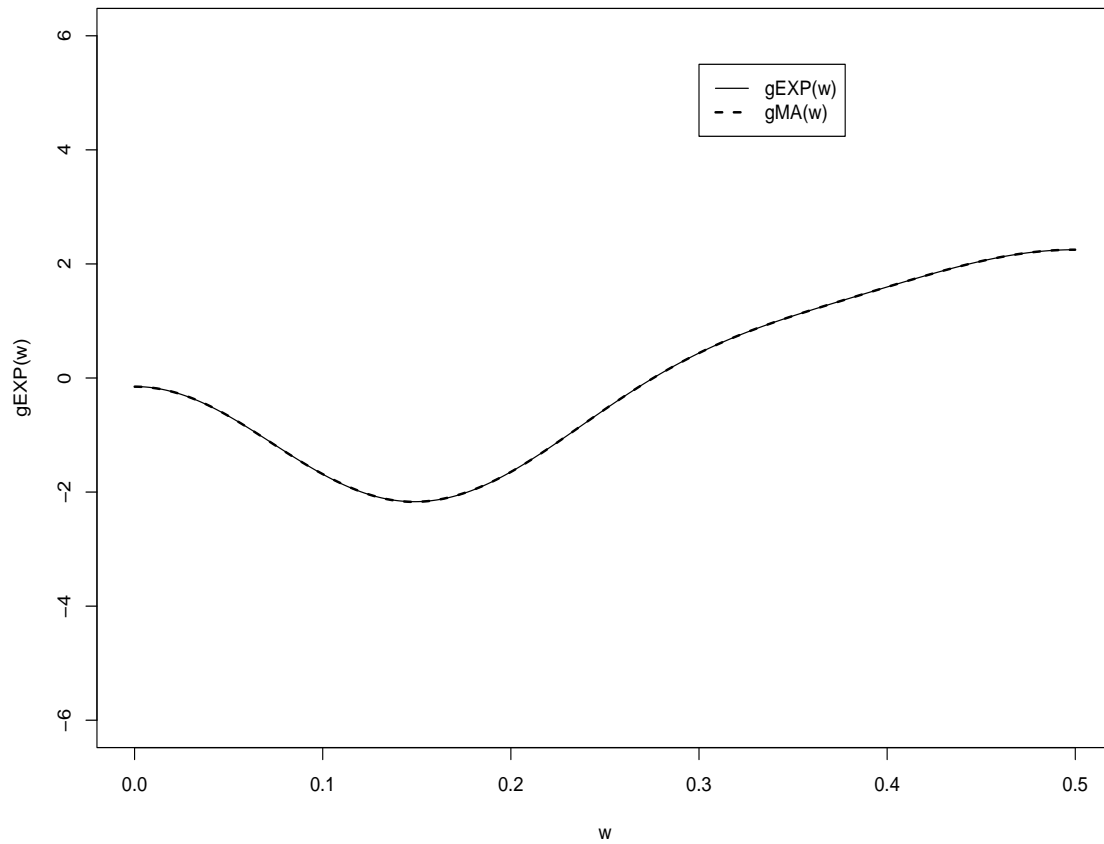


Figure 2. An example of the log EXP spectral density and its corresponding truncated log MA spectral density. The log EXP(4) spectral density given by expression (5.22) and the log spectral density of the truncated MA model given by expression (5.23) are virtually indistinguishable.

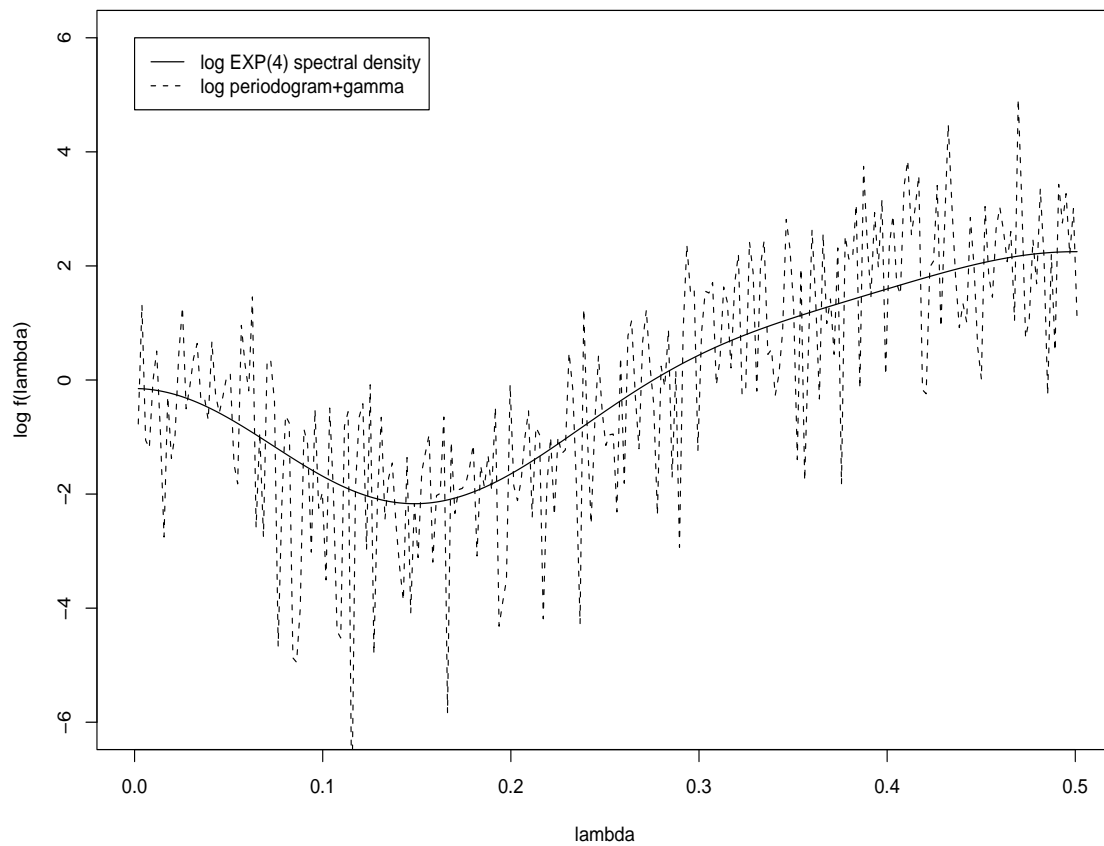


Figure 3. Plot of $\log f(\lambda)$ and $\log \tilde{f}(\lambda) + \gamma$. This plot provides a visual confirmation that the process is being simulated from the correct model. This conclusion is based on the fact that the log sample spectral density plus a known constant overlays the true log spectral density, as expected. Note that $\gamma \approx .57721$ is Euler's constant.

CHAPTER VI

METHODS FOR SHORT MEMORY EXP PARAMETER ESTIMATION

6.1 Introduction

Several methods of parameter estimation exist for short memory EXP models. Among these methods are maximum likelihood, minimization of Renyi information divergence, and log periodogram regression. In addition to these methods, we propose a new fully nonparametric approach based on wavelets.

The main emphasis of this chapter will be on introducing the new parameter estimation method based on wavelets. However, for completeness, the methods based on minimization of Renyi information divergence and log periodogram regression will also be described. The maximum likelihood method, Bloomfield (1973), was fully discussed in Chapter II. Note, the algorithm presented in that chapter can either be implemented directly or equivalently through the use of iteratively reweighted least squares with the aid of a standard software regression program that incorporates the Nelder-Wedderburn algorithm for generalized linear models, see Cameron and Turner (1987).

6.2 Parameter Estimation by Minimum Renyi Information Divergence

Modelling an observed time series $Y(t)$ is equivalent to finding a filter which transforms the time series $Y(t)$ to a no memory (white noise) innovation series $Y^\nu(t)$. Now, let the parameters of the model be denoted θ , and $g_\theta(\lambda)$ be the square modulus of the frequency transfer function of the whitening filter. The spectral density is given by

$$f_\theta(\lambda) = \frac{\sigma^2}{g_\theta(\lambda)} \quad (6.1)$$

where $\sigma^2 = \text{var}[Y^\nu(t)]$, the innovation variance, and $\log \sigma^2 = \int_0^1 \log f_\theta(\lambda) d\lambda$. Additionally, assume $\text{var}[Y(t)] = 1$, i.e. $f^\sim(\lambda) = f_n^\sim(\lambda)$. From the observed time series $Y(t)$, compute the normalized (to integrate to one) sample spectral density $f^\sim(\lambda)$. Parameter estimates can be found by minimizing the formula for Renyi information divergence

$$\text{IR}_{-1}[f^\sim(\lambda)/f_\theta(\lambda)] = \int_0^1 \left(\frac{f^\sim(\lambda)}{f_\theta(\lambda)} - 1 - \log \left(\frac{f^\sim(\lambda)}{f_\theta(\lambda)} \right) \right) d\lambda. \quad (6.2)$$

See Parzen (1992) and Parzen (1993) for a complete discussion of Renyi information. Minimization of expression (6.2) is equivalent to minimizing

$$\log \sigma^2 + \frac{1}{\sigma^2} \int_0^1 f^\sim(\lambda) g_\theta(\lambda) d\lambda. \quad (6.3)$$

Thus, estimators θ^\wedge are obtained by minimizing with respect to θ

$$\sigma_\theta^2 = \int_0^1 f^\sim(\lambda) g_\theta(\lambda) d\lambda. \quad (6.4)$$

Additionally, σ^2 can be estimated from this minimum value by

$$\hat{\sigma}^2 = \sigma_{\theta^\wedge}^2 = \int_0^1 f^\sim(\lambda) g_{\theta^\wedge}(\lambda) d\lambda. \quad (6.5)$$

Therefore, the true spectral density, $f(\lambda)$, can be estimated by

$$f_{\theta^\wedge}(\lambda) = \frac{\hat{\sigma}^2}{g_{\theta^\wedge}(\lambda)}. \quad (6.6)$$

The parameter estimates obtained through minimizing Renyi information divergence are asymptotically equivalent to maximum likelihood estimators. In the case of the EXP model, minimizing $\text{IR}_{-1}[f^\sim(\lambda)/f_\theta(\lambda)]$ from expression (6.2) is equivalent to estimating θ^\wedge by minimizing

$$\sigma_\theta^2 = \int_0^1 \frac{f^\sim(\lambda)}{f_{\theta,m}(\lambda)} d\lambda \quad (6.7)$$

where

$$f_{\theta,m}(\lambda) = \exp \left(\theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k\lambda) \right). \quad (6.8)$$

Therefore, the parameters can be estimated using the modified Newton-Raphson procedure discussed in Chapter II. Note, in Chapter II the algorithm was defined for spectral densities, $f(\lambda)$, such that $-\pi \leq \lambda \leq \pi$. Hence, using the algorithm in this context requires slight modification.

Both maximum likelihood and minimum Renyi information divergence parameter estimation can suffer from several problems. The first problem is that in both of the estimation schemes the parameter estimation depends on correct model specification. The second problem is that the model is being fit in order to minimize an information theoretic distance between f^\sim , the sample spectral density, and the estimated model, f_{θ^\wedge} . However, in short memory situations where the true spectral density has large dynamic range or rapid variations, f^\sim may suffer from leakage (bias) (see Section 6.4). For example, this can occur when the spectral density is close to being the spectral density of a long memory time series but is still considered short memory. Thus, fitting the model to f^\sim may not produce a whitening model of the spectrum.

6.3 Parameter Estimation by Log Periodogram Regression

Recall the definition of the EXP(m) model given by expression (2.15)

$$\log f_{\theta,m}(\lambda) = \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k\lambda). \quad (6.9)$$

This model can be estimated using an ordinary least squares regression and the sample spectral density, $f^\sim(\lambda)$. First, it is well known that

$$f^\sim(\lambda) \stackrel{d}{=} f(\lambda) \frac{\chi_2^2}{2}, \quad 0 < \lambda < .5. \quad (6.10)$$

Therefore, for $0 < \lambda < .5$, it follows that

$$\begin{aligned} \log f^{\sim}(\lambda) &\stackrel{d}{=} \log \left(f(\lambda) \frac{\chi_2^2}{2} \right) \\ &= \log \left(\frac{f(\lambda)}{2} \right) + \log(\chi_2^2). \end{aligned} \quad (6.11)$$

Thus, using the results of Bartlett and Kendall (1946)

$$\text{Var} [\log f^{\sim}(\lambda)] = \frac{\pi^2}{6} \quad (6.12)$$

and

$$\text{E} [\log f^{\sim}(\lambda)] = \log f(\lambda) - \gamma, \quad (6.13)$$

where $\gamma \approx .57721$ is Euler's constant. Additionally, the random variable

$$\epsilon(\lambda) \equiv \log \left(\frac{f^{\sim}(\lambda)}{f(\lambda)} \right) + \gamma \quad (6.14)$$

has mean zero and variance $\sigma_\epsilon^2 = \pi^2/6$, since

$$\begin{aligned} \epsilon(\lambda) &\equiv \log \left(\frac{f^{\sim}(\lambda)}{f(\lambda)} \right) + \gamma \\ &= \log f^{\sim}(\lambda) - \log f(\lambda) + \gamma \\ &\stackrel{d}{=} \log \left(f(\lambda) \frac{\chi_2^2}{2} \right) - \log f(\lambda) + \gamma \\ &= \log(\chi_2^2) + \gamma - \log(2). \end{aligned} \quad (6.15)$$

Now, for $0 < \lambda < .5$, let

$$g(\lambda) \equiv \log f^{\sim}(\lambda) + \gamma \quad (6.16)$$

then

$$g(\lambda) = \log f(\lambda) + \epsilon(\lambda) \quad (6.17)$$

where $\epsilon(\lambda) \sim (0, \sigma_\epsilon^2) = (0, \pi^2/6)$. Thus, $g(\lambda)$ is equal to the true log spectral density plus noise. Next, consider $f^{\sim}(\lambda_j)$ where $\lambda_j = j/N$ are the Fourier frequencies, N is the sample size and $j = 1, \dots, \lfloor (N-1)/2 \rfloor$. Under the assumption that the $\epsilon(\lambda_j) \equiv \epsilon_j$

are independent and identically distributed, the parameters of the EXP(m) model can be estimated using the following expression

$$\begin{aligned} g(\lambda_j) &= \log f^\sim(\lambda_j) + \gamma \\ &= \theta_0 + 2 \sum_{k=1}^m \theta_k \cos(2\pi k \lambda_j) + \epsilon_j \end{aligned} \quad (6.18)$$

where θ_k are estimated using the ordinary least squares estimators, $\hat{\theta}_k$, in the multiple linear regression of $\{\log f^\sim(\lambda_j)\}_{j=1}^{\lfloor (N-1)/2 \rfloor} + \gamma$ on 1 and the cosine functions $\{2 \cos(2\pi k \lambda_j)\}_{j=1}^{\lfloor (N-1)/2 \rfloor}$. In order to implement this procedure we rewrite expression (6.18) in matrix notation

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} g(\lambda_1) \\ g(\lambda_2) \\ \vdots \\ g(\lambda_{\lfloor (N-1)/2 \rfloor}) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 \cos(2\pi \lambda_1) & \cdots & 2 \cos(2\pi m \lambda_1) \\ 1 & 2 \cos(2\pi \lambda_2) & \cdots & 2 \cos(2\pi m \lambda_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2 \cos(2\pi \lambda_{\lfloor (N-1)/2 \rfloor}) & \cdots & 2 \cos(2\pi m \lambda_{\lfloor (N-1)/2 \rfloor}) \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{pmatrix} \\ &+ \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{\lfloor (N-1)/2 \rfloor} \end{pmatrix} \\ &= \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}. \end{aligned} \quad (6.19)$$

Using any standard regression software, or any program that performs matrix multiplication, $\hat{\boldsymbol{\theta}}$ can be computed from the normal equations

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (6.20)$$

This method has both advantages and disadvantages. The main disadvantage of this method is the same as the previous method. The $\hat{\theta}_k$ are being estimated using $g(\lambda_j) = \log f^\sim(\lambda_j) + \gamma$ which could suffer from (bias) leakage, see Section 6.4. If this is the case, the estimated model may not provide a whitening transformation of the spectral density. On the other hand, one advantage of this method is that regression diagnostics such as Mallows's C_p can be used to aid in order selection.

6.4 Multitaper Spectral Density Estimation

6.4.1 Bias Reduction Through Tapering

Suppose $\{Y(t)\}$ is a discrete parameter real valued stationary time series with mean zero and continuous spectral density $f(\lambda)$. Let $\{R(\nu) : \nu \in Z\}$ be the autocovariance function associated with the time series $\{Y(t)\}$. If $\sum_{\nu=-\infty}^{\infty} |R(\nu)| < \infty$ then

$$f(\lambda) = \sum_{\nu=-\infty}^{\infty} R(\nu) \cos(2\pi\lambda\nu), \quad \lambda \in [0, 1]. \quad (6.21)$$

Given an observed time series $Y(1), Y(2), \dots, Y(N)$, usually, estimation of $f(\lambda)$ begins with the sample spectral density (periodogram) $f^\sim(\lambda)$. Let

$$\hat{R}(\nu) = \frac{1}{N} \sum_{t=1}^{N-|\nu|} Y(t)Y(t+|\nu|) \quad (6.22)$$

be the sample autocovariance function then

$$\begin{aligned} f^\sim(\lambda) &= \frac{1}{N} \left| \sum_{t=1}^N Y(t) \exp(-2\pi i\lambda t) \right|^2 \\ &= \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^N Y(j)Y(k) \exp[-2\pi i\lambda(k-j)] \\ &= \frac{1}{N} \sum_{\nu=-(N-1)}^{N-1} \sum_{t=1}^{N-|\nu|} Y(t)Y(t+|\nu|) \exp(-2\pi i\lambda\nu) \\ &= \sum_{\nu=-(N-1)}^{(N-1)} \hat{R}(\nu) \exp(-2\pi i\lambda\nu) \end{aligned}$$

$$= \sum_{\nu=-(N-1)}^{(N-1)} \widehat{R}(\nu) \cos(2\pi\lambda\nu). \quad (6.23)$$

Moreover,

$$\begin{aligned} \mathbb{E} [\widehat{R}(\nu)] &= \mathbb{E} \left(\frac{1}{N} \sum_{t=1}^{N-|\nu|} Y(t)Y(t+|\nu|) \right) = \frac{1}{N} \sum_{t=1}^{N-|\nu|} R(\nu) \\ &= \left(1 - \frac{|\nu|}{N} \right) R(\nu). \end{aligned} \quad (6.24)$$

Therefore, it follows that

$$\begin{aligned} \mathbb{E} [f^{\sim}(\lambda)] &= \sum_{\nu=-(N-1)}^{(N-1)} \mathbb{E} [\widehat{R}(\nu)] \cos(2\pi\lambda\nu) \\ &= \sum_{\nu=-(N-1)}^{(N-1)} \left(1 - \frac{|\nu|}{N} \right) R(\nu) \cos(2\pi\lambda\nu). \end{aligned} \quad (6.25)$$

Thus, $f^{\sim}(\lambda)$ is asymptotically unbiased. However, this does not mean that for any particular N that $\text{bias}[f^{\sim}(\lambda)]$ is necessarily small. Additionally,

$$\begin{aligned} &\sum_{\nu=-(N-1)}^{(N-1)} \left(1 - \frac{|\nu|}{N} \right) R(\nu) \cos(2\pi\lambda\nu) \\ &= \sum_{\nu=-(N-1)}^{(N-1)} \left(1 - \frac{|\nu|}{N} \right) \left(\int_0^1 f(\omega) \cos(2\pi\nu\omega) d\omega \right) \cos(2\pi\lambda\nu) \\ &= \int_0^1 \sum_{\nu=-(N-1)}^{(N-1)} \left(1 - \frac{|\nu|}{N} \right) \cos[2\pi\nu(\lambda - \omega)] f(\omega) d\omega \\ &= \int_0^1 F_N(\lambda - \omega) f(\omega) d\omega \end{aligned} \quad (6.26)$$

where

$$\begin{aligned} F_N(\omega) &= \sum_{\nu=-(N-1)}^{N-1} \left(1 - \frac{|\nu|}{N} \right) \cos(2\pi\nu\omega) \\ &= \frac{1}{N} \left(\frac{\sin(\pi N\omega)}{\sin(\pi\omega)} \right)^2. \end{aligned} \quad (6.27)$$

F_N is known as Fejér's kernel. Thus, the expected value of $f^{\sim}(\lambda)$ is the convolution of Fejér's kernel and the spectral density. Hence, the bias in $f^{\sim}(\lambda)$ can be attributed to the sidelobes of the Fejér kernel. This type of bias usually occurs in processes having spectral densities with rapid variations and/or large dynamic range as defined by expression (4.5). Therefore, for processes with large peaks and troughs in the spectral density bias can be substantial. This bias is known as leakage and can be attributed to the transfer of power from one frequency to another via the convolution of Fejér's kernel and the true spectral density. One method for decreasing the bias in the periodogram is called tapering.

Tapering is a method for reducing the sidelobes in the Fejér kernel. If the sidelobes are reduced, then there will be a substantial decrease in the bias of $f^{\sim}(\lambda)$ due to leakage. It should be noted, however, that this decrease in bias is at the cost of increasing the variance of our estimator. Now in order to use the method of tapering, we first form the product $h_t Y(t)$ for each t , where h_t is called the data taper. Next, let

$$J(\lambda) = \sum_{t=1}^N h_t Y(t) e^{-2\pi i \lambda t} \quad (6.28)$$

then by the spectral representation theorem

$$J(\lambda) = \int_0^1 H(\lambda - \omega) dZ(\omega) \quad (6.29)$$

where $\{Z(\cdot)\}$ is an orthogonal process and $\{h_t\}$ and $H(\cdot)$ are Fourier transform pairs. Moreover, if $\{h_t\}$ is an infinite sequence such that for $t < 1$ and $t > N$ $h_t = 0$ then

$$H(\lambda) = \sum_{t=1}^N h_t e^{-2\pi i \lambda t}. \quad (6.30)$$

Additionally,

$$f_{(t)}^{\sim}(\lambda) = |J(\lambda)|^2 = \left| \sum_{t=1}^N h_t Y(t) e^{-2\pi i \lambda t} \right|^2 \quad (6.31)$$

which implies

$$\mathbb{E} [f_{(t)}^{\sim}(\lambda)] = \int_0^1 H(\lambda - \omega) f(\omega) d\omega \quad (6.32)$$

(see Percival and Walden 1993 for proof). Hence, the goal is to select $\{h_t\}$ and hence $H(\cdot)$ so that the side lobes in the Fejér kernel are reduced. However, this reduction in the sidelobes causes an increase in the width of the main (central) lobe, and thus a variance bias tradeoff.

6.4.2 Multitaper

In the previous section it is noted that one method for reducing bias due to leakage is to apply a data taper to the process prior to computing the estimator for the spectral density. In this section we describe a slightly modified version of a method of multitapering used by Walden et al. (1998). Furthermore, in what follows, the notation used is similar to that found in Percival and Walden (2000), where a more comprehensive discussion of this method can be found. The idea behind multitaper spectral estimates is to form a small number, K , of tapered periodograms, each having a different data taper, and then average them together. If the K data tapers are all pairwise orthogonal and each of them sufficiently prevents leakage then the resulting spectral density estimator will have both a reduction in bias and variance compared to the raw periodogram.

In order to apply this method, let $\{Y(t)\}$ be a real valued mean zero stationary process. Additionally, let $\{h_{n,t} : t = 0, \dots, N - 1\}$ $n = 0, \dots, K - 1$, denote the K different data tapers to be used in forming the multitaper spectral density estimator $f_{(mt)}^{\sim}(\lambda)$. Then

$$f_{(mt)}^{\sim}(\lambda) \equiv \frac{1}{K} \sum_{n=0}^{K-1} f_{(mt),n}^{\sim}(\lambda) \quad (6.33)$$

where

$$f_{(mt),n}^{\sim}(\lambda) \equiv \left| \sum_{t=0}^{N-1} h_{n,t} Y(t) e^{-2\pi i \lambda t} \right|^2. \quad (6.34)$$

Note, the K data tapers are chosen to be orthonormal and the functions $f_{(mt),n}^{\sim}(\lambda)$ are known as direct spectral estimators. The set of tapers used in this work, are the set of sine tapers originally developed by Riedel and Sidorenko (1995); they are defined by

$$h_{n,t} = \left(\frac{2}{N+1} \right)^{\frac{1}{2}} \sin \left(\frac{(n+1)\pi(t+1)}{N+1} \right), \quad t = 0, \dots, N-1. \quad (6.35)$$

Under suitable conditions and for large N

$$f_{(mt)}^{\sim}(\lambda) \stackrel{d}{=} \frac{f(\lambda) \chi_{2K}^2}{2K}, \quad 0 < \lambda < .5. \quad (6.36)$$

Using the results of Bartlett and Kendall (1946), corresponding to the properties of $\log(\chi_{2K}^2)$, it follows that for $0 < \lambda < .5$

$$\mathbb{E} [\log f_{(mt)}^{\sim}(\lambda)] = \log [f(\lambda)] + \psi(K) - \log(K) \quad (6.37)$$

and

$$\text{Var} [\log f_{(mt)}^{\sim}(\lambda)] = \psi'(K) \quad (6.38)$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions respectively. Furthermore, by the results of Bartlett and Kendall (1946), for $0 < \lambda_j < .5$ the random variables

$$\eta(\lambda_j) = \log \left(\frac{f_{(mt)}^{\sim}(\lambda_j)}{f(\lambda_j)} \right) - \psi(K) + \log(K) \quad (6.39)$$

are correlated and approximately $N(0, \psi'(K))$. Moreover, Bartlett and Kendall suggest guidelines for when this approximation is appropriate; it can safely be used for $K \geq 5$ and tentatively for $K = 4$.

Finally, for $0 < \lambda_j < .5$, let

$$g_{(mt)}^{\sim}(\lambda_j) = \log [f_{(mt)}^{\sim}(\lambda_j)] - \psi(K) + \log(K) \quad (6.40)$$

then

$$g_{(mt)}^{\sim}(\lambda_j) = \log [f(\lambda_j)] + \eta(\lambda_j). \quad (6.41)$$

Therefore, the log spectral density estimator is equal to the true spectral density plus noise. Moreover, the noise is correlated and can be well approximated by Gaussian random variable with mean zero and known variance. Thus, this construction provides an ideal setting for wavelet smoothing, which can then be interpreted using the log exponential model and its corresponding MA(∞) representation. A more complete discussion of the properties of $\{h_{n,t}\}$ and $f_{(mt)}^{\sim}$ can be found in Riedel and Sidorenko (1995) and in Percival and Walden (2000) respectively.

6.5 EXP(m) via Multitaper Wavelet Spectrum Estimation

In order to estimate the coefficients of the EXP(m) model, the multitaper wavelet spectral estimation method of Walden et al. (1998) is adapted and extended to permit parameter estimation. Let $Y(0), \dots, Y(N-1)$ be a mean zero stationary process with spectral density $f(\lambda)$, and $2M = 2^{J+1}$ be the smallest power of two greater than or equal to the sample size N . Again, using notation similar to that found in Percival and Walden (2000), the parameter estimation method can be formulated as follows:

1. Form K ($K \geq 4$) series of length $2M$ by using the sine tapers and padding the tapered series with $2M - N$ zeros.

$$\{h_{n,0}Y(0), \dots, h_{n,N-1}Y(N-1), 0, \dots, 0\}, \quad n = 0, \dots, K-1. \quad (6.42)$$

Note, the sequence in expression (6.42) has $2M - N$ zeros and if $Y(t)$ is not assumed to have mean zero replace $Y(t)$ by $Y(t) - \bar{Y}$. Additionally, in the original algorithm proposed by Walden et al. (1998) K is assumed to be greater than or equal to 5 and 2^M is any power of two greater than or equal to the sample size N .

2. Using the fast Fourier transform compute

$$f_{(mt),n}^{\sim}(\lambda_k) \equiv \left| \sum_{t=0}^{2M-1} h_{n,t} Y(t) e^{-2\pi i \lambda_k t} \right|^2, \quad k = 0, \dots, 2M-1 \quad (6.43)$$

where $\lambda_k = \frac{k}{2M}$ and $h_{n,t} Y(t) \equiv 0$ for $t \geq N$.

3. Average the K direct spectral estimators, $f_{(mt),n}^{\sim}(\lambda_k)$, to form the raw multitaper estimate $f_{(mt)}^{\sim}(\lambda_k)$, see expression (6.33).
4. For $k = 0, \dots, 2M-1$, form

$$g_{mt}^{\sim}(\lambda_k) \equiv \log[f_{(mt)}^{\sim}(\lambda_k)] - \psi(K) + \log(K) \quad (6.44)$$

and place the $2M$ values in the column vector

$$\begin{aligned} \mathbf{g}_{\mathbf{mt}}^{\sim} &= \begin{pmatrix} Y_{mt}(\lambda_0) \\ \vdots \\ Y_{mt}(\lambda_{2M-1}) \end{pmatrix} = \begin{pmatrix} \log[f(\lambda_0)] \\ \vdots \\ \log[f(\lambda_{2M-1})] \end{pmatrix} + \begin{pmatrix} \eta(\lambda_0) \\ \vdots \\ \eta(\lambda_{2M-1}) \end{pmatrix} \\ &= \mathbf{D} + \boldsymbol{\eta}. \end{aligned} \quad (6.45)$$

5. Using the Daubechies' Least Asymmetric (LA(8)) wavelet, apply the discrete wavelet transform (DWT) out to level J_0 ($J_0 < J$) to obtain the partial DWT coefficients $\mathbf{W}_{\mathbf{mt},1}, \dots, \mathbf{W}_{\mathbf{mt},J_0}$ and $\mathbf{V}_{\mathbf{mt},J_0}$. Note that J_0 is chosen such that the three coarsest levels are eliminated. Additionally, $\mathbf{W}_{\mathbf{mt},j}$ are the j^{th} level DWT coefficients and $\mathbf{V}_{\mathbf{mt},J_0}$ are the level J_0 scaling coefficients.
6. Apply a soft threshold, δ , to $\mathbf{W}_{\mathbf{mt},1}, \dots, \mathbf{W}_{\mathbf{mt},J_0}$ to obtain $\mathbf{W}_{\mathbf{t},1}, \dots, \mathbf{W}_{\mathbf{t},J_0}$ where

- (a) δ is a level independent threshold

$$\delta = \delta^{(u)} \equiv \sqrt{2\sigma_{\eta}^2 \log(M)} \quad (6.46)$$

and $\sigma_{\eta}^2 = \psi'(K)$. $\delta^{(u)}$ is known as a level independent universal threshold.

(b) δ is a level dependent threshold

$$\delta = \delta_j^{(u)} \equiv \sqrt{2\sigma_j^2 \log(M)} \quad (6.47)$$

$\delta_j^{(u)}$ is known as a level dependent universal threshold. Note that σ_j can either be calculated exactly, see Percival and Walden (2000), or estimated using the MAD estimate for each level j

$$\hat{\sigma}_{(MAD)j} \equiv \frac{\text{median}(|\mathbf{W}_{\mathbf{j},\mathbf{k}} - \text{median}(\mathbf{W}_{\mathbf{j},\mathbf{k}})|)}{.6745}. \quad (6.48)$$

Note, the $\mathbf{V}_{\mathbf{mt},\mathbf{J}_0}$ remain unchanged; see Vidakovic (1999) for a comprehensive discussion surrounding wavelet thresholding.

7. Estimate \mathbf{D} via $\widehat{\mathbf{D}}$ by taking the inverse discrete wavelet transform (IDWT) of $\mathbf{W}_{\mathbf{t},1}, \dots, \mathbf{W}_{\mathbf{t},\mathbf{J}_0}$ and $\mathbf{V}_{\mathbf{mt},\mathbf{J}_0}$. Let \widehat{D}_k denote the k^{th} element of $\widehat{\mathbf{D}}$ and note \widehat{D}_k gives an estimate of $g(\lambda_k)$, the true log spectral density evaluated at the Fourier frequency λ_k .
8. Let $g_{(mthr)}^{\sim}(\lambda_k)$ be the multitaper wavelet log spectral estimator. Then

$$g_{(mthr)}^{\sim}(\lambda_k) = \begin{cases} \widehat{D}_0, & k=0; \\ \frac{1}{2} \left(\widehat{D}_k + \widehat{D}_{(2M-k)} \right), & k=1, \dots, M-1; \\ \widehat{D}_M, & k=M. \end{cases} \quad (6.49)$$

Note, since the thresholded multitaper estimate is not necessarily symmetric about $\lambda = \frac{1}{2}$, $g_{(mthr)}^{\sim}(\lambda_k)$ is taken to be the average of \widehat{D}_k and $\widehat{D}_{(2M-k)}$. Additionally, in this step Walden et al. (1998) estimate the spectrum rather than the log spectrum by using $\exp(\widehat{D}_k)$ in place of \widehat{D}_k .

9. For $h = 0, \dots, M$ approximate the true cepstral coefficients

$$\begin{aligned} c(h) &= \int_0^1 \exp(2\pi i h \lambda) \log [f(\lambda)] d\lambda \\ &= \int_0^1 \cos(2\pi h \lambda) \log [f(\lambda)] d\lambda \end{aligned} \quad (6.50)$$

by using the fast Fourier transform and $g_{\tilde{mthr}}(\lambda_k)$ to obtain $\{\hat{c}(0), \hat{c}(1), \dots, \hat{c}(M)\}$.

The coefficients $\hat{c}(h)$ are the wavelet cepstral coefficients.

10. Let

$$\log f_{\theta^{\wedge m}}(\lambda_j) = \hat{\theta}_0 + 2 \sum_{k=1}^m \hat{\theta}_k \cos(2\pi k \lambda_j) \quad (6.51)$$

and

$$\hat{h}_m(\lambda_j) = [g_{\tilde{mthr}}(\lambda_j) - \log f_{\theta^{\wedge m}}(\lambda_j)]^2 \quad (6.52)$$

where $\lambda_j = j/M, j = 0, 1, \dots, 2M - 1$ and $\hat{\theta}_k \equiv \hat{c}(k)$. Note that $g_{\tilde{mthr}}(\lambda_j) = g_{\tilde{mthr}}(\lambda_{2M-j})$ by symmetry. Compute

$$ISE_m^{(d)} = \frac{1}{2M} \sum_{j=0}^{2M-1} \hat{h}_m(\lambda_j) = \text{mean} [\hat{h}_m(\lambda_j)], \quad j = 0, 1, \dots, 2M - 1. \quad (6.53)$$

11. Choose different orders m_* based on the estimated model parameters and $ISE_m^{(d)}$, see Chapter VII. Then the model

$$\log f(\lambda) = \hat{\theta}_0 + 2 \sum_{k=1}^{m_*} \hat{\theta}_k \cos(2\pi k \lambda) \quad (6.54)$$

where

$$\hat{\theta}_k \equiv \hat{c}(k) \quad (6.55)$$

forms a candidate model.

In practice, for large N , this method is fairly robust to the exact number of tapers used ($K = 4, \dots, 8$). However, overall $K = 4$ performed at least as well as other choices of K for the majority of cases investigated. Additionally, for the cases we investigated, the choice of threshold did not have a significant impact on the procedure. Choosing the SURE, cross validation (CV), or generalized cross validation (GCV) threshold yields similar results. See Donoho and Johnstone (1994), Nason

(1996) and Jansen et al. (1997) respectively for further details on these types of wavelet thresholds.

The wavelet method of parameter estimation possesses many advantages. The first is that the method is fully nonparametric, assuming only that $\{Y(t)\}$ constitutes a realization from a stationary time series. Another advantage of the wavelet method is that the parameter estimates are not obtained from a spectral estimator suffering from substantial leakage. This is especially important when little is known about the true spectral density of the process generating the data. Additionally, for processes with raw spectral estimators suffering from leakage, the wavelet cepstral coefficients often have smaller mean square error than the cepstral coefficients estimated using traditional methods. Furthermore, the wavelet method, of parameter estimation, can be implemented in a fairly automatic manner. This makes this method attractive in situations where the goal is to model many individual time series.

6.6 Simulation Results

In this section we use simulated data to compare the performance of the maximum likelihood, ordinary least squares, and wavelet parameter estimates. Using the indirect simulation method described in Chapter V, 1000 independent samples from two Gaussian time series models are generated for several sample sizes ($N = 100, 256, 512, \text{ and } 1000$). Both models are stationary short memory time series, however, the dynamic range of the two models differs significantly. The first model has small dynamic range, while the dynamic range of second models is much larger. Therefore, in the second model there is bias, due to leakage, present in the periodogram. The models considered are

$$\log f(\lambda) = -3.50 + 2 [.5 \cos(2\pi\lambda) + .25 \cos(2\pi 2\lambda) + .10 \cos(2\pi 3\lambda)] \quad (6.56)$$

and

$$\log f(\lambda) = -1.0 + 2 [2.5 \cos(2\pi\lambda) - 1.5 \cos(2\pi 2\lambda)], \quad (6.57)$$

see Figure 4.

The starting values for the maximum likelihood parameter estimation are obtained by approximating the cepstral correlations using $\log \tilde{f}(\lambda)$ and the fast Fourier transform. The criterion for convergence is based on the sum of squares of the modifications to the parameters divided by the number of parameters being estimated multiplied by the series length. The iterations are stopped the first time this quantity falls below .00001. Note, this is similar to the criterion used in Bloomfield (1973). However, Bloomfield (1973) terminates iterations the first time this quantity falls below .01. The criterion used by Bloomfield may not be strong enough to guarantee the parameter estimates reach convergence, in all cases. Making the modification to .00001 insures convergence, in most cases, with little effect to the number of iterations. For the wavelet parameters, both the level independent and level dependent universal thresholds are used see expressions (6.46) and (6.47). For the level dependent threshold, σ_j is estimated using the MAD estimate, see expression (6.48).

In this simulation study the maximum likelihood and ordinary least squares parameter estimates have small mean square error only for the spectrum with small dynamic range. That is the spectral density whose periodogram does not suffer from bias problems due to leakage. For the first model, the model with small dynamic range, all three methods of parameter estimation exhibit similar mean squared errors. However for model 2, the model suffering from leakage, the wavelet parameter estimates have substantially smaller mean squared errors than both the maximum likelihood and the ordinary least squared estimation methods. Thus, in cases where there is little known about the true spectral density of the process generating the

data, the wavelet method may provide superior parameter estimates to those obtained using either the method of maximum likelihood or least squares regression.

The results of this simulation can be found in Tables 2-9.

Table 2. Model 1 with $N=1000$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively.

Note that all values have been rounded to three decimal places.

	N=1000	$\theta_0 = -3.50$	$\theta_1 = .50$	$\theta_2 = .25$	$\theta_3 = .10$
Mean					
mle (4.056)		-3.502	0.499	0.249	0.099
ols		-3.501	0.500	0.247	0.101
wave(i)		-3.503	0.497	0.249	0.096
wave(d)		-3.503	0.497	0.249	0.095
Bias					
mle		-0.003	-0.001	-0.001	-0.001
ols		-0.001	0.000	-0.003	0.001
wave(i)		-0.003	-0.003	-0.001	-0.005
wave(d)		-0.003	-0.003	-0.001	-0.005
Variance					
mle		0.002	0.001	0.001	0.001
ols		0.003	0.002	0.002	0.002
wave(i)		0.002	0.001	0.001	0.001
wave(d)		0.002	0.001	0.001	0.001
MSE					
mle		0.002	0.001	0.001	0.001
ols		0.003	0.002	0.002	0.002
wave(i)		0.002	0.001	0.001	0.001
wave(d)		0.002	0.001	0.001	0.001

Table 3. Model 1 with $N=512$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively.

Note that all values have been rounded to three decimal places.

	N=512	$\theta_0 = -3.50$	$\theta_1 = .50$	$\theta_2 = .25$	$\theta_3 = .10$
Mean					
	mle (4.241)	-3.505	0.500	0.251	0.097
	ols	-3.500	0.501	0.251	0.097
	wave(i)	-3.507	0.497	0.250	0.093
	wave(d)	-3.507	0.497	0.250	0.093
Bias					
	mle	-0.005	0.000	0.001	-0.003
	ols	0.000	0.001	0.001	-0.003
	wave(i)	-0.007	-0.003	0.000	-0.006
	wave(d)	-0.007	-0.003	0.000	-0.007
Variance					
	mle	0.004	0.002	0.002	0.002
	ols	0.007	0.003	0.003	0.003
	wave(i)	0.005	0.003	0.002	0.002
	wave(d)	0.005	0.003	0.002	0.002
MSE					
	mle	0.002	0.001	0.001	0.001
	ols	0.003	0.002	0.002	0.002
	wave(i)	0.002	0.001	0.001	0.001
	wave(d)	0.002	0.001	0.001	0.001

Table 4. Model 1 with $N=256$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=256	$\theta_0 = -3.50$	$\theta_1 = .50$	$\theta_2 = .25$	$\theta_3 = .10$
Mean					
mle (4.457)		-3.508	0.497	0.245	0.099
ols		-3.499	0.502	0.240	0.101
wave(i)		-3.508	0.494	0.246	0.010
wave(d)		-3.508	0.494	0.246	0.099
Bias					
mle		-0.008	-0.003	-0.005	-0.001
ols		0.001	0.002	-0.010	0.001
wave(i)		-0.008	-0.006	-0.004	0.000
wave(d)		-0.008	-0.006	-0.004	-0.001
Variance					
mle		0.008	0.004	0.004	0.004
ols		0.013	0.007	0.007	0.006
wave(i)		0.009	0.005	0.005	0.004
wave(d)		0.009	0.005	0.005	0.004
MSE					
mle		0.008	0.004	0.004	0.004
ols		0.013	0.007	0.007	0.006
wave(i)		0.009	0.005	0.005	0.004
wave(d)		0.009	0.005	0.005	0.004

Table 5. Model 1 with $N=100$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=100	$\theta_0 = -3.50$	$\theta_1 = .50$	$\theta_2 = .25$	$\theta_3 = .10$
Mean					
mle (4.964)		-3.530	0.498	0.236	0.096
ols		-3.499	0.509	0.227	0.107
wave(i)		-3.525	0.491	0.241	0.097
wave(d)		-3.524	0.491	0.242	0.096
Bias					
mle		-0.030	-0.002	-0.014	-0.004
ols		0.001	0.009	-0.023	0.007
wave(i)		-0.025	-0.009	-0.009	-0.003
wave(d)		-0.024	-0.009	-0.008	-0.004
Variance					
mle		0.021	0.011	0.011	0.011
ols		0.033	0.018	0.018	0.017
wave(i)		0.025	0.012	0.012	0.011
wave(d)		0.025	0.012	0.012	0.011
MSE					
mle		0.022	0.011	0.011	0.011
ols		0.033	0.018	0.019	0.018
wave(i)		0.025	0.012	0.012	0.011
wave(d)		0.025	0.012	0.012	0.011

Table 6. Model 2 with $N=1000$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=1000	$\theta_0 = -1.0$	$\theta_1 = 2.5$	$\theta_2 = -1.5$
Mean				
mle (7.33)		-0.497	2.067	-1.179
ols		-0.496	2.034	-1.131
wave(i)		-1.011	2.501	-1.500
wave(d)		-1.011	2.501	-1.500
Bias				
mle		0.503	-0.432	0.321
ols		0.504	-0.465	0.369
wave(i)		-0.011	0.006	0.000
wave(d)		-0.011	0.016	0.000
Variance				
mle		0.114	0.077	0.037
ols		0.111	0.089	0.049
wave(i)		0.002	0.001	0.001
wave(d)		0.002	0.001	0.001
MSE				
mle		0.366	0.264	0.140
ols		0.365	0.306	0.185
wave(i)		0.003	0.001	0.001
wave(d)		0.003	0.001	0.001

Table 7. Model 2 with $N=512$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=512	$\theta_0 = -1.0$	$\theta_1 = 2.5$	$\theta_2 = -1.5$
Mean				
mle (7.85)		-0.337	1.934	-1.091
ols		-0.329	1.889	-1.032
wave(i)		-1.013	2.504	-1.499
wave(d)		-1.013	2.504	-1.499
Bias				
mle		0.663	-0.566	0.409
ols		0.671	-0.611	0.468
wave(i)		-0.013	0.004	0.001
wave(d)		-0.013	0.004	0.001
Variance				
mle		0.143	0.095	0.043
ols		0.140	0.109	0.055
wave(i)		0.005	0.002	0.002
wave(d)		0.005	0.002	0.003
MSE				
mle		0.582	0.416	0.210
ols		0.589	0.483	0.275
wave(i)		0.005	0.002	0.002
wave(d)		0.005	0.002	0.002

Table 8. Model 2 with $N=256$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=256	$\theta_0 = -1.0$	$\theta_1 = 2.5$	$\theta_2 = -1.5$
Mean				
mle (7.962)		-0.160	1.793	-1.000
ols		-0.140	1.734	-0.936
wave(i)		-1.009	2.498	-1.492
wave(d)		-1.009	2.498	-1.492
Bias				
mle		0.840	-0.707	0.497
ols		0.860	-0.766	0.564
wave(i)		-0.009	-0.002	0.008
wave(d)		-0.009	-0.002	0.008
Variance				
mle		0.194	0.124	0.052
ols		0.198	0.144	0.066
wave(i)		0.009	0.005	0.005
wave(d)		0.009	0.005	0.005
MSE				
mle		0.898	0.623	0.299
ols		0.938	0.731	0.384
wave(i)		0.009	0.005	0.005
wave(d)		0.009	0.005	0.005

Table 9. Model 2 with $N=100$: The number in parenthesis is the mean number of iterations for the maximum likelihood parameter estimation. Additionally, (i) and (d) correspond to level independent and level dependent thresholding respectively. Note that all values have been rounded to three decimal places.

	N=100	$\theta_0 = -1.0$	$\theta_1 = 2.5$	$\theta_2 = -1.5$
Mean				
	mle (7.895)	0.076	1.592	-0.882
	ols	0.129	1.510	-0.802
	wave(i)	-0.863	2.358	-1.406
	wave(d)	-0.863	2.358	-1.405
Bias				
	mle	1.076	-0.908	0.618
	ols	1.129	-0.990	0.698
	wave(i)	0.137	-0.142	0.094
	wave(d)	0.137	-0.142	0.094
Variance				
	mle	0.251	0.153	0.064
	ols	0.270	0.179	0.076
	wave(i)	0.030	0.016	0.014
	wave(d)	0.030	0.016	0.014
MSE				
	mle	1.409	0.977	0.446
	ols	1.546	1.160	0.563
	wave(i)	0.049	0.036	0.023
	wave(d)	0.049	0.036	0.023

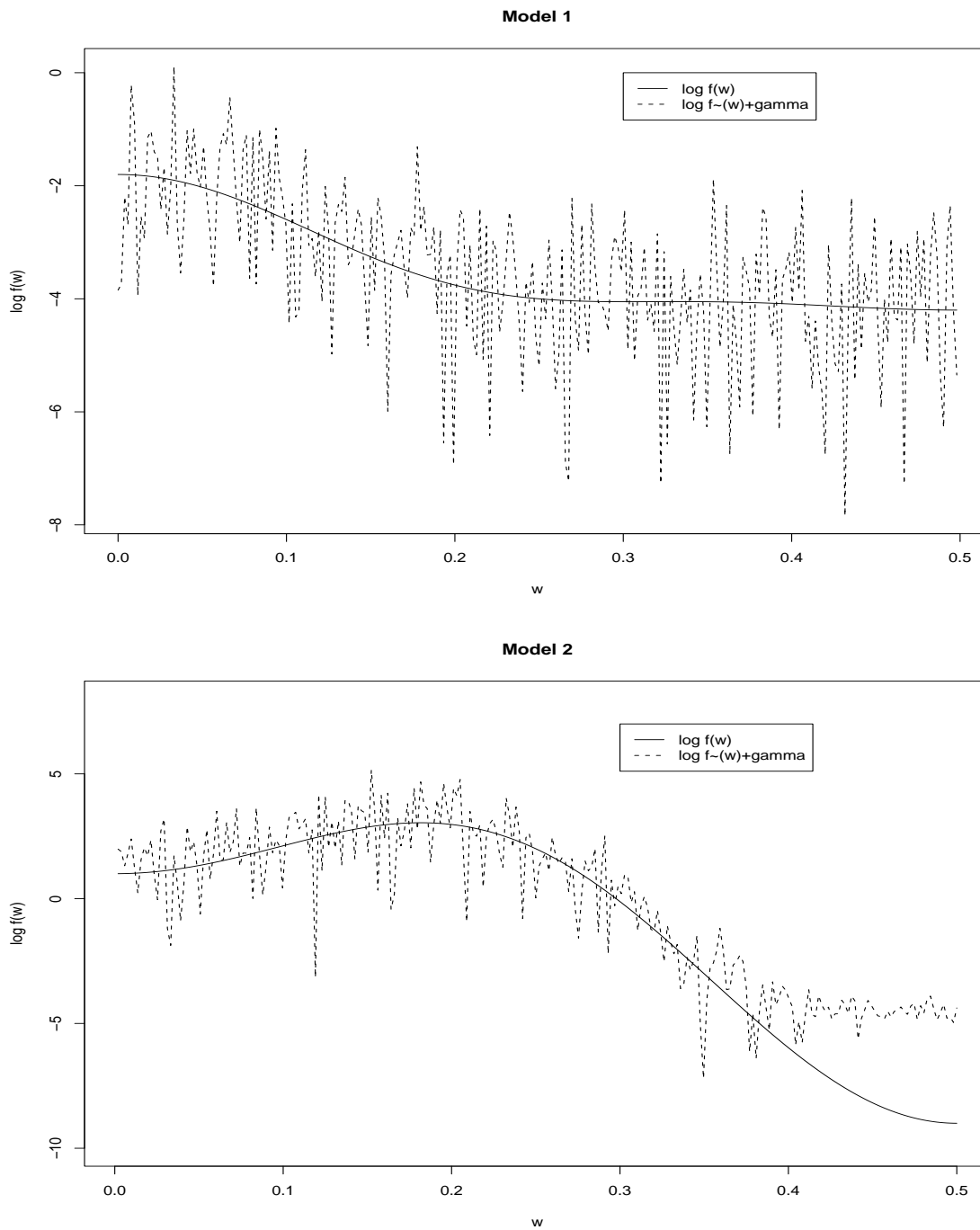


Figure 4. An example plot of $\log f(\lambda)$ and $\log \tilde{f}(\lambda) + .57721$. $\tilde{f}(\lambda)$ is generated using the indirect simulation method with $N=512$. Both Models 1 and 2 are plotted. Note, in the plot of model 2 it is apparent that there is leakage in the periodogram.

CHAPTER VII

ORDER SELECTION AND MODEL VALIDATION

7.1 Introduction and Philosophy

Let $\{Y(t), t \in Z\}$ be a Gaussian mean zero observed time series. Time domain time series modelling can be described by the estimation of a whitening filter that transforms the time series to a white noise innovation series $\{Y^\nu(t)\}$. Conversely, the EXP model is a spectral model (i.e. a frequency domain model). Thus, EXP modelling can be characterized by the estimation of a whitening model of the spectrum. Let $f^w(\lambda)$ be the whitening spectral density, where

$$f^w(\lambda) = \frac{f(\lambda)}{f_{\theta^\wedge}(\lambda)}. \quad (7.1)$$

Thus, $f^w(\lambda)$ is the ratio of the true spectral density and the estimated model. Similarly, let $f_n^w(\lambda)$ be the normalized (normalized to integrate to one) whitening spectral density. Then

$$f_n^w(\lambda) = \frac{f^w(\lambda)}{\sigma^{2w}} \quad (7.2)$$

where

$$\sigma^{2w} = \int_0^1 f^w(\lambda) d\lambda. \quad (7.3)$$

Furthermore, the Fourier transform of $f_n^w(\lambda)$ is called the residual (whitening) correlations and are denoted $\rho^w(h)$,

$$\rho^w(h) = \int_0^1 f_n^w(\lambda) \cos(2\pi h\lambda) d\lambda. \quad (7.4)$$

The goal of model validation (fitting) is to determine if $f^w(\lambda) = \sigma^{2w}$ (a constant); the spectral density of white noise. Equivalently, model validation can be formulated

in terms of $\rho^w(h)$ for $h = 1, 2, \dots$. The goal is then to determine if, for $h = 1, 2, \dots$, $\rho^w(h)$ are the correlations of white noise (i.e. $\rho^w(h) = 0$ for $h = 1, 2, \dots$).

Unfortunately, in practice $f^w(\lambda)$ is not observable. Thus, we formulate the sample whitening spectral density, $f^{w*}(\lambda)$ which is defined by

$$f^{w*}(\lambda) = \frac{f^\sim(\lambda)}{f_{\theta^\wedge}(\lambda)} \quad (7.5)$$

where $f^\sim(\lambda)$ is the sample spectral density given by expression (4.24). Although the sample whitening spectral density is defined for all $\lambda \in [0, 1]$, in practice this quantity is only calculated at the Fourier frequencies. For example,

$$f^{w*}(\lambda_j) = \frac{f^\sim(\lambda_j)}{f_{\theta^\wedge}(\lambda_j)} \quad (7.6)$$

where $f^\sim(\lambda_j)$ is the sample spectral density and $\lambda_j = j/N$, $j = 0, \dots, N-1$, are the Fourier frequencies. Additionally, we assume that the data has been mean centered and thus $f^{w*}(0) \equiv 0$. Similarly, in practice we define the normalized sample whitening spectral density, $f_n^{w*}(\lambda_j)$, by

$$f_n^{w*}(\lambda_j) = \frac{f^{w*}(\lambda_j)}{\hat{\sigma}^{2w}} \quad (7.7)$$

where

$$\begin{aligned} \hat{\sigma}^{2w} &= \frac{1}{N} \sum_{j=0}^{N-1} f^{w*}(\lambda_j) \\ &= \frac{1}{N} \sum_{j=0}^{N-1} \frac{f^\sim(\lambda_j)}{f_{\theta^\wedge}(\lambda_j)}. \end{aligned} \quad (7.8)$$

Thus,

$$\frac{1}{N} \sum_{j=0}^{N-1} f_n^{w*}(\lambda_j) = 1 \quad (7.9)$$

by construction. Moreover, let $\hat{\rho}^w(h)$ be the sample whitening correlations then we define $\hat{\rho}^w(h)$ by the discrete Fourier transform of $f_n^{w*}(\lambda_j)$, $j = 0, \dots, N-1$,

$$\hat{\rho}^w(h) \equiv \frac{1}{N} \sum_{j=0}^{N-1} f_n^{w*}(\lambda_j) \cos(2\pi h \lambda_j). \quad (7.10)$$

For model validation, we test the null hypothesis that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise. This measures if an adequate smoother of $f^{w*}(\lambda)$ is a constant function. Thus, we assume, that under the null hypothesis $f^{w*}(\lambda)$ has the same distribution as the sample spectral density of white noise. Therefore,

$$f^{w*}(\lambda) \stackrel{d}{=} \sigma^{2w} \begin{cases} \chi_2^2/2 & \lambda \neq .5 \\ \chi_1^2 & \lambda = .5 \end{cases} \quad (7.11)$$

by analogy. Moreover, under the null hypothesis and the previous assumption, it follows that $f_n^{w*}(\lambda_j)$ has the same distribution as the normalized periodogram of white noise.

Now, in order to effectively model a time series using the EXP(m) model, a method is needed for choosing the order of candidate models. In this chapter we present a method for choosing candidate models when using the wavelet estimation approach. Additionally, we propose general tools for model validation. Note, in practice, it is desirable to have m be small in order to minimize spurious spectral peaks.

7.2 Linhart and Volkers Criterion

Linhart and Volkers(1985) formulate a general approach to model fitting of stationary Gaussian time series, satisfying certain regularity conditions. The foundations of their approach are based on minimizing an expected discrepancy function. Let $\Delta(f, f_\theta)$ be the discrepancy function and $\Delta(\hat{\theta})$ be the overall discrepancy

$$\Delta(\hat{\theta}) = \Delta(f, f_{\hat{\theta}}). \quad (7.12)$$

The overall discrepancy takes into account both error due to approximation and error due to estimation. The expected overall discrepancy is denoted $E\Delta(\hat{\theta})$; the estimator

$\hat{E}\Delta(\hat{\theta})$ of $E\Delta(\hat{\theta})$ is called the criterion and is used for judging model fit. Models are chosen which minimize the quantity $\hat{E}\Delta(\hat{\theta})$. Since Linhart and Volkers (1985) assume the goal of fitting the model is forecasting, the discrepancy is chosen to be the (one step ahead) mean square error of prediction. Note, in the case of autoregressive models this discrepancy coincides with Akaike's FPE criterion (if a certain approximation is made), see Linhart and Volkers (1985). Now under f (f unknown), the best linear predictor of $Y(N+1)$ given $Y(N), Y(N-1), \dots$ with spectral density g_θ has (one step ahead) mean square error of prediction defined by

$$\Delta(f, g_\theta) = \int_0^1 \frac{f(\lambda)}{k(\lambda, \theta)} d\lambda \quad (7.13)$$

where

$$k(\lambda, \theta) = \frac{g_\theta}{\sigma_g^2} \quad (7.14)$$

and σ_g^2 is the innovation variance of the approximating model. Recall,

$$\sigma_g^2 = \exp\left(\int_0^1 \log g_\theta(\lambda) d\lambda\right). \quad (7.15)$$

The empirical discrepancy is

$$\Delta_N(\theta) = \frac{1}{N} \sum_{j=0}^{N-1} \frac{f^\sim(\lambda_j)}{k(\lambda_j, \theta)} \quad (7.16)$$

where $\lambda_j = j/N$, $j = 0, \dots, N-1$ (the Fourier frequencies). The minimum discrepancy estimator is given by

$$\hat{\theta} = \arg \min\{\Delta_N(\theta) : \theta \in \Theta\}. \quad (7.17)$$

Furthermore, $\hat{\theta}$ obtained in expression (7.17) are asymptotically equivalent to the maximum likelihood estimates described in Chapter II. Assuming regularity conditions and making appropriate approximations, the authors arrive at a general criterion, see Linhart and Zucchini (1986). They recommend as a rule the simple criterion

$$\Delta_N(\hat{\theta}) \left(1 + \frac{2m}{N}\right) \quad (7.18)$$

be used and to choose the order m minimizing the criterion. In terms of the exponential model

$$k^{-1}(\lambda_j, \theta) = \exp \left(-2 \sum_{k=1}^m \hat{\theta}_k \cos(2\pi k \lambda_j) \right). \quad (7.19)$$

Thus, the criterion becomes

$$\left[\frac{1}{N} \sum_{j=0}^{N-1} f^{\sim}(\lambda_j) \exp \left(-2 \sum_{k=1}^m \hat{\theta}_k \cos(2\pi k \lambda_j) \right) \right] \left(1 + \frac{2m}{N} \right). \quad (7.20)$$

Following the philosophy described in Section 7.1, when using maximum likelihood parameter estimation, this criterion can be used as an aid in finding candidate models. These models should then be evaluated to determine whether they provide a whitening transformation of the spectrum; among models which provide a whitening transformation, select the model with the smallest order. Conversely, in the approach of Linhart and Volkers (1985), the fitting procedure is judged solely on the basis of value of the criterion, selecting the model which possesses the minimum estimated expected discrepancy.

7.2.1 Simulation Study

In this section, we examine the performance of the Linhart-Volkers criterion (LVC), for EXP models, as a sole means of order selection. Using the indirect method of simulation described in Chapter V, 1000 independent samples are generated from two Gaussian time series models for several sample sizes ($N=100, 256, 512, 1000$). The models used are the same models used in the simulations in Chapter VI and are given by expressions (6.56) and (6.57). The model given by expression (6.56), model 1, has small dynamic range and thus does not present any problems of bias due to leakage. Conversely, model 2, the model given by expression (6.57) has larger dynamic range than model 1 and suffers from bias problems due to leakage.

As previously noted, the Linhart-Volkers criterion was originally developed in

conjunction with parameter estimates asymptotically equivalent to the maximum likelihood estimates. However, in this simulation, parameters are estimated using both the maximum likelihood method and the wavelet method, with level independent thresholding as described in Chapter VI. As expected this simulation provides empirical evidence that an alternative method for selecting candidate models needs to be considered when using the wavelet method of parameter estimation.

Table 10. Linhart-Volkers criterion: Frequency of order chosen for model 1 using both maximum likelihood and wavelet parameter estimation.

Model 1: EXP(3)	m	$N = 100$	$N = 256$	$N = 500$	$N = 1000$
Maximum Likelihood	1	169	4	0	0
	2	75	5	0	0
	3	453	678	707	714
	4	94	136	104	122
	5	67	61	76	65
	6	41	34	37	38
	7	39	23	27	22
	8	19	25	22	13
	9	21	15	16	14
	10	22	19	11	12
Wavelet Method	1	140	4	0	0
	2	535	428	198	48
	3	288	444	652	787
	4	17	99	127	144
	5	8	16	17	18
	6	9	3	5	3
	7	2	5	1	0
	8	1	0	0	0
	9	0	0	0	0
	10	0	1	0	0

For model 1, the distribution of orders selected differs drastically for the maximum likelihood and wavelet method. For models of this type having order m , this occurs because the first m parameter estimates for both the maximum likelihood and wavelet methods are approximately equal. However, for parameters $m + 1, m + 2, \dots$

Table 11. Linhart-Volkers criterion: Frequency of order chosen for model 2 using both maximum likelihood and wavelet parameter estimation. Note that this model suffers from leakage.

Model 2: EXP(2)	m	$N = 100$	$N = 256$	$N = 500$	$N = 1000$
Maximum Likelihood	1	0	0	0	0
	2	0	0	0	0
	3	528	232	127	118
	4	271	456	465	270
	5	73	149	220	342
	6	54	78	67	125
	7	27	44	61	63
	8	12	16	29	41
	9	18	14	18	23
	10	17	11	13	18
Wavelet Method	1	654	576	144	246
	2	140	170	252	319
	3	156	147	147	184
	4	31	85	109	162
	5	11	20	47	86
	6	6	2	0	2
	7	1	0	1	1
	8	1	0	0	0
	9	0	0	0	0
	10	0	0	0	0

the wavelet parameter estimates are significantly smaller than the maximum likelihood parameters.

As illustrated in Tables 10 and 11, when parameter estimation is done using maximum likelihood, the Linhart-Volkers criterion works well for short memory time series not suffering from bias due to leakage (i.e. small dynamic range and lack of rapid fluctuations in the spectrum). However, when the periodogram is a significantly biased estimate of the spectrum, as in model 2, this method breaks down. Furthermore, even when the correct order is selected the model provides a poor fit, since the parameter estimates are biased. Therefore, other methods for selecting the

orders of candidate models should be considered in situations where little information is known about the spectrum generating the observed time series. For example, situations where the wavelet parameter estimation method may be advantageous.

7.3 Order Selection

In this section we suggest a method for selecting the orders of candidate EXP models when using the wavelet method of parameter estimation. The method proposed is based on the integrated squared error between the nonparametric wavelet estimate and its interpreted log exponential model. It should be emphasized that the goal, in selecting candidate models, is only to identify potentially appropriate models which are then tested to determine goodness of fit.

Recall, that the discrete integrated squared error, $ISE_m^{(d)}$, has the form

$$ISE_m^{(d)} = \frac{1}{2M} \sum_{j=0}^{2M-1} \hat{h}_m(\lambda_j) = \text{mean} \left[\hat{h}_m(\lambda_j) \right], \quad j = 0, 1, \dots, 2M - 1 \quad (7.21)$$

where

$$\hat{h}_m(\lambda_j) = [g_{(mthr)}^{\sim}(\lambda_j) - \log f_{\theta^{\wedge m}}(\lambda_j)]^2. \quad (7.22)$$

The quantity defined by $ISE_m^{(d)}$ is the trapezoid approximation to the continuous integrated squared error calculated at the Fourier frequencies. This can be seen by considering

$$\int_0^1 h(\lambda) d\lambda. \quad (7.23)$$

Suppose we are given $\lambda = 0, 1/2M, \dots, 1$ and $h(0), h(1/2M), \dots, h(1)$, then by the trapezoid approximation

$$\begin{aligned} \int_0^1 h(\lambda) d\lambda &\approx \sum_{j=1}^{2M} \left[\frac{1}{2} \left\{ h\left(\frac{j}{2M}\right) + h\left(\frac{j-1}{2M}\right) \right\} \right] \frac{1}{2M} \\ &= \frac{1}{4M} \left(\sum_{j=1}^{2M} h\left(\frac{j}{2M}\right) + \sum_{j=1}^{2M} h\left(\frac{j-1}{2M}\right) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4M} \left(h(0) + 2 \sum_{j=1}^{2M-1} h\left(\frac{j}{2M}\right) + h(1) \right) \\
&= \frac{1}{2M} \sum_{j=0}^{2M-1} h\left(\frac{j}{2M}\right) \\
&= \text{mean} \left[h\left(\frac{j}{2M}\right) \right], \quad j = 0, 1/2M, \dots, 1
\end{aligned} \tag{7.24}$$

since $h(0) = h(1)$ by symmetry.

In order to use $ISE_m^{(d)}$ as an aid in selecting candidate models we calculate $ISE_m^{(d)}$ for $m = 1, \dots, T$, where T is suitably chosen. Candidate models of order m are chosen such that both $\hat{\theta}_{j+1}$ and $ISE_j^{(d)}$, $j = m, m+1, \dots$, are sufficiently small and $ISE_j^{(d)}$ is not changing significantly. Thus, the goal is to continue adding terms in the model until the terms being added are not significantly different from zero. Note that interpreting the nonparametric representation using the log exponential model has the effect of removing both spurious peaks due to undersmoothing and due to artifacts of the wavelet estimator. However, the removal of spurious peaks may also lead to removal of small peaks that are caused by the underlying process generating the data. Thus, once candidate models are selected, they must be evaluated to determine if they provide a whitening transformation of the spectrum.

In order to illustrate the use of $ISE_m^{(d)}$ we simulated a Gaussian time series ($N = 256$) using the indirect simulation method described in Chapter V. The model considered is given by

$$\log f(\lambda) = 1 + 2 [.65 \cos(2\pi\lambda) - .38875 \cos(2\pi 2\lambda) - .298458 \cos(2\pi 3\lambda)]. \tag{7.25}$$

The estimated coefficients $\hat{\theta}_0, \dots, \hat{\theta}_{15}$ and $ISE_m^{(d)}$ can be found in Table 12.

Lastly Figure 5 shows both a plot of $\log \tilde{f}(\lambda) + \gamma$ ($\gamma = .57721$) with the wavelet nonparametric representation superimposed and a plot of the EXP(3) model with the nonparametric representation superimposed.

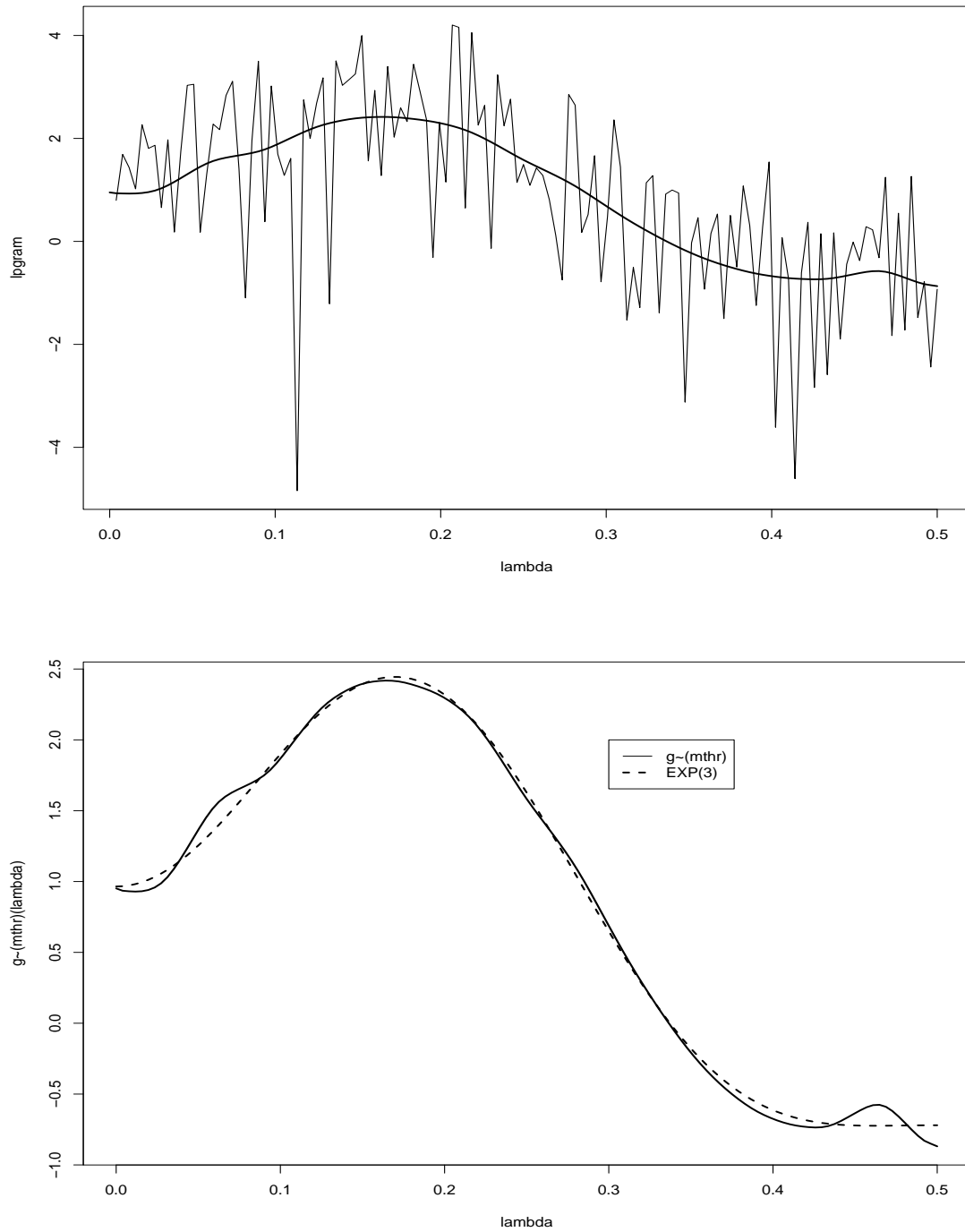


Figure 5. Plot of $\log f_{\sim}(\lambda) + \gamma$ superimposed with $g_{\sim}(mthr)$ and plot of $EXP(3)$ with $g_{\sim}(mthr)$ superimposed.

Table 12. This table displays the parameter estimates and $ISE_m^{(d)}$ for the simulated model given in expression (7.25). The candidate model chosen based on this information is an $EXP(3)$.

m	$\hat{\theta}_m$	$ISE_m^{(d)}$	m	$\hat{\theta}_m$	$ISE_m^{(d)}$
0	.8751	NA	8	-.0182	.0020
1	.6782	.4107	9	-.0006	.0020
2	-.3763	.1367	10	-.0134	.0016
3	-.2569	.0035	11	.0049	.0016
4	.0071	.0035	12	-.0135	.0011
5	-.0169	.0028	13	.0128	.0009
6	-.0058	.0027	14	.0034	.0009
7	.0008	.0027	15	.0152	.0004

7.4 Distribution of the Sample Whitening Correlations

In Section 7.3 a method is developed for choosing candidate models; the next step is model validation. Some of the methods that are suggested in subsequent sections require the distribution of $\hat{\rho}^w(h)$ under the null hypothesis. Recall, the null hypothesis states that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise.

Let $\{\epsilon(t) : t \in Z\}$ be a mean zero covariance stationary time series having spectral distribution function $F_n^w(\lambda)$, where $F_n^w(\lambda)$ is the spectral distribution function associated with the normalized whitening spectral density $f_n^w(\lambda)$. Then, by the spectral representation theorem

$$\begin{aligned}
\rho^w(h) &= \int_0^1 e^{-2\pi i \lambda h} f_n^w(\lambda) d\lambda \\
&= \int_0^1 e^{2\pi i \lambda t} e^{-2\pi i \lambda (t+h)} dF_n^w(\lambda) \\
&= E \left(\int_0^1 e^{2\pi i \lambda t} dZ(\lambda) \overline{\int_0^1 e^{2\pi i \mu (t+h)} dZ(\mu)} \right) \\
&= E \left(\epsilon(t) \overline{\epsilon(t+h)} \right) \\
&= E(\epsilon(t)\epsilon(t+h))
\end{aligned} \tag{7.26}$$

where $\{Z(\lambda), 0 \leq \lambda \leq 1\}$ is an orthogonal increment process. Expression (7.26) suggests that many time domain goodness of fit tests based on correlations may be adapted to the frequency domain. In practice we define the sample whitening correlations, $\hat{\rho}^w(h)$, to be the Fourier coefficients obtained from the discrete Fourier transform of the normalized sample whitening spectral density, recall,

$$\hat{\rho}^w(h) = \frac{1}{N} \sum_{h=0}^{N-1} f_n^{w*}(\lambda_j) \cos(2\pi h \lambda_j). \quad (7.27)$$

Additionally, assume that the process associated with $f^w(\lambda)$ is a Gaussian white noise process with variance σ^{2w} . Similar to Priestly (1981), the random variables $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(m)$ (m fixed) are approximately independent and identically distributed $N(0,1)$ random variables under the null hypothesis of white noise.

One method of verifying the asymptotic distribution of $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(m)$ under the null hypothesis is through the use of simulation. Recall, from Section 7.1 that under the null hypothesis we assume that $f_n^{w*}(\lambda_j)$ has the same distribution as the normalized periodogram of white noise. Therefore, the simulation can be conducted using $f_n^{\sim}(\lambda_j)$ instead of $f_n^{w*}(\lambda_j)$, $j = 0, \dots, N - 1$. The first step in the simulation is to generate M ($M \geq 1000$) independent and identically distributed samples of size N from a $N(0, \sigma^2)$ distribution. Next, for each sample, calculate the normalized sample spectral density $f_n^{\sim}(\lambda_j)$, $j = 0, \dots, N - 1$, where the λ_j 's are the Fourier frequencies. Then, for each sample, calculate $\hat{\rho}^w(h)$, $h = 1, \dots, \lfloor N/2 \rfloor$ using $f_n^{\sim}(\lambda_j)$, $j = 0, \dots, N - 1$ and the fast Fourier transform. Summary statistics are calculated and both the Shapiro-Wilk test of normality and the runs test for independence are performed. Simulations, for several different values of m, N , and σ^2 , confirm the claimed asymptotic distribution under the null hypothesis.

The distribution of the whitening correlations, under the null hypothesis, can also be demonstrated empirically through the direct use of the normalized sample

whitening sample spectral density, as in the following example. For this illustration, the model

$$\log f(\lambda) = .1 + 2 [.65 \cos(2\pi\lambda) - .4 \cos(2\pi 2\lambda) - .3 \cos(2\pi 3\lambda)] \quad (7.28)$$

is considered. First, 10,000 independent samples are generated with $N = 100$ and $f_n^{w*}(\lambda_j)$, $j = 0, \dots, N-1$ is formed using the true model in the denominator. Next, the quantities $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$ are computed using $f_n^{w*}(\lambda_j)$, $j = 0, \dots, N-1$, and the fast Fourier transform. The Shapiro-Wilk test for normality and the runs test for independence are then performed. The results of this empirical study confirm the claimed distribution and can be found in Table 13. Furthermore, simulations for sample sizes $N = 512$ and $N = 1000$, yield similar results.

Table 13. Descriptive statistics for $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$, and Shapiro-Wilk p-value. These values demonstrate that the random variables are $N(0,1)$ and that the variability of the mean and variance is small. Note that $N = 100$ with 10,000 repetitions and Q is the quantile for the Shapiro-Wilk p-value.

Descriptive Statistics	Value
mean(p-value: Shapiro-Wilk)	.50
var(p-value: Shapiro-Wilk)	.08
mean(mean($\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$))	0
var(mean($\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$))	.02
mean(var($\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$))	1.0
var(var($\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(N/2)$))	.08
mean(p-value: runs test)	.40
$Q(.75)$.75
$Q(.50)$.49
$Q(.25)$.25

7.5 Information Diagnostic

Recall, the goal of model validation is to determine if $f^w(\lambda)$ is equal to σ^{2w} (a constant), the spectral density of white noise. One possible test statistic for this hypoth-

esis is motivated by

$$I_\infty = - \int_0^1 \log f_n^w(\lambda) d\lambda, \quad (7.29)$$

where I_∞ is called the “information diagnostic”. A comprehensive discussion on time series and information can be found in Parzen (1983a) and Parzen (1992). Additionally, this quantity can be regarded as a measure of the distance from the whitening spectral density to a constant function. Now, under the null hypothesis $I_\infty = 0$ with larger values of this quantity being indicative of greater departures from white noise.

In order to use I_∞ as motivation for a test statistic, we first consider the following lemma.

Lemma 1. Let $\{Y(t)\}$ be a mean centered covariance stationary process with sample spectral density $f^{\sim}(\lambda)$. Furthermore, suppose the underlying process associated with $f^w(\lambda)$ is Gaussian white noise with variance σ^{2w} . Then

$$\log(\hat{\sigma}^{2w}) \xrightarrow{P} \log(\sigma^{2w}). \quad (7.30)$$

Proof:

Case 1: (N even)

$$\begin{aligned} \hat{\sigma}^{2w} &= \frac{1}{N} \sum_{j=0}^{N-1} f^{w*}(\lambda_j) = \frac{1}{N} \sum_{j=1}^{N-1} f^{w*}(\lambda_j) \\ &= \frac{1}{N} \left[2 \sum_{j=1}^{N/2-1} f^{w*}(\lambda_j) + f^{w*}(1/2) \right]. \end{aligned} \quad (7.31)$$

Therefore,

$$\begin{aligned} \mathbb{E}(\hat{\sigma}^{2w}) &= \frac{1}{N} [2(N/2 - 1)\sigma^{2w}\mathbb{E}(\chi_2^2/2) + \sigma^{2w}\mathbb{E}(\chi_1^2)] \\ &= \frac{1}{N} [2(N - 2)\sigma^{2w} + \sigma^{2w}] = \left(\frac{N - 1}{N}\right) \sigma^{2w} \end{aligned} \quad (7.32)$$

and

$$\begin{aligned}
\text{Var}(\hat{\sigma}^{2w}) &= \text{Var}\left[\frac{1}{N}\sum_{j=1}^{N-1}f^{w*}(\lambda_j)\right] \\
&= \frac{1}{N^2}\text{Var}\left[2\sum_{j=1}^{N/2-1}f^{w*}(\lambda_j)+f^{w*}(1/2)\right] \\
&= \frac{1}{N^2}4(N/2-1)\text{Var}[f^{w*}(\lambda_j)]+\frac{1}{N^2}\text{Var}[f^{w*}(1/2)] \\
&\quad +\frac{1}{N^2}4(N/2-1)(N/2-2)\text{Cov}[f^{w*}(\lambda_j),f^{w*}(\lambda'_j)] \\
&\quad +\frac{1}{N^2}2(2)(N/2-1)\text{Cov}[f^{w*}(\lambda_j),f^{w*}(1/2)] \\
&= \frac{\sigma^{4w}}{N^2}[4(N/2-1)+2]=\frac{2(N-1)\sigma^{4w}}{N^2}. \tag{7.33}
\end{aligned}$$

This implies that

$$\begin{aligned}
\text{MSE}(\hat{\sigma}^{2w}) &= \frac{2(N-1)\sigma^{4w}}{N^2}+\left(-\frac{1}{N}\sigma^{2w}\right)^2 \\
&= \frac{2\sigma^{4w}}{N}\longrightarrow 0 \text{ as } N\longrightarrow\infty. \tag{7.34}
\end{aligned}$$

Thus,

$$\hat{\sigma}^{2w}\xrightarrow{\text{MSE}}\sigma^{2w} \tag{7.35}$$

which implies

$$\log(\hat{\sigma}^{2w})\xrightarrow{p}\log(\sigma^{2w}). \tag{7.36}$$

Case 2: (N odd)

$$\begin{aligned}
\hat{\sigma}^{2w} &= \frac{1}{N}\sum_{j=0}^{N-1}f^{w*}(\lambda_j)=\frac{1}{N}\sum_{j=1}^{N-1}f^{w*}(\lambda_j) \\
&= \frac{1}{N}\left[2\sum_{j=1}^{(N-1)/2}f^{w*}(\lambda_j)\right]. \tag{7.37}
\end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}(\hat{\sigma}^{2w}) &= \frac{1}{N} \left\{ 2 \left(\frac{N-1}{2} \right) \mathbb{E}[f^{w*}(\lambda_j)] \right\} \\ &= \left(\frac{N-1}{N} \right) \sigma^{2w} \end{aligned} \quad (7.38)$$

and

$$\begin{aligned} \text{Var}(\hat{\sigma}^{2w}) &= \frac{1}{N^2} \text{Var} \left[2 \sum_{j=1}^{(N-1)/2} f^{w*}(\lambda_j) \right] \\ &= \frac{4}{N^2} \left(\frac{N-1}{2} \right) \text{Var}[f^{w*}(\lambda_j)] \\ &\quad + \frac{4}{N^2} \left(\frac{N-1}{2} - 1 \right) \left(\frac{N-1}{2} - 2 \right) \text{Cov}[f^{w*}(\lambda_j), f^{w*}(\lambda'_j)] \\ &= \frac{2}{N^2} [(N-1)\sigma^{4w} \text{Var}(\chi_2^2/2)] \\ &= \frac{2(N-1)\sigma^{4w}}{N^2}. \end{aligned} \quad (7.39)$$

This implies

$$\text{MSE}(\hat{\sigma}^{2w}) \longrightarrow 0 \text{ as } N \longrightarrow \infty. \quad (7.40)$$

Therefore,

$$\log(\hat{\sigma}^{2w}) \xrightarrow{p} \log(\sigma^{2w}) \quad (7.41)$$

and thus the lemma has been proven.

Furthermore, for $0 < \lambda_j < 1$, let

$$g(\lambda_j) = \log f^\sim(\lambda_j) + C_j \quad (7.42)$$

where $C_j = \gamma \approx .57721$ for $\lambda_j \neq .5$ and $C_j = (\log 2 + \gamma)/\pi$ for $\lambda_j = .5$. Then

$$g(\lambda_j) = \log f(\lambda_j) + \epsilon(\lambda_j) \quad (7.43)$$

where $\epsilon(\lambda_j) = \log[f^\sim(\lambda_j)/f(\lambda_j)] + C_j = \log(U_j) + C_j$ and

$$U_j \stackrel{d}{=} \begin{cases} \chi_2^2/2 & \lambda_j \neq .5 \\ \chi_1^2 & \lambda_j = .5 \end{cases} \quad (7.44)$$

Therefore,

$$\begin{aligned}\log f^\sim(\lambda_j) + C_j &= \log f(\lambda_j) + \log(U_j) + C_j \\ &= \log f(\lambda_j) + \epsilon(\lambda_j)\end{aligned}\tag{7.45}$$

which implies

$$\log f^\sim(\lambda_j) + C_j - \log f_{\theta^\wedge}(\lambda_j) = \log f(\lambda_j) + \epsilon(\lambda_j) - \log f_{\theta^\wedge}(\lambda_j)\tag{7.46}$$

or

$$\log f^{w*}(\lambda_j) + C_j = \log f^w(\lambda_j) + \epsilon(\lambda_j).\tag{7.47}$$

Now, since $C_j = \gamma$ for $\lambda_j \neq .5$ and $[(\log 2 + \gamma)/\pi]$ for $\lambda_j = .5$, it follows that

$$-\frac{1}{N} \sum_{j=1}^{N-1} C_j = - \left[\frac{(N-2)}{N} \gamma + \frac{1}{N} \left(\frac{\log 2 + \gamma}{\pi} \right) \right].\tag{7.48}$$

This implies

$$-\frac{1}{N} \sum_{j=1}^{N-1} C_j \longrightarrow -\gamma \text{ as } N \longrightarrow \infty\tag{7.49}$$

or

$$-\frac{1}{N} \sum_{j=1}^{N-1} C_j \xrightarrow{p} -\gamma.\tag{7.50}$$

Furthermore, $E[\epsilon(\lambda_j)] = 0$ and $\text{Var}[\epsilon(\lambda_j)] < \infty$ for all λ_j , see Wahba(1980). Therefore,

$$\frac{1}{N} \sum_{j=1}^{N-1} \epsilon(\lambda_j) \xrightarrow{\text{MSE}} 0\tag{7.51}$$

and thus we have the following theorem.

Theorem 1. Let I_∞^* be the empirical information diagnostic, where

$$I_\infty^* = -\frac{1}{N-1} \sum_{j=1}^{N-1} \log f_n^{w*}(\lambda_j) - \gamma.\tag{7.52}$$

Then

$$I_\infty^* \xrightarrow{p} I_\infty.\tag{7.53}$$

Proof:

By the trapezoid approximation to the integral, it follows that

$$\frac{1}{N} \sum_{j=0}^{N-1} \log f^w(\lambda_j) \longrightarrow \int_0^1 \log f^w(\lambda) d\lambda \text{ as } N \longrightarrow \infty. \quad (7.54)$$

Moreover,

$$\frac{1}{N} \sum_{j=0}^{N-1} \log f^w(\lambda_j) = \frac{1}{N} f^w(0) + \frac{1}{N} \sum_{j=1}^{N-1} \log f^w(\lambda_j) \quad (7.55)$$

which implies

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=0}^{N-1} \log f^w(\lambda_j) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^{N-1} \log f^w(\lambda_j) \\ &= \int_0^1 \log f^w(\lambda) d\lambda \end{aligned} \quad (7.56)$$

or

$$\frac{1}{N} \sum_{j=1}^{N-1} \log f^w(\lambda_j) \xrightarrow{p} \int_0^1 \log f^w(\lambda) d\lambda. \quad (7.57)$$

Now, for $0 < \lambda_j < 1$,

$$\log f^{w*}(\lambda_j) + C_j - \epsilon(\lambda_j) = \log f^w(\lambda_j). \quad (7.58)$$

Thus,

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^{N-1} \log f^w(\lambda_j) + \frac{1}{N} \sum_{j=1}^{N-1} C_j - \frac{1}{N} \sum_{j=1}^{N-1} \epsilon(\lambda_j) &= \frac{1}{N} \sum_{j=1}^{N-1} \log f^w(\lambda_j) \\ &\xrightarrow{p} \int_0^1 \log f^w(\lambda) d\lambda. \end{aligned} \quad (7.59)$$

Therefore,

$$-\frac{1}{N} \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma \xrightarrow{p} \int_0^1 \log f^w(\lambda) d\lambda \quad (7.60)$$

and so

$$\begin{aligned}
-\frac{1}{N-1} \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma &= -\left(\frac{N}{N-1}\right) \left(\frac{1}{N}\right) \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma \\
&\longrightarrow -\frac{1}{N} \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma \\
&\xrightarrow{p} \int_0^1 \log f^w(\lambda) d\lambda.
\end{aligned} \tag{7.61}$$

Hence,

$$-\frac{1}{N-1} \sum_{j=1}^{N-1} \log f_n^{w*}(\lambda_j) - \gamma \xrightarrow{p} \int_0^1 \log f_n^w(\lambda) d\lambda \tag{7.62}$$

since

$$\log(\hat{\sigma}^{2w}) \xrightarrow{p} \log(\sigma^{2w}). \tag{7.63}$$

and thus we have proven the theorem.

Therefore, in practice I_∞ can be approximated by

$$I_\infty^* = -\frac{1}{N-1} \sum_{j=1}^{N-1} \log f_n^{w*}(\lambda_j) - \gamma \tag{7.64}$$

and used as a test statistic. Critical values for this test statistic can be found using simulation. First, simulate the time series $\{Y(t)\}$ from N iid $N(0, 1)$ random variables, where N equals the length of the series. Next, setting $f_n^{w*}(\lambda_j) \equiv f_n^\sim(\lambda_j)$, the quantity in expression (7.7) is calculated. The reason for setting $f_n^{w*}(\lambda_j) = f_n^\sim(\lambda_j)$ is that under the null hypothesis $f_n^{w*}(\lambda_j)$ has the same distribution as the normalized periodogram of white noise. This procedure is repeated M ($M = 10,000$) times and selected quantiles are recorded. Table 14 contains simulation results for sample sizes $N = 100, 256, 512,$ and 1000 . Additionally, p-values can be found for any sample size by using the same simulation procedure and calculating the proportion of values greater than or equal to the computed test statistic (empirical information diagnostic).

Table 14. Critical values for the test statistic given in expression (7.64). Values found using simulation with 10,000 repetitions. All values have been rounded to three decimal places.

N	$Q(.05)$	$Q(.25)$	$Q(.50)$	$Q(.75)$	$Q(.90)$	$Q(.95)$	$Q(.975)$
100	-.184	-.086	-.009	.074	.153	.206	.249
256	-.113	-.049	-.002	.045	.092	.121	.146
512	-.080	-.036	-.003	.031	.064	.083	.101
1000	-.057	-.025	-.001	.024	.045	.059	.071

Several things are important to note regarding this testing procedure. First, theoretically the value of the test statistic increases as the departure from no memory increases. Second, any precision lost in estimating I_∞ using expression (7.64) is not problematic, since both the test statistic and the critical values obtained by simulation are calculated using the same estimate.

Again, assuming the process associated with $f^{w*}(\lambda)$ is Gaussian white noise, we can find the asymptotic distribution of I_∞^* using a calculation method similar to Davis and Jones (1968). First, from Bartlett and Kendall (1946), it follows that

$$E [\log (\chi_N^2/N)] = -\log(N/2) + \psi(N/2) \quad (7.65)$$

$$\text{Var} [\log (\chi_N^2/N)] = \psi'(N/2) \quad (7.66)$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions respectively. Furthermore, let

$$I_{u,\infty}^* = -\frac{1}{N-1} \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma \quad (7.67)$$

be the unnormalized empirical information diagnostic. Under the null hypothesis that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise, f^{w*} is assumed to have the same distribution as the periodogram of white noise. Thus, under the

null hypothesis,

$$\log f^{w*}(\lambda_j) \stackrel{d}{=} \begin{cases} \log(\chi_2^2) + \log(\sigma^{2w}) - \log 2 & \lambda_j \neq .5 \\ \log(\chi_1^2) + \log(\sigma^{2w}) & \lambda_j = .5. \end{cases} \quad (7.68)$$

Moreover, assuming the process associated with $f^w(\lambda)$ is Gaussian white noise, with variance σ^{2w} , it follows that $f^{w*}(\lambda_j)$ and $f^{w*}(\lambda'_j)$ are independent for $\lambda_j \neq \lambda'_j$ (for $0 < \lambda_j \leq .5$). Therefore, we have the following theorem.

Theorem 2. Assuming the process associated with f^{w*} is Gaussian white noise, I_∞^* has the following expectation and variance.

N even:

$$\mathbb{E}(I_\infty^*) = \psi\left(\frac{N-1}{2}\right) - \log\left(\frac{N}{2}\right) + \frac{\log 2 + \gamma}{(N-1)\pi} - \frac{\gamma}{N-1} \quad (7.69)$$

$$\text{Var}(I_\infty^*) = \frac{(2N-1)\pi^2}{6(N-1)^2} - \psi'\left(\frac{N-1}{2}\right) \quad (7.70)$$

N odd:

$$\mathbb{E}(I_\infty^*) = \psi\left(\frac{N-1}{2}\right) - \log\left(\frac{N}{2}\right) \quad (7.71)$$

$$\text{Var}(I_\infty^*) = \frac{2}{N-1} \left(\frac{\pi^2}{6}\right) - \psi'\left(\frac{N-1}{2}\right) \quad (7.72)$$

Proof: We present here only the proof for the case where N is even, as the case where N is odd follows by a similar argument. First,

$$\begin{aligned} \mathbb{E}(I_{u,\infty}^*) &= \mathbb{E}\left[-\frac{1}{N-1} \sum_{j=1}^{N-1} \log f^{w*}(\lambda_j) - \gamma\right] \\ &= -\frac{1}{N-1} \mathbb{E}\left[2 \sum_{j=1}^{N/2-1} \log f^{w*}(\lambda_j) + \log f^{w*}(1/2)\right] \\ &= -\frac{1}{N-1} \{(N-2)\mathbb{E}[\log(\chi_2^2) + \log(\sigma^{2w}) - \log 2]\} \\ &\quad -\frac{1}{N-1} \{\mathbb{E}[\log(\chi_1^2) + \log(\sigma^{2w})]\} - \gamma \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{N-1} \left[-(N-2)\gamma + (N-1) \log(\sigma^{2w}) - \left(\frac{\log 2 + \gamma}{\pi} \right) \right] - \gamma \\
&= -\log(\sigma^{2w}) - \frac{\gamma}{N-1} + \frac{1}{N-1} \left(\frac{\log 2 + \gamma}{\pi} \right). \tag{7.73}
\end{aligned}$$

Let $Y^*(\lambda_j) = \log f^{w^*}(\lambda_j)$, then

$$\begin{aligned}
\text{Var}(I_{u,\infty}^*) &= \text{Var} \left[-\frac{1}{N-1} \sum_{j=1}^{N-1} Y^*(\lambda_j) - \gamma \right] \\
&= \frac{1}{(N-1)^2} \text{Var} \left[\sum_{j=1}^{N-1} Y^*(\lambda_j) \right] \\
&= \frac{1}{(N-1)^2} \{ \text{Var} [Y^*(1/2)] + 4(N/2 - 1) \text{Var} [Y^*(\lambda_j)] \} \\
&\quad + \frac{4(N/2 - 1)(N/2 - 2)}{(N-1)^2} \text{Cov} [Y^*(\lambda_j), Y^*(\lambda'_j)] \\
&\quad + \frac{2(2)(N/2 - 1)}{(N-1)^2} \text{Cov} [Y^*(\lambda_j), Y^*(1/2)] \\
&= \frac{1}{(N-1)^2} \{ \text{Var} [\log(\chi_1^2)] + 2(N-2) \text{Var} [\log(\chi_2^2/2)] \} \\
&= \frac{1}{(N-1)^2} [\psi'(1/2) + 2(N-2)\psi'(1)] \\
&= \frac{1}{(N-1)^2} \left[\left(\frac{\pi^2}{2} \right) + 2(N-2) \frac{\pi^2}{6} \right] \\
&= \frac{(2N-1)\pi^2}{6(N-1)^2}; \tag{7.74}
\end{aligned}$$

since $\psi'(1/2) = \pi^2/2$ and $\psi'(1) = \pi^2/6$. Now,

$$\begin{aligned}
\hat{\sigma}^{2w} &= \frac{1}{N} \sum_{j=0}^{N-1} f^{w^*}(\lambda_j) \\
&= \frac{1}{N} \left[\sum_{j=1}^{N/2-1} 2f^{w^*}(\lambda_j) + f^{w^*}(1/2) \right] \\
&\stackrel{d}{=} \frac{\sigma^{2w}}{N} [\chi_{2(N/2-1)}^2 + \chi_1^2] = \frac{\sigma^{2w}}{N} \chi_{N-1}^2 \\
&= \frac{\sigma^{2w}(N-1)}{N} \left(\frac{\chi_{N-1}^2}{(N-1)} \right) \tag{7.75}
\end{aligned}$$

which implies

$$\log(\hat{\sigma}^{2w}) \stackrel{d}{=} \log(\sigma^{2w}) + \log\left(\frac{N-1}{N}\right) + \log\left[\frac{\chi_{N-1}^2}{(N-1)}\right]. \quad (7.76)$$

Therefore,

$$\begin{aligned} \mathbb{E}[\log(\hat{\sigma}^{2w})] &= \log(\sigma^{2w}) + \log\left(\frac{N-1}{N}\right) - \log\left(\frac{N-1}{2}\right) + \psi\left(\frac{N-1}{2}\right) \\ &= \log(\sigma^{2w}) - \log\left(\frac{N}{2}\right) + \psi\left(\frac{N-1}{2}\right) \end{aligned} \quad (7.77)$$

and

$$\text{Var}[\log(\hat{\sigma}^{2w})] = \psi'\left(\frac{N-1}{2}\right). \quad (7.78)$$

Thus,

$$\begin{aligned} \mathbb{E}(I_\infty^*) &= \mathbb{E}[\log(\hat{\sigma}^{2w}) + I_{u,\infty}^*] \\ &= \log(\sigma^{2w}) - \log\left(\frac{N}{2}\right) + \psi\left(\frac{N-1}{2}\right) \\ &\quad - \log(\sigma^{2w}) - \frac{\gamma}{N-1} + \frac{\log 2}{(N-1)\pi} - \frac{\gamma}{(N-1)\pi} \\ &= \psi\left(\frac{N-1}{2}\right) - \log(N/2) + \frac{\log 2 + \gamma}{(N-1)\pi} - \frac{\gamma}{N-1} \end{aligned} \quad (7.79)$$

and

$$\begin{aligned} \text{Var}(I_\infty^*) &= \text{Var}[I_{u,\infty}^* + \log(\hat{\sigma}^{2w})] \\ &= \text{Var}(I_{u,\infty}^*) + \text{Var}[\log(\hat{\sigma}^{2w})] + 2\text{Cov}[I_{u,\infty}^*, \log(\hat{\sigma}^{2w})]. \end{aligned} \quad (7.80)$$

Furthermore, by theorem B.4.1 Mardia et al. (1979),

$$\frac{f^{w*}(\lambda_j)}{\sum_{j=1}^{N-1} f^{w*}(\lambda_j)} \quad \text{and} \quad \sum_{j=1}^{N-1} f^{w*}(\lambda_j)$$

are independent. Therefore,

$$\begin{aligned}
\text{Var} \left[\log f^{w^*}(\lambda_j) - \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] &= \text{Var} \left[\log \left(\frac{f^{w^*}(\lambda_j)}{\sum_{j=1}^{N-1} f^{w^*}(\lambda_j)} \right) \right] \\
&= \text{Var} [\log f^{w^*}(\lambda_j)] + \text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \\
&\quad - 2\text{Cov} \left[\log f^{w^*}(\lambda_j), \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \quad (7.81)
\end{aligned}$$

and

$$\begin{aligned}
\text{Var} [\log f^{w^*}(\lambda_j)] &= \text{Var} \left[\log \left(\frac{f^{w^*}(\lambda_j)}{\sum_{j=1}^{N-1} f^{w^*}(\lambda_j)} \right) + \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \\
&= \text{Var} \left[\log \left(\frac{f^{w^*}(\lambda_j)}{\sum_{j=1}^{N-1} f^{w^*}(\lambda_j)} \right) \right] \\
&\quad + \text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right]. \quad (7.82)
\end{aligned}$$

This implies

$$\begin{aligned}
\text{Var} \left[\log \left(\frac{f^{w^*}(\lambda_j)}{\sum_{j=1}^{N-1} f^{w^*}(\lambda_j)} \right) \right] &= \text{Var} [\log f^{w^*}(\lambda_j)] - \text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \\
&= \text{Var} [\log f^{w^*}(\lambda_j)] + \text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \\
&\quad - 2\text{Cov} \left[\log f^{w^*}(\lambda_j), \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right]. \quad (7.83)
\end{aligned}$$

Hence,

$$2\text{Cov} \left[\log f^{w^*}(\lambda_j), \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] = 2\text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \quad (7.84)$$

which implies

$$\begin{aligned}
\text{Cov} \left[\log f^{w^*}(\lambda_j), \log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] &= \text{Var} \left[\log \sum_{j=1}^{N-1} f^{w^*}(\lambda_j) \right] \\
&= \psi' \left(\frac{N-1}{2} \right). \quad (7.85)
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Var}(I_\infty^*) &= \text{Var} [I_{u,\infty}^* + \log(\hat{\sigma}^{2w})] \\
&= \text{Var}(I_{u,\infty}^*) + \text{Var}[\log(\hat{\sigma}^{2w})] + 2\text{Cov}[I_{u,\infty}^*, \log(\hat{\sigma}^{2w})] \\
&= \text{Var}(I_{u,\infty}^*) + \text{Var}[\log(\hat{\sigma}^{2w})] + 2\text{Cov}[\log f^{w*}(\lambda_j), \log(\hat{\sigma}^{2w})] \\
&= \frac{(2N-1)\pi^2}{6(N-1)^2} + \psi' \left(\frac{N-1}{2} \right) - 2\psi' \left(\frac{N-1}{2} \right) \\
&= \frac{(2N-1)\pi^2}{6(N-1)^2} - \psi' \left(\frac{N-1}{2} \right). \tag{7.86}
\end{aligned}$$

Note that $E(I_\infty^*) \rightarrow \infty$ as $N \rightarrow \infty$. That is, I_∞^* is an asymptotically unbiased estimate of I_∞ . Also by an application of the central limit theorem I_∞^* is asymptotically normal. Therefore, asymptotic p-values can be found using the quantiles of the normal distribution with mean and variance given by the previous theorem. It should be noted that for small sample sizes we recommend calculating the p-values using simulation.

7.6 Whitening Spectral Distribution

Additional tests for goodness of fit can be formulated using the normalized whitening spectral distribution function. Tests of this sort were primarily developed by Bartlett. However, we introduce them in this section as they can be used in conjunction with the methods we develop for model validation when using EXP models. It should be noted that these tests have not previously been used in the context of model validation for EXP models.

The whitening spectral distribution function, $F^w(\lambda)$, $0 \leq \lambda \leq .5$, is defined by

$$F^w(\lambda) = 2 \int_0^\lambda f_n^w(\lambda') d\lambda' \tag{7.87}$$

where $f_n^w(\lambda)$ is the whitening spectral density defined in expression (7.2). In practice the sample whitening spectral distribution function is used when constructing tests

involving $F^w(\lambda)$ and is defined by

$$\begin{aligned} F^{w*}(\lambda_k) &= \frac{\sum_{j=1}^k f_n^{w*}(\lambda_j)}{\sum_{j=1}^M f_n^{w*}(\lambda_j)} \\ &= \frac{\sum_{j=1}^k f^{w*}(\lambda_j)}{\sum_{j=1}^M f^{w*}(\lambda_j)}, \quad k = 1, \dots, M \end{aligned} \quad (7.88)$$

where $M = [N/2] + 1$ and λ_j is the j^{th} Fourier frequency. Recall, under the null hypotheses, we assume $f^{w*}(\lambda)$ has the same distribution as the sample spectral density of white noise. Thus, under the null hypothesis $F^{w*}(\lambda_k)$, $k = 1, \dots, M$, has the same distribution as the sample spectral distribution of white noise.

Moreover, under the null hypothesis that $f^w(\lambda)$ is the spectral density of white noise, $F^w(\lambda) \equiv 2\lambda$, $\lambda \in [0, .5]$. Therefore, under H_0 the sample whitening spectral distribution function $F^{w*}(\lambda_k) \approx k/M$ where $k = 1, \dots, M$. Thus, a quick diagnostic for determining if $f^w(\lambda)$ is the spectral density of white noise is the graph of $F^{w*}(\lambda_k)$ versus λ_k . In Chapter IV suggestions are made for determining memory type using the graph of the sample spectral distribution function. Similarly in terms of the graph of $F^{w*}(\lambda_k)$

$$\begin{array}{lll} F^{w*} & \text{uniform} & \longrightarrow \text{no memory} \\ F^{w*} & \text{otherwise} & \longrightarrow \text{short memory} \\ F^{w*} & \text{has sharp jumps} & \longrightarrow \text{long memory.} \end{array} \quad (7.89)$$

Therefore, if the graph of F^{w*} is the graph of a uniform distribution, then $F^w(\lambda)$ is the spectral distribution function of a white noise process.

In addition to being a quick graphical diagnostic, $F^{w*}(\lambda_k)$ can be used in place of the sample spectral distribution function in Bartlett's test, which may be considered a time series analog of the Kolmogorov-Smirnov test. Under the null hypothesis that $f^w(\lambda)$ is the spectral density of white noise, it can be shown that

$F^{w*}(\lambda_1), \dots, F^{w*}(\lambda_{M-1})$ have the same distribution as the ordered random variables $U_1 \leq \dots \leq U_{M-1}$ of sample size $M - 1$ from a Uniform (0,1) population. Furthermore, the result

$$\lim_{M \rightarrow \infty} \Pr \left(\max_{1 \leq k \leq M} \sqrt{M} \left| U_k - \frac{k}{M} \right| \leq a \right) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2a^2 j^2} \equiv G(a) \quad (7.90)$$

can be used to determine whether $F^{w*}(\lambda_k)$ has a maximum deviation from the expected straight line $F^{w*}(\lambda_k) = k/M$ that is larger than reasonably expected under the null hypothesis. Let

$$B^* = \max_{1 \leq k \leq M} \sqrt{\frac{M}{2}} \left| F^{w*}(\lambda_k) - \frac{k}{M} \right| \quad (7.91)$$

and $F(b)$ be the cdf

$$F(b) = \Pr(B^* \leq b). \quad (7.92)$$

A complete discussion of Bartlett's test including the values of $F(b)$ for $b = .40, .41, \dots, 2.00$ can be found in Newton (1988).

Lastly, a compromise between the quick graphical diagnostic and the modified Bartlett's test is a graph of $F^{w*}(\lambda_k)$ along with 95% confidence bands. If $F^{w*}(\lambda_k)$ is the spectral distribution function of white noise then $F^{w*}(\lambda_k)$ will have an approximate 95% chance of falling within the lines $y = 2x \pm 1.36/\sqrt{M}$ where $M = [N/2] + 1$. Note that 1.36 is the ninety fifth percentile of Bartlett's statistic and that this plot is similar to the plot constructed by Newton (1988), differing only in that $F^{w*}(\lambda_k)$ is used in place of the sample spectral distribution function $F^\sim(\lambda_k)$.

7.7 EXP Model Validation by AIC

Let $Y(1), \dots, Y(N)$ be an observed, short memory, covariance stationary time series. Then $\{Y(t)\}$ can be approximated by a stationary autoregressive scheme of order p

$$Y(t) = \sum_{j=1}^p \alpha(j)Y(t-j) + Y^\nu(t), \quad t \in Z \quad (7.93)$$

where $\alpha(1), \dots, \alpha(p)$ are constants and the zeros of $1 - \alpha(1)z - \dots - \alpha(p)z^p$ are all outside the unit circle in the complex plane, and $\{Y^\nu(t)\}$ is a white noise innovation process. Therefore, an autoregressive process is a white noise process if and only if the process has order zero.

One method for testing for white noise is by using a data driven method of order selection. The test would reject the null hypothesis of white noise if and only if the selected order is greater than zero. According to Hart (1997), Parzen appears to have been the first person in statistics to propose using an order selection criterion as a form of model validation.

A widely used order determining criterion is Akaike's information criterion (AIC). AIC is defined by

$$\text{AIC}(k) = \log \hat{\sigma}_k^2 + \frac{2k}{N}, \quad k \geq 1 \quad (7.94)$$

where $\hat{\sigma}_k^2$ is the Yule-Walker estimate of the innovation variance for an approximating $\text{AR}(k)$ scheme. Possible definitions of $\text{AIC}(0)$ are $\text{AIC}(0) = 0$ or $\text{AIC}(0) = -1/N$. Shibata (1976) shows that if $\text{AIC}(0)$ is defined to be equal to zero then using AIC as a test of white noise yields an approximate probability of type I error equal to .29, for large N . Defining $\text{AIC}(0)$ equal to $-1/N$ slightly reduces the probability of type I error.

In this section we propose a new method for using AIC in conjunction with exponential models. This method uses $\hat{\rho}^w(h)$ to form autoregressive smoothers of $f^{w*}(\lambda)$. Note that this method can be used as a means of model validation as well as an aid in order selection, where the goal is to choose the smallest order m such that $f^{w,m}(\lambda)$ is the spectral density of white noise.

The first step in testing the null hypothesis is to calculate the sample whitening correlations, $\hat{\rho}^{w,m}(h)$ associated with the approximating $\text{EXP}(m)$ scheme. Recall,

$\rho^{w,m}(h)$ is defined as the Fourier transform of $f_n^{w,m}(\lambda)$. In practice, provided the sample size is moderate to large, $\rho^{w,m}(h)$ can be approximated by $\hat{\rho}^{w,m}(h)$ using $f_n^{w,m*}(\lambda_j)$ $j = 0, \dots, N - 1$ and the fast Fourier transform. The second step is to calculate the whitening residual variances, $\hat{\sigma}_{m,k}^{2w}$ $k = 1, 2, \dots$. Calculation of $\hat{\sigma}_{m,k}^{2w}$, $k = 1, 2, \dots$ given $\hat{\rho}^{w,m}(h)$ proceeds by a recursive algorithm called the Levinson-Durbin recursion, see Parzen (1983b).

Let $\pi(1), \dots, \pi(m)$ represent the partial correlation coefficients. Note that $|\pi(j)| < 1$, $j = 1, \dots, m$. Using the Levinson-Durbin recursion partial correlation coefficients, autoregressive coefficients, and residual variances can be determined from correlation coefficients as follows:

for $k = 1$

$$\begin{aligned}\alpha_1(1) &= \rho(1) \\ \sigma_1^2 &= 1 - \rho^2(1)\end{aligned}$$

while for $k = 2, 3, \dots, m$

$$\begin{aligned}\alpha_k(k) &= -\pi(k) = \frac{1}{\sigma_{k-1}^2} \sum_{j=0}^{k-1} \alpha_{k-1}(j) \rho(k-j) \\ \sigma_k^2 &= \sigma_{k-1}^2 \{1 - \pi^2(k)\} \\ \alpha_k(j) &= \alpha_{k-1}(j) - \pi(k) \alpha_{k-1}(k-j).\end{aligned}$$

To obtain the quantities $\hat{\sigma}_{m,k}^{2w}$ the estimates $\hat{\rho}^{w,m}(h)$ are substituted into the above recursion. In order to implement the AIC test define $\text{AIC}_m(0) \equiv -1/N$ and calculate

$$\text{AIC}_m(k) = \log \hat{\sigma}_{m,k}^{2w} + \frac{2k}{N}, \quad k = 1, 2, \dots, T. \quad (7.95)$$

where T is suitably chosen. If the minimum $\text{AIC}_m(k) = \text{AIC}_m(0) \equiv -1/N$ then there is insufficient evidence to conclude $f^{w,m*}(\lambda)$ does not behave like the sample spectral density of white noise.

7.8 Frequency Domain Data-Driven Portmanteau Test

Recall, one method of model validation is to test the null hypothesis that the sample whitening correlations behave like the correlations of white noise. This hypothesis can be tested using statistics directly constructed from $\hat{\rho}^w(h)$, the sample whitening correlations. In this section the data-driven portmanteau test (data-driven Q test), Hart (1997), is extended to the frequency domain by way of the sample whitening correlations. This test provides an additional means for assessing goodness of fit when using the EXP model. The original Q test is a time domain test of white noise based on the statistic

$$Q(m) = (N + 2) \sum_{j=1}^m \frac{\hat{\rho}^2(j)}{1 - j/N} \quad (7.96)$$

where $\hat{\rho}(j)$ is the autocorrelation at lag j , see expression (4.26). Under the hypothesis of white noise $\sqrt{N}\hat{\rho}(1), \dots, \sqrt{N}\hat{\rho}(m)$ (m fixed) are approximately independent and identically distributed $N(0, 1)$ random variables, see Priestly (1981). This implies that $Q(m)$ has a limiting χ_m^2 distribution under the null hypothesis.

The problem encountered in practice when using this test is the need to choose m , poor choices of m can reduce the power of the test. One method of addressing this problem is to choose the value of m based on the data and is implemented in the data-driven version of the portmanteau test developed by Hart (1997). Define

$$Q(\hat{m}) = (N + 2) \sum_{j=1}^{\hat{m}} \frac{\hat{\rho}^2(j)}{(1 - j/N)} \quad (7.97)$$

where \hat{m} is the maximizer of $\hat{R}(m)$ and

$$\hat{R}(m) = \begin{cases} 0, & m = 0 \\ \sum_{j=1}^m N\hat{\rho}^2(j) - 2m, & m = 1, \dots, N - 1. \end{cases} \quad (7.98)$$

Under appropriate regularity conditions, $Q(\hat{m})$ converges to the random variable T , defined by the following theorem.

Theorem 3. (Hart 1997) Let Z_1, Z_2, \dots be independent and identically distributed $N(0, 1)$ random variables, and define

$$S(k) = \sum_{j=1}^k (Z_j^2 - 2), \quad k = 1, 2, \dots$$

In addition, define $E_j(x)$ to be the event

$$\{0 < S(j) \leq x - 2j; S(k) \leq S(j), \quad k = 1, \dots, j - 1\} \quad j = 1, 2, \dots$$

and let n_x be the largest integer less than $x/2$. Under the hypothesis of white noise, $Q(\hat{m})$ converges in distribution to the following distribution function

$$P(T \leq x) = \begin{cases} 0, & x < 0 \\ .71, & 0 \leq x \leq 2 \\ .71 \left\{ 1 + \sum_{j=1}^{n_x} P[E_j(x)] \right\}, & x > 2 \end{cases} \quad (7.99)$$

Further details regarding this theorem and its' proof can be found in Hart (1997).

The nature of the distribution of $Q(\hat{m})$ is that the limiting size of a nonrandomized test is less than or equal to .29. However, in practice, this is not a problem since the size of a test is rarely chosen to be as large as .29. Note that the percentiles of the distribution given in Theorem 3 can be found through simulation, with a partial listing in Table 15.

In order to extend this test for the purpose of validating spectral models, we formulate the test in terms of $\hat{\rho}^w(h)$. Formulating the test statistic in terms of $\hat{\rho}^w(h)$ will produce a statistic that can be used to test the hypothesis that the sample whitening correlations behave like the correlations of white noise. This is equivalent to testing the hypothesis that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise, or that the estimated model provides a whitening transformation of the spectrum.

Table 15. Percentiles of the limiting distribution of $Q_w(\hat{m})$. Note that all values are found using simulation. Since the values are found via simulation, they are approximate and therefore not strictly monotone increasing.

x	$P(T \leq x)$	x	$P(T \leq x)$	x	$P(T \leq x)$	x	$P(T \leq x)$
2.5	.740	8.0	.886	14.25	.947	22.5	.978
2.6	.747	8.25	.885	14.5	.957	23.5	.978
2.7	.750	8.5	.894	14.75	.940	24.5	.974
2.8	.756	8.75	.893	15.0	.955	25.5	.990
2.9	.757	9.0	.901	15.25	.954	26.5	.983
3.0	.764	9.25	.902	15.5	.962	27.5	.987
3.25	.771	9.5	.902	15.75	.956	28.5	.980
3.5	.778	9.75	.905	16.0	.957	29.5	.991
3.75	.785	10.0	.915	16.25	.958	30.5	.993
4.0	.791	10.25	.913	16.5	.959	31.5	1.00
4.25	.801	10.5	.922	16.75	.965	32.5	.996
4.5	.813	10.75	.925	17.0	.965	33.5	.996
4.75	.815	11.0	.924	17.25	.965	34.5	.997
5.0	.824	11.25	.928	17.5	.963	35.5	.993
5.25	.831	11.5	.928	17.75	.964	36.5	.995
5.5	.835	11.75	.932	18.0	.976	37.5	.989
5.75	.837	12.0	.933	18.25	.973	38.5	.993
6.0	.846	12.25	.937	18.5	.970	39.5	.994
6.25	.851	12.5	.944	18.75	.971	40.5	1.00
6.5	.861	12.75	.938	19.0	.964	41.5	.997
6.75	.863	13.0	.946	19.25	.977	42.5	.998
7.0	.872	13.25	.950	19.5	.971	43.5	.987
7.25	.874	13.5	.950	19.75	.974	44.5	.997
7.5	.875	13.75	.943	20.5	.974	45.5	.995
7.75	.881	14.0	.949	21.5	.968	46.5	.993

Recall from Section 7.4, that under the assumption that the process associated with $f^w(\lambda)$ is Gaussian white noise with mean zero and variance σ^{2w} , the random variables $\sqrt{N}\hat{\rho}^w(1), \dots, \sqrt{N}\hat{\rho}^w(m)$ (m fixed) are approximately independent and identically distributed $N(0,1)$ random variables. Therefore, we define $Q_w(\hat{m})$ by

$$Q_w(\hat{m}) = (N + 2) \sum_{j=1}^{\hat{m}} \frac{[\hat{\rho}^w(j)]^2}{(1 - j/N)} \quad (7.100)$$

where \hat{m} is the maximizer of $\hat{R}_w(m)$ and

$$\hat{R}_w(m) = \begin{cases} 0, & m = 0 \\ \sum_{j=1}^m N [\hat{\rho}^w(j)]^2 - 2m, & m = 1, \dots, \lfloor N/2 \rfloor. \end{cases} \quad (7.101)$$

Under the appropriate regularity conditions, see Hart (1997), $Q_w(\hat{m})$ also converges to the random variable T given by Theorem 3 and can be used in the same manner as its time domain counterpart. Thus, the frequency domain portmanteau test $Q_w(\hat{m})$ provides another method for validating time series exponential models.

7.9 Data Analysis

In this section a data analysis is performed as a means of illustrating the previously developed methods. The data chosen for this purpose are the annual tree-ring indices of *Widdingtonia cedarbergensis*, measured as $\log(\text{index}/100)$. This data is initially analyzed by Zucchini and Hiemstra, in order to investigate the relationship between tree ring indices of *Widdingtonia cedarbergensis* and the annual rainfall totals at a neighboring location (as sited in Linhart and Zucchini, 1986). In the work of Zucchini and Hiemstra the logarithms of the index/100 are identified as an AR(2) process.

Subsequently, Linhart and Volkers (1985) use this data to illustrate their method of order selection. In their analysis both ARMA models and Bloomfield's EXP models are considered and an AR(2) model is chosen using their order selection criterion.

However, if their analysis is restricted to consider only EXP models then minimizing their order selection criterion yields an EXP(5) model.

Ultimately, in the analysis performed using our method, the model chosen is an EXP(3) model. This model captures the important features of our nonparametric wavelet estimate as well as satisfies the criterion of providing a whitening transformation to the spectral density. Furthermore, the AR representation associated with the chosen EXP(3) model is similar to the AR model chosen by Linhart and Volkers (1985).

The first step in analyzing this data is to construct a time plot of the data, see Figure 6. This plot suggests that the time series is stationary with constant variance. Additionally, using the diagnostics described in Chapter IV, it is determined that the time series is short memory.

The next step in our analysis is to estimate the parameters associated with candidate EXP(m) models. Using the wavelet method of parameter estimation described in Chapter VI, parameters for EXP(m) models, $m = 1, 2, \dots, M$ are estimated simultaneously. To facilitate choosing candidate models we also compute $ISE_m^{(d)}$, $m = 1, \dots, 9$. The values of $\hat{\theta}_m$ and $ISE_m^{(d)}$ can be found in Table 16.

Table 16. This table displays the parameter estimates and $ISE_m^{(d)}$ for the estimated model.

m	$\hat{\theta}_m$	$ISE_m^{(d)}$	m	$\hat{\theta}_m$	$ISE_m^{(d)}$
0	-3.562	NA	5	-.013	.0067
1	.520	.1639	6	-.002	.0067
2	.272	.0159	7	-.000	.0067
3	.062	.0081	8	.013	.0064
4	.022	.0071	9	-.006	.0063

Using the values of $\hat{\theta}$ and $ISE_m^{(d)}$ we initially choose EXP models of orders three

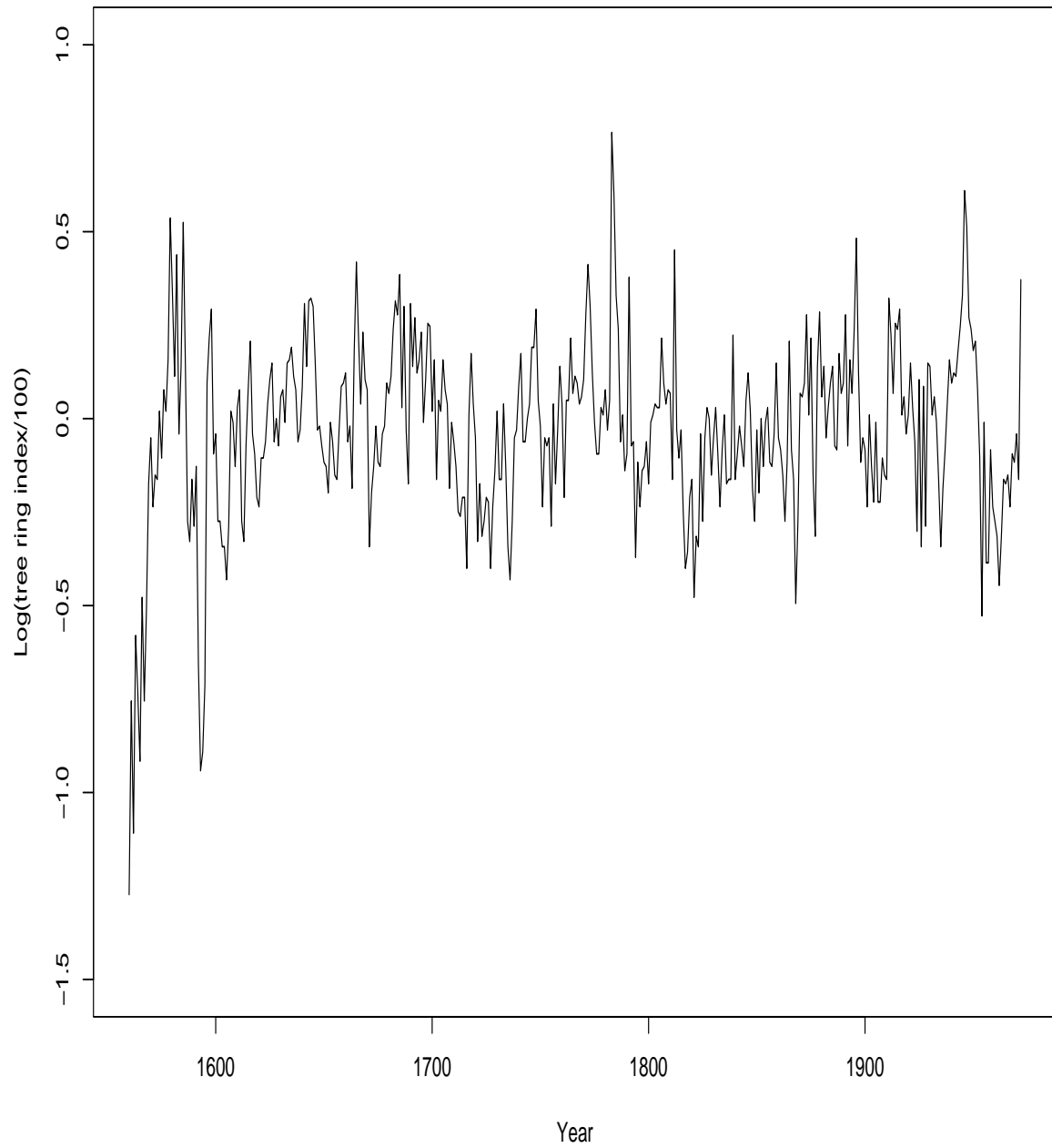


Figure 6. Time plot of logarithm of tree ring indices/100, for *Widdingtonia cedarbergensis*.

and four as candidate models. The EXP(3) model is given by

$$\log f(\lambda) = -3.562 + 2[.520 \cos(2\pi\lambda) + .272 \cos(2\pi 2\lambda) + .062 \cos(2\pi 3\lambda)] \quad (7.102)$$

A plot of the log periodogram $+.57721$ with the chosen EXP(3) model superimposed can be seen in Figure 7.

Using the previously developed methods of model validation we test the two equivalent hypotheses that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise and that $\hat{\rho}^w$ behave like the correlations of white noise for $h = 1, 2, \dots$. The first method of model validation used for this purpose is the information diagnostic described in Section 7.5. Recall, the test statistic used for this diagnostic is

$$I_\infty^* = - \left(\frac{1}{N-1} \right) \sum_{j=1}^{N-1} \log f_n^{w*}(\lambda_j) - \gamma \quad (7.103)$$

where $\gamma \approx .57721$. Using the EXP(3) model given by expression (7.102), $I_\infty^* = -.0192$. The p-value associated with this test statistic is found using simulation and equals .7999. Thus, we have strong evidence to conclude that $f^w(\lambda)$ is consistent with the spectral density of white noise.

Another diagnostic used, in order to validate the model given by expression (7.102), is the graph of the sample whitening spectral distribution with 95% confidence bands, see Figure 8. This graph clearly shows that $F^{w*}(\lambda_k)$, $k = 1, \dots, M$ is contained entirely in the 95% confidence bands $y = 2x \pm 1.36/\sqrt{M}$ where $0 \leq x \leq .5$ and $0 \leq y \leq 1$. Thus, at $\alpha = .05$ there is insufficient evidence to conclude that $f^w(\lambda)$ is not the spectral density of white noise.

The last two methods of model validation considered are the EXP model validation by AIC and the frequency domain data-driven portmanteau test. EXP model validation by AIC is described in Section 7.7, this method of model validation rejects the null hypothesis that $f^{w*}(\lambda)$ behaves like a white noise sample spectral density if and only if $\min \text{AIC}(k) < \text{AIC}(0) = -1/N$, for $k \geq 1$. Fitting the EXP(3)

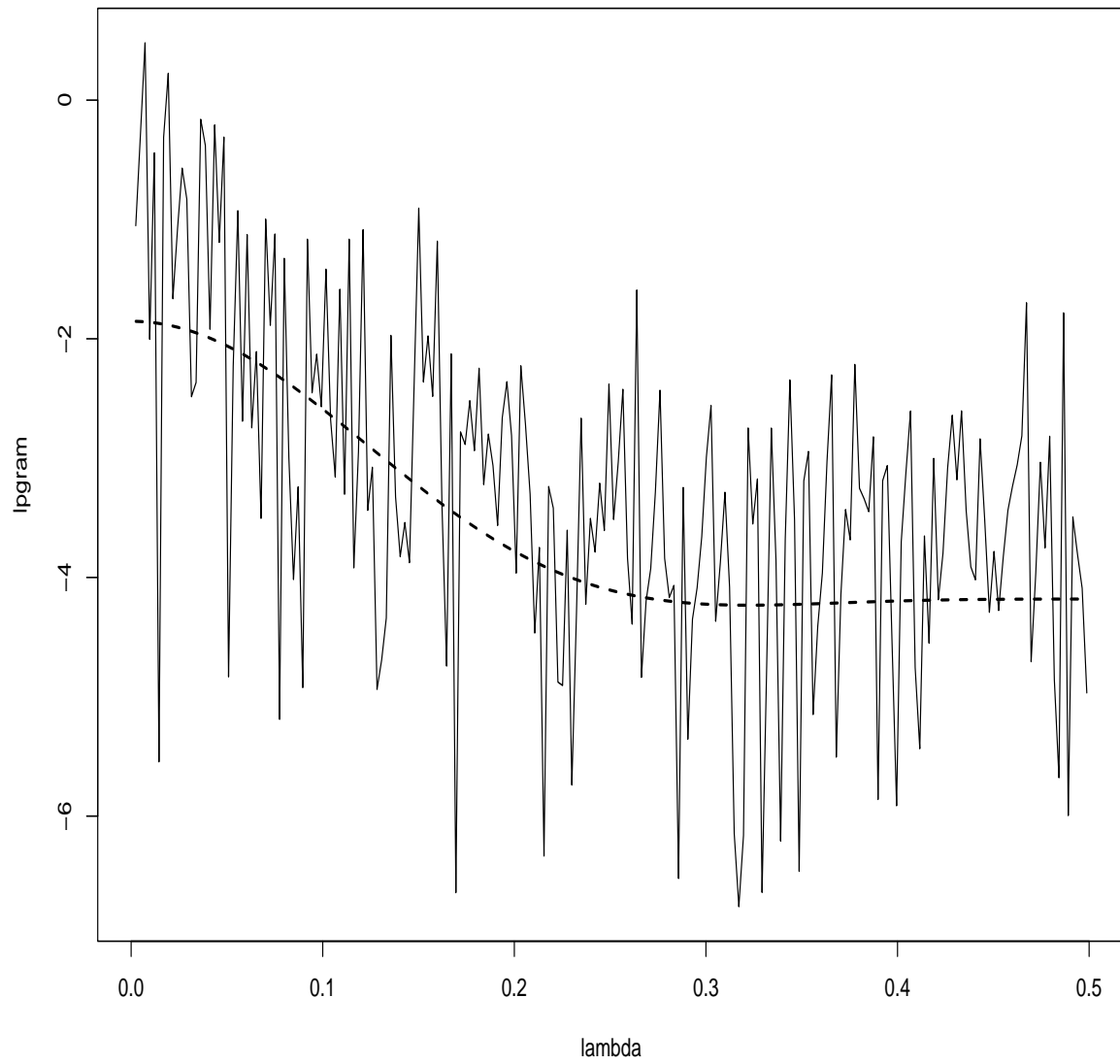


Figure 7. Plot of log periodogram $+.57721$ with fitted $EXP(3)$ model. The $EXP(3)$ model is given by expression (7.102).

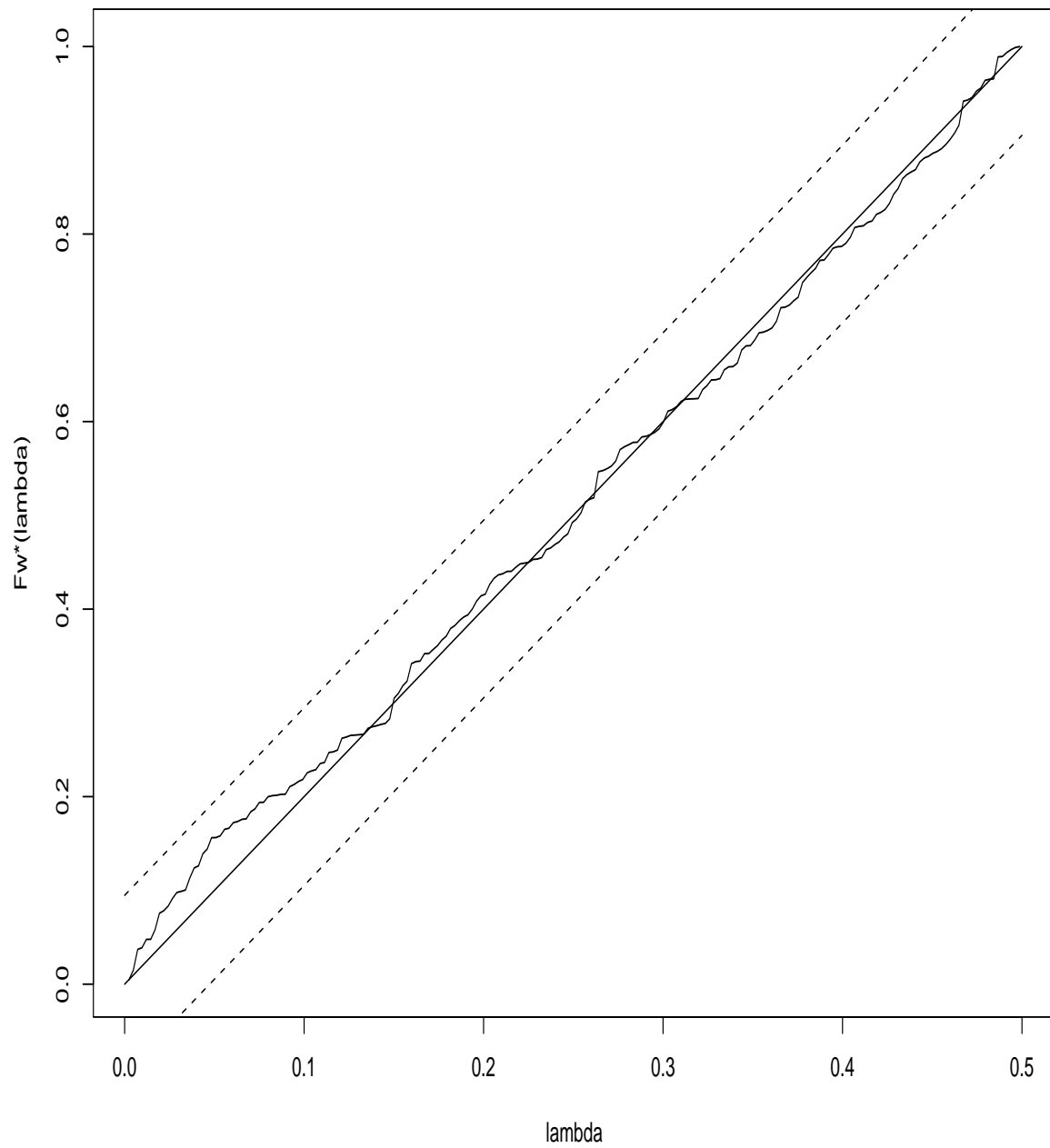


Figure 8. Plot of the sample whitening spectral distribution function with asymptotic 95% confidence bands from fitting the EXP(3) model.

model given by expression (7.102) results in a whitening spectral density for which $AIC(0) = -1/N < AIC(k)$, for all $k \geq 1$, see Table 17. Thus, there is insufficient

Table 17. Values of EXP model validation criteria using AIC. Note $AIC(0) = -1/N$ and $\min AIC(k) > AIC(0)$, ($k = 1, 2, \dots$).

k	$AIC(k)$	k	$AIC(k)$	k	$AIC(k)$	k	$AIC(k)$
0	-.0024	4	.0109	8	.0182	12	.0237
1	.0048	5	.0111	9	.0195	13	.0280
2	.0061	6	.0125	10	.0177	14	.0303
3	.0104	7	.0149	11	.0189	15	.0348

evidence to conclude $f^w(\lambda)$ is not the spectral density of white noise.

The frequency domain data-driven portmanteau test is described in Section 7.8. The value of the test statistic $Q_w(\hat{m})$ is equal to .00375 and is found using $\hat{m} = 1$. The p-value associated with $Q_w(\hat{m})$ is .71. Therefore, this test again suggests that, $f^w(\lambda)$ is consistent with the spectral density of white noise.

All of the methods of model validation fail to reject the null hypothesis that $f^{w*}(\lambda)$ behaves like the sample spectral density of white noise. Thus, we conclude that the model given by expression (7.102) provides an adequate fit. The associated AR representation is given by

$$Y(t) - .520Y(t-1) - .137Y(t-2) + .057Y(t-3) + .035Y(t-4) + \dots = Y^\nu(t) \quad (7.104)$$

with innovation variance equal to .028. The AR(2) model chosen by Linhart and Volkens (1985) is given by

$$Y(t) - .51Y(t-1) - .18Y(t-2) = Y^\nu(t) \quad (7.105)$$

and has innovation variance equal to .03584. The model we chose is similar to the model chosen by Linhart and Volkens (1985) differing primarily in the value of the innovation variance and the spectrum near the zero frequency, with our model being

slightly smaller in both cases. Note that residual (innovation) variance estimates the one-step ahead prediction error variance which would be achieved using the fitted model. Thus our model not only provides a whitening transformation of the spectrum, but also has slightly smaller innovation variance than the model obtained by Linhart and Volkers (1985).

CHAPTER VIII

AREXP MODELS

8.1 Overview and Algorithm

In this chapter we introduce a new iterative model for time series analysis and forecasting. The models considered are called AREXP (Autoregressive Exponential) models because the model fitted to a long memory time series $\{Y(t)\}$ is based on time series analysis of AR schemes obtained using a “best-lag” nonstationary autoregression followed by short memory EXP modelling. This new model was proposed for investigation by Parzen (2002) and is based on the ARARMA (ARAR) model originally developed by Newton and Parzen (1984) and Parzen (1982). Furthermore, the ARAR algorithm is described in Brockwell and Davis (2002).

The first step in AREXP modelling is to fit an AR model which transforms the time series $\{Y(t)\}$ to short memory residuals $\{Y^s(t)\}$ whose unnormalized spectral density can be computed by

$$f^{s\sim}(\lambda) = |A(\lambda)|^2 f^{\sim}(\lambda) \quad (8.1)$$

where

$$A(\lambda) = \sum_{j=0}^k a_{j,k} \exp(2\pi i \lambda) \quad (8.2)$$

The second step is to fit a log exponential model to $f^{s\sim}(\lambda)$. One major benefit of fitting the model iteratively is in terms of model validation. More specifically, frequency domain model validation for long memory processes is rather difficult. However, in this context, we only need to determine if the log exponential model fit to $f^{s\sim}(\lambda)$ provides a whitening transformation of the spectrum. Thus, model validation is reduced to the context of short memory.

Given a data set $\{Y(t), t = 1, 2, \dots, N\}$, the first step is to decide if whether the underlying process is “long memory”, and if so apply a memory shortening transformation before attempting to fit an EXP model. The technique used for transforming the time series to short memory is called “memory shortening” and was pioneered by Parzen (1982). In order to use this technique we must first consider the following differencing operations

1. $\tilde{Y}(t) = Y(t) - \hat{\phi}(\hat{\tau})Y(t - \hat{\tau})$
2. $\tilde{Y}(t) = Y(t) - \hat{\phi}_1 Y(t - 1) - \hat{\phi}_2 Y(t - 2)$.

Then with the aid of a five step algorithm, we classify the time series $\{Y(t)\}$ and take one of the following courses of action:

- L - Declare $\{Y(t)\}$ to be long memory and form the transformed time series $\{\tilde{Y}(t)\}$ using (1).
- M - Declare $\{Y(t)\}$ to be moderately long memory and form $\{\tilde{Y}(t)\}$ using (2).
- S - Declare $\{Y(t)\}$ to be short memory.

If the alternative L or M is chosen, then the transformed series $\{\tilde{Y}(t)\}$ is again checked. If the transformed series is again found to be long-memory or moderately long-memory, then a further transformation is performed. The process continues until the transformed series is classified as short-memory. In order to display the algorithm for deciding L, M, and S, some definitions are needed. Consider the “best-lag” nonstationary autoregression given by

$$\tilde{Y}(t) = Y(t) - \hat{\phi}(\hat{\tau})Y(t - \hat{\tau}). \quad (8.3)$$

For each τ , $\hat{\phi}(\tau)$ is chosen to minimize over ϕ

$$\sum_{t=\tau+1}^N [Y(t) - \phi(\tau)Y(t - \tau)]^2. \quad (8.4)$$

Parzen (1981) refers to this quantity as the “squariance”, which is the sum of square of the residuals after fitting the “best-lag” nonstationary autoregression. The next step is to then choose the lag $\hat{\tau}$ to minimize over τ

$$\text{Err}(\tau) = \frac{\sum_{t=\tau+1}^N [Y(t) - \hat{\phi}(\hat{\tau})Y(t - \hat{\tau})]^2}{\sum_{t=\tau+1}^T Y^2(t)}. \quad (8.5)$$

This quantity is the normalized mean squared error. The five step algorithm proceeds as follows:

1. Compute $\hat{\phi}(\tau)$ and $\text{Err}(\tau)$ for $\tau = 1, 2, \dots, M_\tau$ (where M_τ is suitably chosen) and choose $\hat{\tau} = \min_{\tau} \text{Err}(\tau)$.
2. If $\text{Err}(\hat{\tau}) \leq 8/n$ go to L.
3. If $\hat{\phi}(\hat{\tau}) \geq .9$ and $\hat{\tau} > 2$ go to L.
4. If $\hat{\phi}(\hat{\tau}) \geq .9$ and $\hat{\tau} = 1$ or 2 determine the values of $\hat{\phi}_1$ and $\hat{\phi}_2$ that minimize

$$\sum_{t=3}^N [Y(t) - \phi_1 Y(t-1) - \phi_2 Y(t-2)]^2; \quad (8.6)$$

then go to M.

5. If $\hat{\phi}(\hat{\tau}) < .9$ go to S.

Intuitively, these steps determine how well the model given by the “best-lag” nonstationary autoregression does in predicting the current value of our time series. This information is then used to determine the transformation needed. Note that the above values are chosen because they perform well for a wide variety of time series. Thus, the values displayed here are flexible and can be adjusted. In fact, the algorithm described in Brockwell and Davis (2002) contains slightly modified cut off values. Finally, in order to proceed from S we use the following three step algorithm:

1. Form

$$\log f_{\theta^\wedge, m}(\lambda) = \hat{\theta}_0 + 2 \sum_{k=1}^m \hat{\theta}_k \cos(2\pi k\lambda), \quad m = 1, \dots, M$$

where the $\hat{\theta}_k$'s are the estimated model parameters.

2. Choose candidate models of order $m = \hat{m}$ using a suitable order selection method.
3. Using model validation criteria described previously test the hypothesis $H_0 : f^{w, \hat{m}}(\lambda) = \sigma^{2w}$, where $f^{w, \hat{m}}(\lambda) = f^{s\sim}(\lambda) / f_{\theta^\wedge, m}(\lambda)$.

8.2 Wolf Sunspot Data

In this section we illustrate the use of the AREXP model by analyzing the Wolf Sunspot data from 1700-1988. In order to fit the AREXP model we employ the maximum likelihood parameter estimation method for short memory EXP models. It should be noted that any of the estimation methods previously described would be reasonable, since fitting the AR model in the first step of the procedure acts as a pre-whitening filter. Therefore, in practice, when fitting the EXP portion of the model we are not usually confronted with the problem of bias due to leakage. Moreover, it was verified specifically that in this case the EXP model is being fit to a raw estimator not suffering from leakage. The advantage of using maximum likelihood estimation in this setting is that the Linhart-Volkers criterion can be used as an aid in selecting the orders of candidate models. This may be desirable since one of the goals of the AREXP model is forecasting.

The first step of our process compute $\hat{\phi}(\tau)$ and $\text{Err}(\tau)$, this was done for $\tau = 1, \dots, 15$, see Table 18. From this information we determined that $\hat{\tau} = 1$, $\text{Err}(\hat{\tau}) = .140$, $8/n = .028$ and $\hat{\phi}(\hat{\tau}) = .931$. Thus, we proceed to step 4 of the algorithm and

Table 18. The values of $\hat{\phi}(\tau)$ and $Err(\tau)$ for values of $\tau = 1, 2, \dots, 15$.

τ	$\hat{\phi}(\tau)$	$Err(\tau)$	τ	$\hat{\phi}(\tau)$	$Err(\tau)$	τ	$\hat{\phi}(\tau)$	$Err(\tau)$
1	.931	.140	6	.464	.787	11	.874	.275
2	.786	.387	7	.546	.706	12	.805	.386
3	.626	.612	8	.670	.562	13	.694	.544
4	.505	.747	9	.795	.389	14	.585	.677
5	.446	.803	10	.872	.274	15	.506	.758

find ϕ_1 and ϕ_2 that minimize expression (8.6). We classify our series as moderately long-memory and perform the appropriate memory shortening transformation. This series only required one transformation in order to be classified as short memory. The model fitted to the Sunspot series is given by

$$\hat{f}(\lambda) = |A(\lambda)|^{-2} f_{\theta^\wedge}(\lambda) \quad (8.7)$$

where $A(\lambda) = 1 - 1.488 \exp(2\pi i \lambda) + .598 \exp(2\pi i 2\lambda)$ and $f_{\theta^\wedge}(\lambda)$ is an EXP(12) with parameter values found in Table 19. Lastly, a plot of the log sample spectral density

Table 19. The values of the EXP(12) model parameters for the AREXP model fit to the Wolf Sunspot data 1700-1988. Note that $\hat{\theta}_0 = 5.456$.

m	$\hat{\theta}_m$	m	$\hat{\theta}_m$	m	$\hat{\theta}_m$	m	$\hat{\theta}_m$
1	-.249	4	.069	7	-.043	10	.285
2	-.184	5	-.169	8	.014	11	.236
3	-.273	6	-.078	9	.267	12	.126

with the fitted log AREXP model and a plot of the whitening sample spectral distribution function with asymptotic 95% confidence bands can be found in Figures 9 and 10 respectively. Note that the plot of the whitening spectral distribution function indicates that the fitted model provides a whitening transformation of the spectrum. Moreover, the estimated AREXP model contains a primary peak at frequency .090, corresponding to the famous sunspot cycle of 11.1 years and a secondary peak at frequency .193, corresponding to a cycle of 5.18 years. Lastly, in order to use this model

for the purpose of forecasting we would calculate the AR representation associated with the EXP portion of the model. However, further research is needed for finding the order of candidate models when the primary purpose is forecasting. More specifically we need a method of choosing candidate models that provide a more “gentle” whitening transformation of the spectrum.

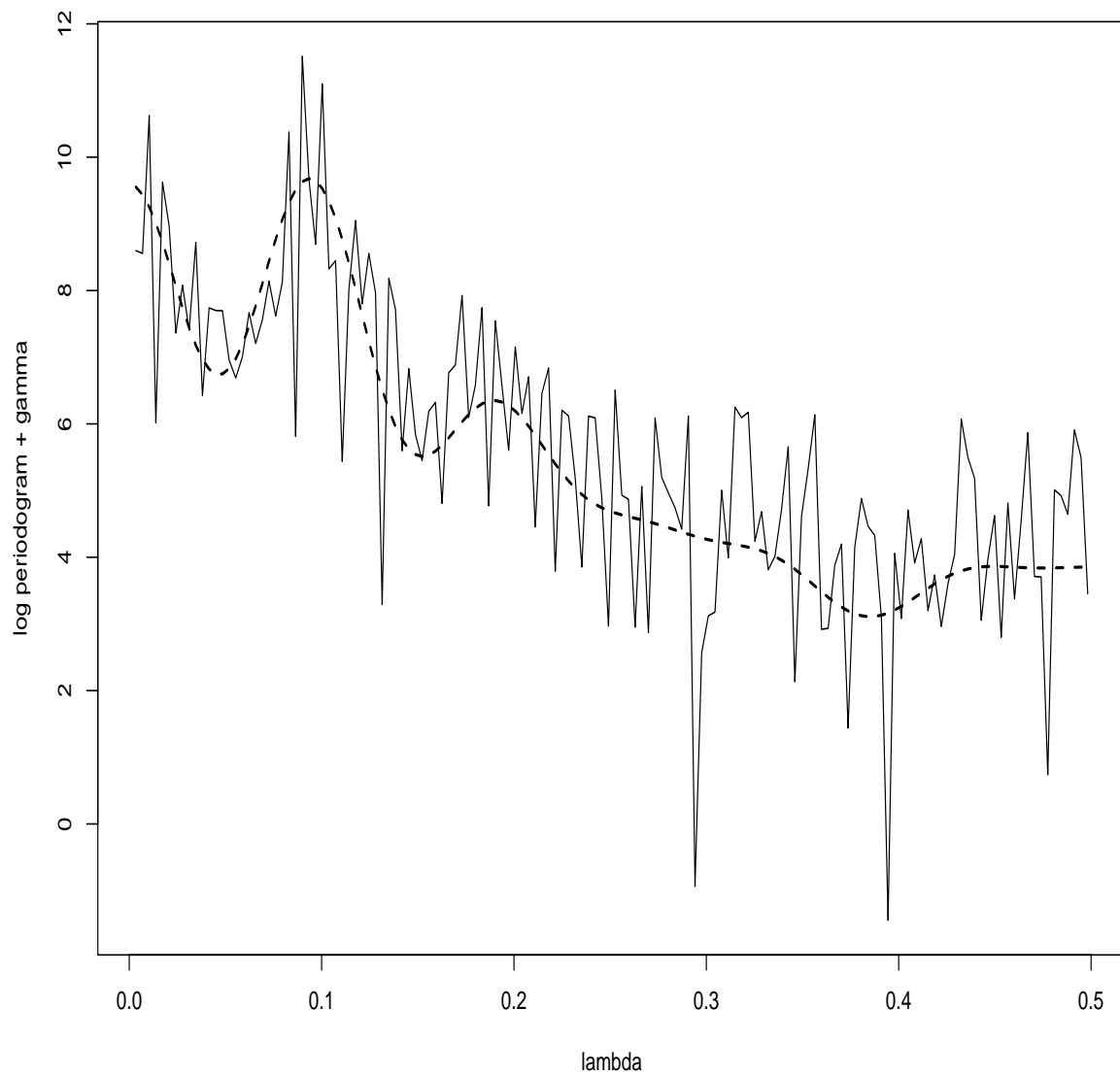


Figure 9. Plot of log periodogram $+ .57721$ with fitted log AREXP model for the Wolf Sunspot data 1700-1988.

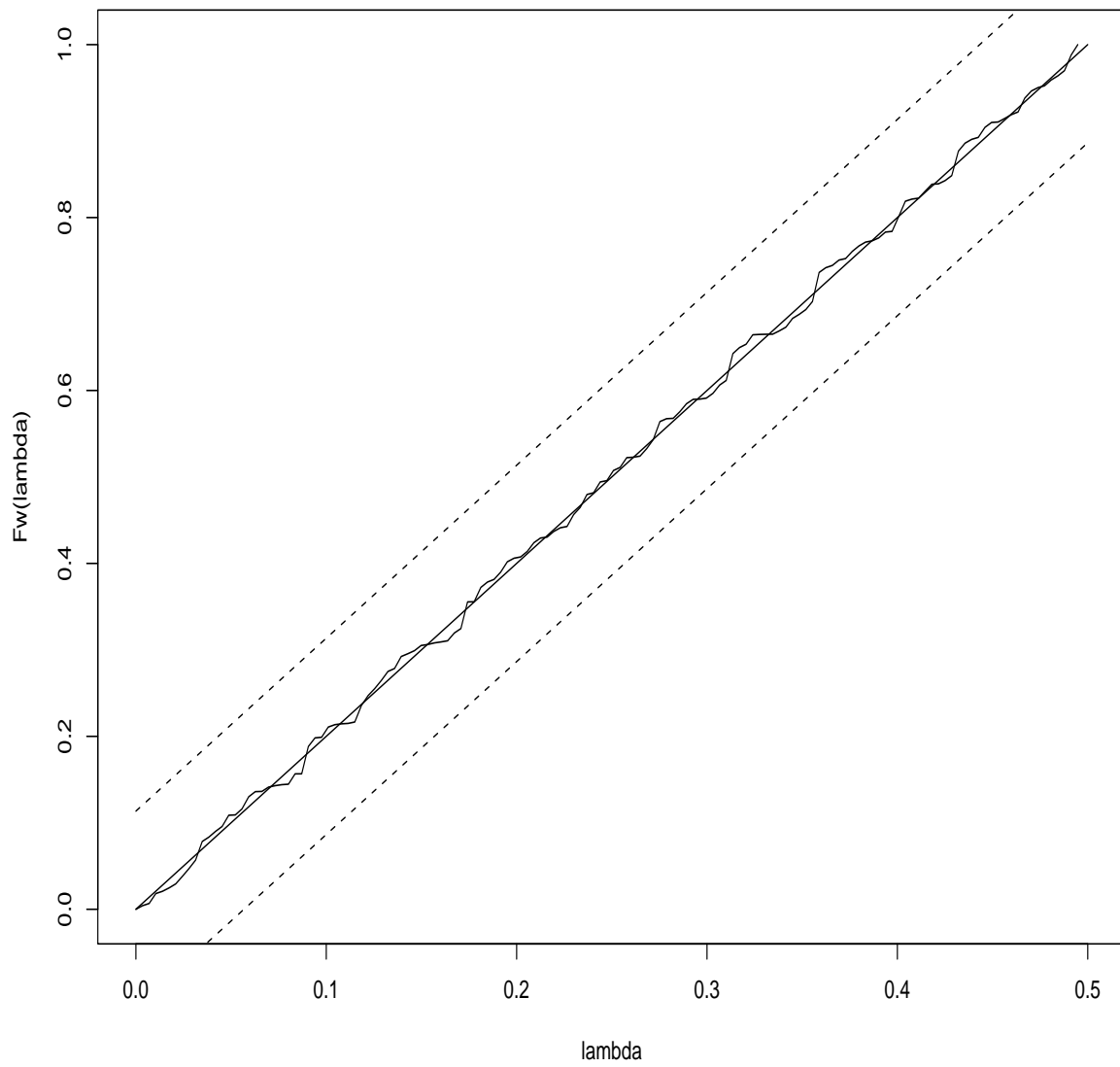


Figure 10. Plot of the sample whitening spectral distribution function with asymptotic 95% confidence bands from fitting the AREXP model.

CHAPTER IX

CONCLUSION

9.1 Concluding Remarks

The exponential model of Bloomfield(1973) has become increasingly important partly due to its recent applications to long memory time series. However, practical methods for using the EXP model in the analysis of observed time series data have been limited. In this dissertation, we have developed several new statistical procedures that improve the utility and robustness of the EXP model.

First, we have introduced a new method of simulating data from the $EXP(m)$ model. This method is based on simulating from a truncated $MA(\infty)$ representation of the $EXP(m)$ scheme, as provided by Pourahmadi's formula. This method has the benefit of being less computationally intensive and less error prone than existing methods, while being extremely easy to implement.

Next, we proposed a new nonparametric approach to parameter estimation based on wavelets. This method extends the work of Walden et al. (1998), allowing for parameter estimation. The advantage of this method is that, for many spectra, parameter estimates are less susceptible to biases associated with methods of parameter estimation based directly on the raw periodogram.

Additionally, in this dissertation, several new methods were introduced for the validation of spectral models. These methods test the hypothesis that the estimated model provides a whitening transformation of the spectrum; this is equivalent to the time domain notion of producing a model with white noise residuals. The new methods introduced include EXP model validation by AIC, a frequency domain data driven portmanteau test, and the information diagnostic.

9.2 Problems for Further Study

The AREXP models introduced in Chapter VIII provide an excellent algorithmic approach to spectral modelling of long memory time series. However, in the context of forecasting the model may fit “too well” and not provide a small mean squared error of prediction for large forecast horizons. Thus, one open problem is to develop alternative order selection criteria for fitting the EXP portion of the iterated model. The alternative order selection criteria would provide a more “gentle” whitening transformation of the spectrum.

Another problem for further research is the application of exponential models in the context of periodically correlated time series. The proposed research would examine the role of exponential models to spectral estimation for multiple time series. This research would provide an alternative to using periodic autoregressions for multiple spectral estimation.

Additionally, in this dissertation, model validation procedures are developed for spectral estimates not suffering from bias due to leakage. One problem for further research is the development of a band limited frequency domain goodness of fit test. Band limited tests can be restricted to consider only frequency bands of the spectral estimate not suffering from leakage. This testing can be especially useful in several different applications; for example, in some physics applications only the low frequency values of the spectrum are important.

Lastly, an open problem for research is the development of frequency domain goodness of fit tests based on multitaper spectral estimation. The developed tests would apply to both short memory and long memory time series. In this context, the use of multitapers could facilitate superior estimates of the innovation variance and residual correlations.

REFERENCES

- Bartlett, M. S. and Kendall, D. G. (1946), "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *Suppl. J. Roy. Statist. Soc.*, 8, 128-138.
- Beran, J. (1993), "Fitting Long-Memory Models by Generalized Linear Regression," *Biometrika* 80, 817-822.
- (1994), *Statistics for Long-Memory Processes*, Boca Raton, FL: Chapman & Hall/CRC.
- Bloomfield, P. (1973), "An Exponential Model for the Spectrum of a Scalar Time Series," *Biometrika* 60, 217-226.
- Brockwell, P. and Davis, R. (1996), *Time Series: Theory and Methods*, New York: Springer-Verlag.
- (2002). *Introduction to Time Series and Forecasting*, New York: Springer-Verlag.
- Cameron, M. and Turner, R. (1987), "Fitting Models to Spectra Using Regression Packages," *Applied Statistics* 36, 47-57.
- Davies, R. B. and Harte, D. S. (1987), "Tests for Hurst Effect," *Biometrika*, 74, 95-101.
- Davis, H. and Jones, R. (1968), "Estimation of the Innovation Variance of a Stationary Time Series," *Journal of the American Statistical Association*, 63, 141-149.
- Donoho, D. L. and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425-455.
- Hart, J. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer-Verlag.

- Hurvich, C. (2002), "Multistep Forecasting of Long Memory Series Using Fractional Exponential Models," *The International Journal of Forecasting*, 18, 167-179.
- Janacek, G. (1982), "Determining the Degree of Differencing for Time Series via the Log Spectrum," *Journal of Time Series Analysis* 3, 177-183.
- Jansen, M., Malfait, M., and Bultheel, A. (1997). "Generalized Cross Validation for Wavelet Thresholding." *Signal Processing*, 56, 33-44.
- Linhart, H. and Volkers, P. (1985), "On a Criterion for the Selection of Models for Stationary Time Series," *Metrika* 32, 181-196.
- Linhart, H. and Zucchini, W. (1986), *Model Selection*, New York: Wiley.
- Mardia, K. V. and Kent, J. T. and Bibby, J. M. (1979), *Multivariate Analysis*, San Diego, CA: Academic Press.
- Nason, G. (1996), "Wavelet Shrinkage by Cross Validation," *Journal of the Royal Statistical Society, Series B, Methodological*, 58, 463-479.
- Newton, H. J. (1988), *TIMESLAB: A Time Series Analysis Laboratory*, Pacific Groove, California: Wadsworth.
- Newton, H. J. and Parzen, E. (1984), "Forecasting and Time Series Models Types of 111 Economic Time Series," in *The Forecasting Accuracy of Major Time Series Methods*, eds. S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, New York: Wiley, pp. 267-288.
- Parzen, E. (1981), "Time Series Model Identification and Prediction Variance Horizon," in *Applied Time Series Analysis II*, New York: Academic Press.
- (1982), "ARARMA Models for Time Series Analysis and Forecasting," *Journal of Forecasting*, 1, 67-82.
- (1983a), "Time Series Model Identification By Estimating Information." in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. S. Karlin,

- T. Amemiya, L. Goodman, New York: Academic Press, pp. 279-298.
- (1983b), “Autoregressive Spectral Estimation.” in *Handbook of Statistics Volume 3: Time Series in the Frequency Domain*, Amsterdam; New York: North-Holland/Elsevier pp. 221-247.
- (1992), “Time Series, Statistics, and Information,” in *New Directions in Time Series Analysis, Part I*, vol. 45 of *IMA Vol. Math. Appl.*, eds. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, M. Taqqu, New York: Springer-Verlag, pp. 265-287.
- (1993), “Stationary Time Series Analysis Using Information and Spectral Analysis,” in *Developments in Time Series Analysis. In Honour of Maurice B. Priestly*, ed. T.S. Rao, London: Chapman & Hall, pp. 139-148.
- (2002), “Time Series Analysis, SIEVE, Memory, and Exponential Spectral Model.” Presented at the 40th Anniversary Conference, Department of Statistics Texas A&M University.
- Percival, D. B. (1992), “Simulating Gaussian Random Processes With Specified Spectra,” in *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface*, Fairfax Station: VA, Interface Foundation of North America, pp. 534-538.
- Percival, D. B. and Walden, A. T. (1993), *Spectral Analysis for Physical Applications. Multitaper and Conventional Univariate Techniques*, New York: Cambridge University Press.
- (2000), *Wavelet Methods for Time Series Analysis*, New York: Cambridge University Press.
- Pourahmadi, M. (1983), “Exact Factorization of the Spectral Density and Its Application to Forecasting and Time Series Analysis,” *Communications in Statistics, Part A – Theory and Methods* 12, 2085-2094.

- Priestly, M. B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press.
- Riedel, K. S. and Sidorenko, A. (1995), "Minimum Bias Multiple Taper Spectral Estimation," *IEEE Transactions on Signal Processing*, 43, 188-195.
- Shibata, R. (1976), "Selection of the Order of Autoregressive Models by Akaike's Information Criterion," *Biometrika*, 63, 117-126.
- Vidakovic, B. (1993), *Statistical Modeling by Wavelets*, New York: Wiley.
- Wahba, G. (1980), "Automatic Smoothing of the Log Periodogram," *Journal of the American Statistical Association*, 75, 122-132.
- Walden, A., McCoy, E., and Percival, D. (1998), "Spectrum Estimation by Wavelet Thresholding of Multitaper Estimators," *IEEE Transactions on Signal Processing*, 46, 3153-3165.
- Walker, A.M. (1964), "Asymptotic Properties of Least Squares Estimates of Parameters of the Spectrum of a Stationary Non-Deterministic Time Series," *Journal of Australian Mathematical Society* 4, 363-384.

VITA

Scott Holan was born in Skokie, Illinois. He received a Bachelor of Science degree in mathematics in 1997 and a Master of Science degree in mathematics in 1999, both from the University of Illinois at Chicago. In May 2004 he received a Ph.D. degree in statistics from Texas A&M University. His permanent address is

1636 N. Wells #1915

Chicago, IL 60614