INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

RICE UNIVERSITY

**Tests for Harmonic Components in the Spectra of
Categorical Time Series**

by

Monnie McGee

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Katherine B. Ensor, Chair
Associate Professor of Statistics

Dennis D. Cox
Professor of Statistics

Mark Wiesner
Associate Professor of Environmental
Science and Engineering

Houston, Texas

May, 1995

UMI Number: 9610679

UMI Microform 9610679

Copyright 1996, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized
copying under Title 17, United States Code.

UMI

300 North Zeeb Road
Ann Arbor, MI 48103

Tests for Harmonic Components in the Spectra of Categorical Time Series

Monnie McGee

Abstract

The main purpose of spectral analysis in time series is to determine what patterns exist in a particular set of data. To accomplish this, one often calculates the Fourier periodogram of the data and inspects it for peaks. However, since the periodogram is not a consistent estimate of the true spectral density, peaks can be obscured. Therefore, it is necessary to test for significance of the peaks. One of the most widely used tests for significance of periodicities in the periodogram is Fisher's test. The test statistic for Fisher's test is the quotient of the maximum periodogram ordinate and the sum of all the ordinates. If the test statistic is too large, we reject the null hypothesis that the data are white noise.

In this thesis, I develop a test for categorical time series which is an analog of Fisher's test for continuous parameter time series. The test involves finding the Walsh-Fourier periodogram of the data and then calculating the test statistic for Fisher's test. I explain the theory behind Walsh-Fourier analysis and compare the theory to that of Fourier analysis. Asymptotic results for the distribution of Fisher's test for Walsh-Fourier spectra are presented and compared with a simulated distribution. I also perform power studies in order to assess the detection capability of the test. In the presence of multiple peaks, this test tends to lose power. Therefore, I also explore several alternatives to Fisher's test for Walsh-Fourier spectra and apply all of the alternative methods to several real data sets.

Acknowledgments

No dissertation can be completed without the support and suggestions of many people. First of all, I would like to thank my advisor, Katherine B. Ensor, for taking on the challenge of directing a thesis which explores some uncharted territory. Without her insight, questions, suggestions, and encouragement, this project would not have been possible. David S. Stoffer, a man who has done the pioneering work in this field, deserves special mention for his guidance. Also, I thank Dennis D. Cox and Mark Wiesner, the other members of my committee, for their time and input.

Completion of a work such as this also requires the support, suggestions, encouragement, and patience of many good friends, family, and colleagues. An incomplete list of these people follows: Keith Baggerly, Meg Gelder Ehm, Stephen Harper, Maeve McCarthy, Theodora Overfelt, Stephan Sain, and the members and friends of the singles group of West University Church of Christ in Houston, Texas.

A special thanks goes to my parents, Mary and Glynn McGee, and my sister, Melinda McGee. I cannot say enough to express the depth of my gratitude to them.

Contents

Abstract	ii
Acknowledgments	iii
List of Tables	vii
List of Illustrations	ix
1 Introduction	1
1.1 Previous Work	3
1.1.1 Markov Chains	3
1.1.2 DARMA and NDARMA Processes	6
1.2 Tests for Periodicities	10
1.2.1 Schuster's and Walker's Tests	10
1.2.2 Fisher's Test	11
1.2.3 Increasing the Power of Fisher's Test	12
1.2.4 Tests for Walsh-Fourier Analysis	13
1.3 Walsh-Fourier Analysis	15
1.4 Applications of Walsh-Fourier Analysis	18
1.4.1 Ordinal and Nonsinusoidal Data	18
1.4.2 Nonlinear Time Series	19
1.5 The Scope of the Thesis	21
2 Fisher's Test for Walsh-Fourier Spectra	23
2.1 Signal plus Noise Model for Categorical Time Series	23
2.1.1 Asymptotic Distribution of the Maximum I	24
2.1.2 Simulation Results I	26
2.2 Redefining the Walsh-Fourier Periodogram	28
2.2.1 Asymptotic Distribution of the Maximum II	29
2.2.2 Simulation Results II	32
2.3 Discussion	37

3	Performance of the Test Against Alternatives	38
3.1	Power Against Non-sinusoidal Data	38
3.2	Power Against the Clipped AR(p) Model	40
3.2.1	P-values for the Clipped AR(1) Model	41
3.2.2	P-values for Clipped AR(2) Models	44
3.3	Results for Multiple Thresholds	47
3.4	Discussion	49
4	Modifications to Improve the Power of Fisher's Test	50
4.1	The Median	50
4.1.1	Asymptotic Distribution of F_m	50
4.1.2	Simulation Results for F_m	51
4.2	The Trimmed Mean	54
4.2.1	Asymptotic Distribution of F_r	54
4.2.2	Simulation Results for F_r	57
4.3	The Maximum Order Statistic	61
4.4	Performance of the Modified Statistics	62
4.5	Discussion	65
5	Data Analysis	68
5.1	Analysis of the Sleep State Data	68
5.2	Analysis of the Job Load Data	71
5.3	Analysis of the Polarity of the Earth	73
5.4	Analysis of Carbon Dioxide Levels above Mauna Loa	76
5.5	Analysis of AT&T Bond Rates	78
5.6	Discussion	79
6	Future Work	81
6.1	Tests for Multiple Peaks	81
6.2	Tests for Comparing Multiple Spectra	82
A	Computer Code	84
A.1	Walshmax.c : Computes the Walsh-Fourier transform of categorical white noise data	84
A.2	Sort.c: program for sorting the elements of a vector	90

A.3	Walshcat.c: calculates WFP of categorical white noise data via the “signal plus noise” model	91
A.4	Walshclipar1.c: calculates the WFP for clipped AR(1) data	95
A.5	Walshreal.c: calculates the WFP of data read from a file	100
A.6	Walshpulse.c: simulates square-wave data	103
A.7	Walshdarma.c: calculates the logiacal covariance function of a DARMA(1,1) process	105
A.8	Pvalue.S: programs to calculate the power and size of Fisher’s test . .	107
A.9	Darma.S: Splus programs simulating DARMA and DAR models . . .	108
A.10	Other programs: Splus functions to accomplish various small calculations	110

Bibliography	112
---------------------	------------

Tables

2.1	Comparison of Upper Quantiles of the Extreme Value Distribution with the Upper Quantiles of the Simulated Maximum of the WFP from the Signal plus Noise Model, $N = 128$	27
2.2	Comparison of Upper Quantiles of the Extreme Value Distribution with the Upper Quantiles of the Simulated Maximum of the WFP for all Categories and Sample Sizes.	36
2.3	Empirical Size of the Simulated F_w , $N = 512$	36
3.1	Coefficients of γ_1 with corresponding rejection rates for the AR(1) model.	42
3.2	Selected AR(2) coefficients and corresponding rejection rates, $N = 512$	46
4.1	Comparison of Upper Quantiles of the Extreme Value Distribution Scaled by the Median with the Upper Quantiles of the Simulated Maximum of the WFP, $N = 512$	53
4.2	Empirical Size of the Simulated F_m , $N = 512$	53
4.3	Empirical Size of the Simulated F_r for Various Values of r , $N = 512$	60
4.4	Empirical Size of the Simulated F_d , $N = 512$	62
4.5	Coefficients of γ_1 with Corresponding Rejection Rates for F_w , F_m , and F_r , $N = 512$	64
4.6	Coefficients of γ_1 and γ_2 with Corresponding Rejection Rates for F_w , F_m , and F_r , $N = 512$	65
5.1	Test Statistics, P-values, and Empirical Sizes for Modifications of Fisher's test on the Sleep State Data.	70
5.2	Test Statistics, P-values, and Empirical Sizes for the Job Load Data.	73
5.3	Test Statistics, P-values, and Empirical Sizes for the Geomagnetic Reversal Data.	75

5.4	Test Statistics, P-values, and Empirical Sizes for the Mauna Loa Data.	77
5.5	Test Statistics, P-values, and Empirical Sizes for the Bond Rate Data.	79

Illustrations

1.1	A realization of a DARMA(1,1) process of length 200 with $\beta = 0.7$ and $\rho = 0.4$	8
1.2	(a): Realization of a DAR(1) process with $\rho = 0.4$, (b): Autocorrelation function, (c): Spectral Density.	9
1.3	(a): $\tau(j)$ for an NDARMA(1,1) process where $\rho = 0.3, \delta_0 = 0.2, N = 128$. (b): $\rho = 0.9, \delta_0 = 0.2, N = 128$	17
2.1	Quantiles of Extreme-Value Distribution vs. the Maximum of the Walsh-Fourier periodogram for $N = 128$	28
2.2	QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N = 32$ and $C = 2, 3, 4$ and 5	33
2.3	QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N = 128$ and $C = 2, 3, 4$ and 5	34
2.4	QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N=512$ and $C = 2, 3, 4$ and 5	35
3.1	Periodic Nonsinusoidal data and Walsh-Fourier periodograms for 3 Categories with $N = 32$ (top) and 4 Categories with $N = 128$ (bottom).	39
3.2	Strength of correlation of clipped AR(1) process versus the percentage rejected at the 0.05 level for Fisher's test for Walsh-Fourier spectra.	41
3.3	Realization of data (a) and WFP (b) for clipped AR(1) data with $\gamma_1 = 0.35$	43
3.4	Realization of data (a) and WFP (b) for clipped AR(1) data with $\gamma_1 = 0.95$	44
3.5	Perspective Plot of rejection rates for clipped AR(2) models (z-axis) with respect to γ_1 (x-axis) and γ_2 (y-axis), $N = 512$	45

3.6	Realization of data and Walsh–Fourier Periodograms for Clipped AR(2) processes with coefficients .23 and .95, (a) and (b), and .84 and .17, (c) and (d).	47
3.7	Realizations of Data and WFP’s for a Clipped AR(2) Process with Coefficients .23 and .95, (a) and (b), and a Clipped AR(3) Process with Coefficients .75, .15, .35, (c) and (d).	48
4.1	Quantiles of the Extreme Value Distribution for the maximum from the WFP scaled by the median vs. Quantiles of Simulated Distributions, $N = 512$	52
4.2	Quantiles of the Extreme Value Distribution for the maximum from the WFP Scaled by the Trimmed Mean vs. Quantiles of Simulated Distributions, $N = 512$ and $r = 50$	58
4.3	Absolute Error Between 95th Quantile of the Extreme Value Distribution and 95th Quantile of F_r for $r = 5, 15, 50, 100, 200, 256$	59
4.4	QQ Plots for F_d where $N = 512$ and $C = 2, 3, 4$, and 5	61
4.5	Coefficients of γ_1 versus Rejection Rate at the 0.05 Level of Significance for Modified Statistics.	63
4.6	Data (a) and Walsh–Fourier periodogram (b) for a Clipped AR(2) with $\gamma_1 = 0.15$ and $\gamma_2 = 0.5$	66
5.1	(a): Sleep state record of an unexposed infant. (b): Walsh–Fourier periodogram of the same record.	69
5.2	(a): Job Load data, (b): Walsh–Fourier periodogram of the Job Load Data.	72
5.3	Walsh–Fourier Transform of Geomagnetic Reversals over 40 million years.	74
5.4	Walsh–Fourier periodogram of the monthly CO ₂ levels above Mauna Loa, 1979–1990	76
5.5	Walsh–Fourier periodogram of daily bond yields from April 1975 to December 1975	78

Chapter 1

Introduction

Categorical time series are serially correlated data which are gathered in terms of states (or categories) at discrete time points. They appear in many different contexts. For instance, one may monitor the amount of rain for a series of weeks and record only whether or not a particular day was wet or dry, resulting in a binary time series. The pattern of a person's sleep states, where the reading from an EEG is recorded as one of six possible sleep states, is of interest to many neurologists. Another application in the biosciences that involves using the methods of analysis developed for categorical time series is gene characterization. Geoscientists are interested in geomagnetic reversals of the polarity of the earth from "normal" polarity to "reverse" polarity. Marketing researchers might collect weekly data corresponding to the preferences of consumers for different brand name products.

In all of the examples mentioned previously, determining the pattern of the time series (or whether or not a pattern even exists) can provide valuable insight into the problem being studied. Spectral analysis is the tool by which statisticians determine the existence of a pattern in a time series. To understand the objective of spectral analysis, consider the visible spectrum. Most people are familiar with the way white light is decomposed into different colors as it travels through a prism. The different colors of the rainbow correspond to the different wavelengths of light. Some time series are also composed of functions which have different "wavelengths". They can be decomposed into separate parts, analogous to the visible spectrum, in order to interpret the data in a meaningful way. The "prism" which statisticians use to decompose a time series is a transform based on a set of complete orthogonal functions. The most frequently used transform is the discrete Fourier transform, given by

$$a_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{it\omega_j},$$

where $\omega_j = 2\pi j/n \in (-\pi, \pi]$, n is the sample size, and the data are denoted by the x_i 's.

The Fourier periodogram (FP) is defined by

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

[6]. By plotting the values of the periodogram versus the frequencies, we can determine whether or not a periodic pattern exists in the data. If there is a peak in the periodogram at a specified frequency then we may suspect that there is a period of the corresponding length in the original time series signal.

Decomposing continuous time series data into smooth sine and cosine functions usually gives us a reasonably accurate determination of the patterns in the data. However, the categorical time series mentioned in the first paragraph of this chapter do not consist of continuous data. These data sets contain sharp jumps which could be overlooked if the data were correlated with a smooth function. It makes more sense to perform spectral analysis of categorical time series by using a set of orthogonal functions which have sharp jumps. One such set of functions is the Walsh functions, a complete orthogonal set of functions which consist of square waveforms. Because the Walsh functions are not smooth, they are able to follow discrete levels in categorical time series [2].

Instead of revealing a pattern in the data in terms of frequency (cycles per unit time), as does Fourier analysis, Walsh–Fourier analysis gives us a pattern for the data in terms of “zero-crossings per unit time”, or *sequency* [2]. A Walsh–Fourier analysis provides different insight into categorical data than a Fourier analysis; sometimes this approach is more informative. For instance, consider the study of the patterns of sleep states of neonates, where the sleep states were recorded as one of six categorical variables [34]. Both a Fourier analysis and a Walsh–Fourier analysis were performed. The two analyses gave approximately the same results, in the sense that the maximum peak in both periodograms occurred at relatively the same place. This peak was interpreted as the average time for a given infant to cycle through all sleep states. However, the Walsh–Fourier spectrum also showed a peak at another place in its spectrum. This component was interpreted as the average length of time that an infant stays in each sleep state. By looking at the data, the researchers could see that this second peak was meaningful; however, the Fourier analysis missed this second component.

As mentioned earlier, determining whether or not a pattern exists in a given realization of a time series is often as important as defining the pattern. There are numerous tests to ascertain whether or not a pattern exists in time series data. Some

of these tests require the user to test for the existence of a pattern at a specified frequency. Other tests are more general because they test the null hypothesis that the series is white noise versus the alternative that the series is not white noise. One such general test is Fisher's test. The test statistic for Fisher's test is given by

$$F_q = \frac{\max_j [I(\omega_j)]}{\sum_{j=1}^q I(\omega_j)}, \quad (1.1)$$

where $q = \lfloor (n-1)/2 \rfloor$ [4]. For the Fourier periodogram, it is possible to obtain the exact distribution of this statistic.

1.1 Previous Work

In this section, we discuss several methods which have been used to perform spectral analysis of categorical time series. First, we discuss the spectral analysis of Markov chain models as seen in [3]. This type of analysis produces an interpretation of the data in terms of rankings or orderings instead of in terms of sequences. The next method we discuss is based on a discrete formulation of an ARMA model [18]. The spectra produced from this method are Fourier spectra, and have much the same interpretation as the spectra for continuous ARMA models.

1.1.1 Markov Chains

One of the ways in which categorical time series models traditionally have been handled is as time homogeneous Markov chains. Time homogeneous Markov chains have the property that the probability of moving from one state to another is the same for each time interval of length t . In addition, the state to which the present state changes depends only on the present state. No past states are involved [9]. These models are flawed in three important ways [8].

- Categorical time series (or any time series) do not always have a one lag dependence structure. Of course, it is possible to use a Markovian model with more than one lag of dependence.
- With the increase in the lag of dependence, there is a corresponding exponential increase in the number of parameters, unless additional constraints are placed on the model.
- Adding constraints can cause the Markov chain to become nonhomogeneous.

An example of the spectral analysis of Markov chains is given by Beckett and Diaconis [3]. They explain the theory for binary data and then extend the results to polytomous chains. We will sketch the theory for the binary case only.

Consider $X_1, X_2, \dots, X_n$ with $X_i = (X_{i0}, \dots, X_{ip})$ and each X_{ij} taking values in $\{0, 1\}$. The index p indicates the length of the string. For instance, if we were to flip a coin 6 times and record the outcomes as either zeroes or ones, and repeat the experiment twenty times, $p = 6$ and $i = 1, \dots, 20$. The X_i are assumed to be independent realizations from a Markov chain with transition matrix

$$T^i = \begin{pmatrix} \theta_{00}^i & \theta_{01}^i \\ \theta_{10}^i & \theta_{11}^i \end{pmatrix}$$

and arbitrary starting position $\theta_{i0}^i = P(X_{i0} = 1)$. The starting state X_{i0} and the 2×2 count matrix, T^i , are sufficient statistics for the i th string. Then $T_{j,k}^i$, the j, k th entry of the count matrix is given by the number of j to k transitions in the string. For instance, suppose that $X_i = (011001)$. Then, $T_{00} = 1$, $T_{01} = 2$, $T_{10} = 1$, and $T_{11} = 1$. Note that the number of elements in the string is one more than the sum of the count in the matrix T . This is the case for all strings.

In order to employ spectral analysis on binary Markov chains, the authors reparameterize the Markov model. First, they fix a starting state $k \in \{0, 1\}$ and a transition count matrix T . The final state of the string, denoted $l \in \{0, 1\}$, is determined by k and T . The string starts and ends in the same state if and only if $T_{01} + T_{10}$ is even. The set of all strings compatible with k and T can be given by a pair of subsets s and t and their complements. The set s , $s \subset \{1, 2, \dots, T_{00} + T_{01} + l\}$, specifies which transitions go from 0 to 0, and the set t , $t \subset \{1, 2, \dots, T_{10} + T_{11} + l - 1\}$, specifies which transitions go from 1 to 0. The functions which give us the number of compatible strings with parameters s and t belong to a vector space which can be written as the product of symmetric groups.

Before we outline the procedure for spectral analysis on a binary Markov model, we clarify the model with an example. Consider $k = 0$ and

$$T = \begin{pmatrix} 1 & 4 \\ 3 & 3 \end{pmatrix}.$$

The string will have length 12, and $l = 1$. Let $s = \{3\}$ and $t = \{1, 2, 5\}$; therefore, $s^c = \{1, 2, 4\}$ and $t^c = \{3, 4, 6\}$. The string begins at 0 and the first element of s^c is 1, this means that the first transition is from 0 to 1. Now we are in state 1, so we

must examine the elements of t or t^c to see what state is next. The first element of t is 1, so the first transition from state 1 is to state 0. Continuing in this manner, we see that the string is given by (010100111011).

To perform spectral analysis on the binary data, let $X_{p-j,j}$ be the set of binary strings of length p which contain j ones. Let $M^{p-j,j}$ be the set of all functions from $X_{p-j,j}$ into the real numbers. Define $f_j \in M^{p-j,j}$ as the function which gives the number of observed sequences with a specified pattern under the model mentioned previously. We can decompose the vector space $M^{p-j,j}$ into the direct sum of subspaces

$$M^{p-j,j} = S^p \oplus S^{p-1,1} \oplus \dots \oplus S^{p-j,j}.$$

The subspaces into which $M^{p-j,j}$ is decomposed have natural interpretations. S^p is a one-dimensional subspace which consists of the constant functions (such as the mean). $S^{p-1,1}$ has dimension $p-1$. The projection of f_j onto $S^{p-1,1}$ can be interpreted as the number of strings having j ones with a one in position i , $1 \leq i \leq p$. The projection of f_j onto $S^{p-2,2}$ results in the number of strings having j ones in positions k and l , adjusted for strings in which j ones occur in either position k or position l . In general, the projection of f_j onto $S^{p-j,j}$ gives us the number of strings having j ones in positions $i_1, \dots, i_j$, adjusted for the “marginal” frequencies in which the ones do not occur in this manner. Beckett and Diaconis [3] also extend these results to multinomial data. The projections of the functions f_j onto the subspaces can be interpreted in much the same way in the multinomial case.

There are two main differences between spectral analysis under the Markov model and spectral analysis using Walsh–Fourier transforms. Using the group theory decomposition method results in a different interpretation of the data than does Walsh–Fourier analysis. When performing spectral analysis via the Markov chain model, we find patterns in the number of times a certain category occurs at different places in the “strings” of data. With this model, we can also see how many times one category occurs at two, three, or more places within the same string for all strings in the data. Walsh–Fourier analysis allows us to see the existence of a general cycle in the data. Depending on the type of data available, and the information we wish to extract from that data, we may choose which method of spectral analysis to use. The other difference is the complexity level. Although spectral analysis under the Markov chain model does give interesting and meaningful results, it is not easily understood or implemented. Walsh–Fourier analysis is more accessible in terms of both theory and practice.

1.1.2 DARMA and NDARMA Processes

Another model for categorical time series is an extension of the theory for standard time series data. Jacobs and Lewis ([16], [17], [18]) developed a scheme for obtaining a stationary sequence of discrete random variables with a given probability function π and an autocorrelation structure like that of an ARMA(p,N+1) process. The new process is called the discrete autoregressive moving average process, denoted DARMA(p,N+1). Here, p refers to the lag of the AR part of the process, while N refers to the lag of the MA part. Suppose $\{X_n\}$ is a DARMA(p,N+1) process. Then its mathematical formulation is given by the equation

$$X_n = U_n Y_{n-D_n} + (1 - U_n) Z_{n-(N+1)},$$

where the “autoregressive tail” of the process, Z_n , is given by

$$Z_n = V_n Z_{n-A_n} + (1 - V_n) Y_n, \quad n = -N, -N+1, \dots$$

The definitions of the other components of the model are as follows:

- $\{Y_n\}$ is a sequence of independent and identically distributed random variables defined on a discrete state space with $P(Y_n = i) = \pi(i)$.
- π can be any discrete probability function, such as the Poisson.
- $\{D_n\}$ is defined as a sequence of independent and identically distributed random variables taking values $0, 1, 2, \dots, N$, with $P(D_k = n) = \delta_n$, for $n = 0, 1, \dots, N$.
- $\{A_n\}$ is a sequence of independent and identically distributed random variables taking values $1, 2, \dots, p$ with $P(A_k = n) = \alpha_n$, $n = 1, 2, \dots, p$.
- $\{U_n\}$ and $\{V_n\}$ are independent sequences of binary random variables given by

$$\begin{aligned} P(U_n = 1) &= \beta, & 0 \leq \beta \leq 1 \\ P(V_n = 1) &= \rho, & 0 \leq \rho < 1, \end{aligned}$$

[18].

The DARMA process is defined for a mixture of a possibly infinite number of *independent* random variables with a given probability function. Even if random variables with the same probability function π are *dependent*, they still have probability function π . For instance, suppose Y_1 and Y_2 are dependent random variables

each with a Poisson distribution. Define $\{X_n\}$ as a mixture of Y_1 and Y_2 such that $P(X_j = y_i) = .5$, for $i = 1, 2$ and $j = 1, \dots, n$. X_j still has a Poisson distribution, no matter what the dependence structure of Y_1 and Y_2 . Jacobs and Lewis [18] apply this concept to the modeling of discrete time series data and obtain a new process which is a mixture of a finite number of possibly dependent random variables. This process is called the NDARMA(p,N) process. It is defined by

$$X'_n = V_n X'_{n-A_n} + (1 - V_n) Y_{n-D_n},$$

where $\{A_n\}$, $\{V_n\}$, and $\{D_n\}$ are the same as above. The restriction that $\{Y_n\}$ be a sequence of independent and identically distributed random variables is relaxed to include sequences of dependent identically distributed random variables. Note that it is not necessary to define the NDARMA process in terms of an “autoregressive tail”. Since the autoregression is made explicit in the formulation of the NDARMA model, this process is more reminiscent of the familiar ARMA(p,N) process than is the DARMA process.

In order to understand how the DARMA and NDARMA processes work, we provide the formulation of the DARMA(1,1) and NDARMA(1,1) models. Then we give a realization of a DARMA(1,1) process and explain the probabilities associated with that process. If $\{X_n\}$ is a stationary DARMA(1,1) process, then

$$\begin{aligned} X_n &= \begin{cases} Y_n & \text{with probability } \beta \\ Z_{n-1} & \text{with probability } 1 - \beta, \end{cases} \\ \text{where} & \\ Z_n &= \begin{cases} Z_{n-1} & \text{with probability } \rho \\ Y_n & \text{with probability } 1 - \rho. \end{cases} \end{aligned} \quad (1.2)$$

The autocorrelations are given by the simplified formula

$$\text{Corr}(X_n X_{n+k}) = r(k) = \rho^{k-1} (1 - \beta) [\beta(1 - \rho) + (1 - \beta)\rho].$$

A stationary NDARMA(1,1) process is given by

$$X'_n = \begin{cases} X'_{n-1} & \text{with probability } \rho \\ Y_n & \text{with probability } (1 - \rho)\delta_0 \\ Y_{n-1} & \text{with probability } (1 - \rho)(1 - \delta_0), \end{cases} \quad (1.3)$$

with autocorrelations given by

$$\text{Corr}(X_n X_{n+k}) = r'(k) = \rho^{k-1} [\rho + (1 - \rho)^2 \delta_0 (1 - \delta_0)] \quad (1.4)$$

[18]. Both models explain categorical time series as a sequence in which probabilities are associated with moving to another level or staying at the same level. Not only are these models reminiscent of ARMA processes, but also they call to mind Markov chains. However, we can model higher-order dependence in the data parsimoniously with these models, while we cannot do so with Markov chain models.

Figure 1.1 A realization of a DARMA(1,1) process of length 200 with $\beta = 0.7$ and $\rho = 0.4$.

To clarify further, we will give an example of a realization of the DARMA(1,1) process. Suppose that $\beta = 0.7$ and $\rho = 0.4$. Then, from (1.2), we see that we can have $X_n = Y_n$ seventy percent of the time while the other thirty percent of the time, $X_n = Z_{n-1}$. This thirty percent further decomposes so that with probability $0.3 \times 0.4 = 0.12$, $X_n = Z_{n-2}$, and, with probability $0.3 \times 0.6 = 0.18$, $X_n = Y_{n-1}$, for each $n = 0, 1, \dots, N$. A DARMA(1,1) process with these parameters has been plotted in Figure 1.1. We generate this figure for $\{Y_n\}$ a sequence of Poisson random variables with mean $\lambda = 1$ and $\{Z_n\}$ an autoregressive process of lag one. The programs used

to simulate a DARMA(1,1) process and a DAR(1) process are listed in appendix A.9.

Figure 1.2 (a): Realization of a DAR(1) process with $\rho = 0.4$,
(b): Autocorrelation function, (c): Spectral Density.

Jacobs and Lewis [16] briefly mention spectral analysis of the DMA(N) and DAR(1) processes, but not the spectra of DARMA and NDARMA processes. The spectrum for the DMA(N) process depends on the underlying distribution of the D_n 's, while the spectrum for the DAR(1) process does not. As with the MA(N) process, the DMA(N) process has a spectrum which can be interpreted as a polynomial in $\cos(\omega)$, where ω represents frequency. The value of the initial point of the spectrum is related to the amount of long term dependence in the process. In general, the DAR(1) process allows for more long term dependence than the DMA(N) process.

Figure 1.2(a) is a plot of data from a DAR(1) process with $\rho = 0.4$. Its autocorrelation function and spectral density are given in parts (b) and (c), respectively. For graphs (a) the x-axis represents time and the y-axis represents the value of the

process. Graph (b) gives us the value of the autocorrelation function (y-axis) at each lag of the process (x-axis). The x-axis of plot (c) represents the frequencies from 0 to π and the y-axis represents the value of the spectral density at a particular frequency. In Figure 1.2, we can see that both the correlation and the spectral density of the DAR(1) process give us the same information as those of the AR(1) process. For instance, the autocorrelation function is exponentially decaying with its starting point at the value of the lag 1 coefficient, ρ . The spectral density has a small “peak” at the low frequency end of the spectrum, again indicating the presence of low correlation in the data. Therefore, a spectrum of a DAR(1) process gives us the same information as that of an AR(1) process. This information is displayed in terms of cycles per unit time.

1.2 Tests for Periodicities

As mentioned in the introduction, some applications of time series analysis require the knowledge of the location of periodicities within the time series. At first, statisticians attempted to assign significance to every peak in the periodogram. Since the periodogram is not a consistent estimate of the spectral density of the data, this practice was soon abandoned. Statistical tests for periodicities were needed in order to find the significant peaks in the periodogram.

1.2.1 Schuster’s and Walker’s Tests

One of the first tests developed for the significance of peaks in the Fourier periodogram was Schuster’s test [10]. This test was derived from the fact that, for a random sample $u_1, \dots, u_{2n+1}$, from a normal distribution, any linear functions A and B such that

$$\begin{aligned} A &= \sum_{r=1}^{2n+1} (a_r u_r) \\ B &= \sum_{r=1}^{2n+1} (b_r u_r), \end{aligned}$$

where $\{a_r\}$ and $\{b_r\}$ are sequences with zero mean and unit variance, independent of u_r , will also be distributed normally. Moreover, A and B will be independent if $\sum_{r=1}^{2n+1} a_r b_r = 0$. Define $X = A^2 + B^2$. Clearly, X has a χ^2_2 distribution when we scale by the variance of the sequence $\{u_r\}$. We can choose a_r and b_r to be the Fourier coefficients, which when scaled properly, meet the criterion for normality of A and

B. To perform this test, one calculates the periodogram at a frequency at which one suspects a large value exists. The null hypothesis that the ordinate is significant is rejected for large values of that ordinate. This test requires that we select the frequencies of the periodogram which we want to test in advance of calculating the entire periodogram, and that we know the variance of the sequence $\{u_r\}$.

We hardly ever know which ordinate we wish to test before all such ordinates of the periodogram are calculated. It would be better to choose the largest value of the periodogram ordinates by reason of its magnitude when compared to the other ordinates, instead of choosing an arbitrary frequency at which to apply the test. This requires comparing the largest observed value of $I(\omega_j)$ with the value that would have been observed if the data were white noise. Suppose that the chosen level of significance for the test is α . The corresponding value of the periodogram ordinate, x , for which we reject the null hypothesis of white noise, would be given by $\exp(-x/2\sigma^2) = \alpha$ for any particular term, where σ^2 is the variance of the data sequence [10]. But, if we choose x out of the largest of n values of the periodogram, then the probability that all other values are less than x would be given by

$$(1 - e^{-x/2\sigma^2})^n = 1 - \alpha.$$

This criterion is used by Walker to determine the significance of the largest periodogram ordinate [10].

1.2.2 Fisher's Test

Both Schuster's and Walker's criteria require that the population variance of the original data be known. Fisher developed a test in which it is not necessary to know the variance of the population. The test statistic is given by (1.1). It is easy to find the exact distribution of this statistic for the Fourier periodogram under the null hypothesis that the time series is Gaussian white noise. Using the notation of (1.1), it can be shown that

$$Y_i = \frac{\sum_{k=1}^i I(\omega_k)}{\sum_{k=1}^q I(\omega_k)}, \quad i = 1, \dots, q-1 \quad (1.5)$$

are distributed as the order statistics of a sample of $q-1$ independent random variables, where each one is distributed uniformly on the interval $[0, 1]$. Therefore,

$$P(F_q \geq x) = 1 - \sum_{j=0}^q (-1)^j \binom{q}{j} (1-jx)_+^{q-1}$$

where F_q is given by (1.1), $y_+ = \max(y, 0)$, and x is the value of the test statistic [6].

Fisher's test has been shown to be the most powerful test for testing the significance of a harmonic component at a single frequency [30]. However, there is a fundamental problem with this test in that it has poor detection capability in the presence of large values in the periodogram [4]. Unfortunately, the presence of these large ordinates can result in an inflated value for the denominator. This reduces the overall value of F_q , thus increasing the probability of Type II error. We will see evidence of this characteristic of the test in chapter 3.

1.2.3 Increasing the Power of Fisher's Test

There are two major approaches to reducing the probability of type II error in Fisher's test. Both approaches use the fact that the periodogram may contain several "significantly large" components in order to update the value of the test statistic. We can consider the sum of the periodogram elements in the denominator of (1.1) to be an estimate of the variance of the original process. Large values in the periodogram inflate this estimate of variance, thus causing (1.1) to be too small. Bølviken's solution to this problem involves replacing the sum in the denominator of F_q with a more robust estimate [4]. Two possible replacements involve a type of trimmed estimate and a Winsorized estimate. The trimmed estimate uses the smallest $q - a$ ordinates of the periodogram to compute the sample variance. The Winsorized estimate is given by

$$\hat{\sigma}^2 = \sum_{k < q-a} I(\omega_{(k)}) + (a+1)I(\omega_{(q-a)}),$$

where $I(\omega_{(k)})$ indicated the k th largest element of the Fourier periodogram.

Bølviken shows that the way in which the robust estimate of the variance is obtained is immaterial compared to the gains from using a robust estimate. Using either the trimmed sum or the Winsorized sum, one can expect to gain twice as much power over the statistic (1.1) if large components are present in the periodogram. One weakness of this approach is that the choice of a , the amount of "trimming", is quite arbitrary. Bølviken states that "for longer series it is probably wise to trim away three or four times as many amplitudes as the number of suspected cycles". He offers no more structured guidance on the choice of this parameter.

Another extension of Fisher's test is given by Siegel [30]. Even if we modify Fisher's test so that it is more sensitive to possible peaks in the periodogram, the test is still only meant for testing for the existence of one peak. Siegel shows that Fisher's

test is a special case of a larger family of tests which lose power against Fisher's test when testing for a single periodicity, yet gain considerable power when other significant frequencies are present. These tests adaptively consider the presence of multiple frequencies instead of considering only the largest frequency. Siegel begins with considering (1.5) as Fisher's test statistic. He recognizes the problem that Bølviken also recognized, the presence of more than one extremely large periodogram ordinate can severely reduce the power of Fisher's test. The goal is to find a statistic based on all the large order statistics in the periodogram.

Suppose that g_F is the critical value for Fisher's test. We want to choose a "threshold value", $g \leq g_F$. Then, for each value of Y_i that exceeds this threshold, we sum the difference between g and Y_i . Let $\lambda = g/g_F$. The proposed test statistic, which rejects the null hypothesis for large values, is

$$T_\lambda = \sum_{j=1}^n (Y_j - \lambda g_F)_+, \quad (1.6)$$

where $y_+ = \max(y, 0)$. Note that the case where $\lambda = 1$ corresponds to Fisher's test. When one expects the presence of compound periodicities (Siegel's terminology for more than one significant component), one might want to use a smaller value of λ .

Unlike Bølviken, whose choice of a trimming parameter is rather arbitrary, Siegel offers some criteria for the choice of λ and presents some power studies which suggest the "best" λ for his particular examples. Two simulation studies are presented in Siegel's paper, representing two different possible situations: a case when the data contain one significant component, and a case where the data contain two significant components of the same magnitude. In the case where there is only one significant component, the statistic T_λ is shown to lose little power against F_q for $0.4 \leq \lambda \leq 1.0$. For the case where there are two significant components, T_λ actually gains considerable power over F_q for the same range of λ . Therefore, even when one expects a single periodicity, it might be best to test its significance using T_λ with $0.5 \leq \lambda \leq 0.8$ in order to guard against a loss of power in the presence of another periodicity. Other simulations in Siegel's paper show that the power of T_λ is proportional to the strength of the signal from the periodogram.

1.2.4 Tests for Walsh-Fourier Analysis

We have mentioned two of the several tests for the significance of harmonic components in the Fourier domain. However, with the exception of a paper by Goldstein and

Gordon [14], there has not been much work on such tests for Walsh–Fourier spectral analysis. Goldstein and Gordon find the asymptotic distribution of the maximum order statistic from the Walsh–Fourier periodogram, which has immediate applications to Fisher’s test. They approach the problem by considering the Walsh functions as a 2^p factorial design where the errors are independent and identically distributed with mean zero and there are no main effects or interactions.

First, Goldstein and Gordon define the Walsh functions as a series of Kronecker products, denoted by $\otimes$. They define a matrix M_1 as

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Then, the matrix M_p is defined by $M_1 \otimes M_{p-1}$, for $p > 1$. For instance, if $p = 2$, then,

$$M_2 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

If we arrange the rows (columns) of M_2 in ascending order of the number of sign changes, the rows (columns) of M_2 will correspond to $W(0, \lambda)$, $W(1, \lambda)$, $W(2, \lambda)$, and $W(3, \lambda)$, respectively, as described in section 1.3.

Let U_i be the i th column of M_p and $X = X_1, \dots, X_N$ be $N = 2^p$ i. i. d. variates distributed which is finite in the neighborhood of zero. The Walsh–Fourier transform of X is defined as $U^T X$. Goldstein and Gordon state that there exists a constant γ such that

$$\left| P\left\{ \max_{1 \leq j \leq N} |U_j^T X| \leq tN^{1/2} \right\} - e^{2N(1-\Phi(t))} \right| < \gamma(F, \epsilon, \delta) N^{\delta-\epsilon} \quad (1.7)$$

where $0 < \epsilon < 1/2$, $\delta > 0$, and $t > \sqrt{2 \log N - (3 - 2\epsilon) \log \log N}$. The notation $\gamma(F, \epsilon, \delta)$ denotes the dependence of γ on the three items in the parentheses. The maximum of the Walsh–Fourier coefficients (without taking the absolute value) can be approximated by $\exp[-N(1 - \Phi(t))]$ on the same interval by a nearly identical error bound [14].

Goldstein and Gordon display the results of simulations which show that their asymptotic distribution for the maximum ordinate of the Walsh–Fourier periodogram approximates the “true” distribution quite well, even for sample sizes as small as $N = 16$. At such small sample sizes, the approximation is much better in the upper tails than in the lower tails.

1.3 Walsh-Fourier Analysis

Around the same time that Jacobs and Lewis were developing the theory behind the DARMA and NDARMA processes and Siegel and Bølviken were developing extensions to Fisher's test, a new interest in applications of Walsh functions to statistics was beginning. Like Fourier analysis, Walsh analysis involves the orthogonal decomposition of a series, but not into sums of sine and cosine waves. Walsh functions are square waveforms which take on the values $+1$ and -1 [20]. They were first introduced in 1923, but were mostly forgotten because the underlying theory was not as elegant as the theory of Fourier analysis. Then, in the late 1960's, Walsh functions began to be used for engineering applications such as image processing and pattern recognition. The fast Walsh transform is faster to compute than the fast Fourier transform, and the binary nature of Walsh functions was ideally suited to the binary world of computers [20]. Later, when computers became much more efficient, Walsh transforms were again replaced by the more elegant Fourier transforms. Walsh functions have been used recently for the spectral analysis of categorical time series ([20], [21], [24], [32], [34], [35]).

It is important to point out that there are two types of spectral analysis involving Walsh functions and it is not always clear which type of analysis is being discussed in the literature. The theory developed for *Walsh* spectral analysis is for series which are *dyadically* stationary. A dyadically stationary series is one in which the mean is constant and the covariance of two data points which are separated by a dyadic time increment does not depend on time. Before we discuss the nature of dyadic time, it is important to understand dyadic addition, which is defined as

$$m \oplus n = \sum_{j=0}^q |m_j - n_j| 2^j, \quad (1.8)$$

where m_j and n_j correspond to the coefficients of the j th power of two in the binary representations of m and n , and q is the longer of the length of the binary representations of m and n [20]. Examples of dyadically stationary time series are difficult to find in nature: they tend to be man-made. One such series is given by the response of a sequence filter to a white noise input [25].

Dyadic time behaves in a strange way. For example, on the dyadic clock, if it is 5:00 now, then, three hours from now would be 6:00 [34]. Note also that if it is 6:00 now, then in three hours it would be 5:00 again. Therefore, time cycles between two values if we try to step forward in time by equal increments. For our purposes, there is little

practical use for dyadic time, but, because of the binary nature of Walsh functions, the theory developed for dyadically stationary series is much more elegant than the theory developed for real time. For instance, a spectral representation theorem holds for dyadically stationary time series. Unfortunately, such is not the case for *Walsh-Fourier* spectral analysis, which is based on real-time stationarity. Because there are many more applications for Walsh-Fourier analysis, we will concentrate on the theory behind it instead of the theory behind Walsh analysis. Morettin gives a thorough explanation of the differences between the two types of spectral analysis [24].

The Walsh transform, W_j , of a sequence $\{X_n\}$ is given by

$$W_j = \frac{1}{N} \sum_{j=0}^{N-1} X_j W(j, \lambda)$$

where $N = 2^p$ for $p > 0$ is the sample size and $W(j, \lambda)$ is the j th Walsh function defined at sequency λ . The Walsh functions can be ordered by the number of times each function crosses zero. $W(1, \lambda)$ crosses zero one time; likewise, the function $W(4, \lambda)$ crosses zero four times. There are many ways to order the Walsh functions. We will order the functions by ascending number of zero crossings, also known as *sequency* ordering. Sequency ordering is well suited for the purpose of spectral analysis because it corresponds closely with the ordering of the cosine and sine functions in Fourier spectral analysis [2].

Another function of interest is the j th *logical covariance*, given by

$$\tau(j) = \frac{1}{N} \sum_{j=0}^{N-1} \gamma(j \oplus k - k), \quad (1.9)$$

where $\gamma(j) = E[X(N)X(N+j)]$ and the operation " $\oplus$ " is defined by (1.8) [20]. The role of the logical covariance in discrete parameter time series is similar, both in theory and practice, to that of the autocovariance function in continuous parameter time series. For instance, in order for the Walsh-Fourier spectral density function to exist, the $\tau(j)$'s must be absolutely summable. The same relationship exists between the Fourier spectral density and the autocovariance functions. Also, one can represent the logical covariances as the Stieltjes integral

$$\tau(j) = \int_0^1 W(j, \lambda) dG(\lambda) \quad \forall j \geq 0$$

where $G(\lambda)$ is a non-decreasing, right continuous function with $G(0) = 0$, given by

$$G(\lambda) = \lim_{N \rightarrow \infty} \int_0^\lambda \sum_{j=0}^{N-1} \tau(j) W(j, t) dt$$

Figure 1.3 (a): $\tau(j)$ for an NDARMA(1,1) process where $\rho = 0.3, \delta_0 = 0.2, N = 128$. (b): $\rho = 0.9, \delta_0 = 0.2, N = 128$.

[20]. Both of these formulas are analogous to functions for continuous parameter time series involving the autocovariance function. The program which produces logical covariances of a DARMA(1,1) process is given in section A.7.

By looking at a graph of the logical covariances for two NDARMA(1,1) processes in Figure 1.3, we see that the function tends to zero as the number of lags gets large and also provides information about the probabilities associated with the process, much like the autocorrelation function of an ARMA(p,q) process. In the top graph of Figure 1.3, the autocorrelation function is approximately 0.5 at $j = 1$ and is essentially zero after $j = 1$. The quick decay indicates that the process has a high probability of taking on values in the sequence $\{Y_n\}$, which is white noise. Therefore, the process is dominated by the moving average component of the NDARMA model, and consequently, $\rho \ll 0.5$. The bottom picture shows a process with a high proba-

bility of taking on values other than white noise, as indicated by the slow decay. This process is dominated by the autoregressive part of the process which implies $\rho \gg 0.5$.

There is also an estimate of the spectral density analogous to the Fourier periodogram. The *Walsh-Fourier periodogram* is defined as

$$d_N^2(\lambda) = \left[\frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} X_j W(j, \lambda) \right]^2 \quad (1.10)$$

for $0 \leq \lambda \leq 1$. In general, this periodogram is not a consistent estimate of the Walsh spectral density

$$f(\lambda) = \sum_{j=0}^{\infty} \tau(j) W(j, \lambda), \quad (1.11)$$

where $W(j, \lambda)$ is the j th Walsh function in sequence order, since the variance of (1.10) can be shown to be $2f^2(\lambda)$ [20]. However, if λ_N has a finite binary representation and $\lambda_N \oplus \lambda \rightarrow 0$ as $N \rightarrow \infty$, then $E[d_N^2(\lambda_N)] \rightarrow f(\lambda)$. Unlike the ordinates of the Fourier periodogram, the ordinates of the Walsh-Fourier periodogram are not necessarily asymptotically independent. The covariance between two sequences $\lambda_{1,N}$ and $\lambda_{2,N}$ will go to zero with increasing sample size if certain conditions are met, however. First, both sequences must have finite binary representations. Secondly, we require that $|\lambda_{1,N} - \lambda_{2,N}| \geq N^{-1}$ with $\lambda_{i,N} \rightarrow 0$ for $i = 1, 2$ as $N \rightarrow \infty$ [32].

1.4 Applications of Walsh-Fourier Analysis

Walsh-Fourier spectral analysis can be applied to many different problems in many different fields. Literature exists on the application of Walsh-Fourier analysis in engineering, geology, biology, and economics ([2], [11], [22], [23], [26], [31], [35], [36]). Walsh-Fourier analysis works particularly well for data measured at the nominal or ordinal level. It is also possible to use Walsh-Fourier analysis to find thresholds in non-stationary real-valued time series [34]. These are all instances when Fourier analysis can be difficult (or even impossible) to implement and/or interpret.

1.4.1 Ordinal and Nonsinusoidal Data

Walsh functions have been applied to various engineering applications such as signal processing, digital synthesis, and pattern recognition [2]. Spectral analysis using Walsh functions was first suggested in a paper by Polyak and Schreider [27], where they analyzed the spectrum of discontinuous and transient waveforms using Walsh

functions. In fact, it can be empirically verified that Walsh functions give more accurate estimates of the spectral density when the signal is generated by square waveforms, while the Fourier transform results in a more accurate representation of data generated by a sinusoidal process. For applications in which computational efficiency is of importance, such as on-line applications, Walsh functions are often used, even when the signal is sinusoidal [2].

In addition to engineering, Walsh-Fourier analysis has been used in other fields. One such application to geomagnetic reversals in the polarity of the earth is given in [26]. Negi, Tiwari, and Rao maintain that the current techniques used in the geosciences to determine cycles in geological and geophysical time series are not appropriate for analysis of “nonsinusoidal variations in the Earth’s processes, such as geomagnetic reversals and climatic changes”. Modeling such nonsinusoidal series by traditional methods can often result in the presence of excess peaks in the spectrum, which can result in misleading information about the process. Their paper compares the results of Walsh-Fourier and Fourier analyses when applied to binary and nonsinusoidal data. The authors find that Walsh-Fourier analysis results in better resolution of the spectrum and gives sequences which correspond to theoretically predicted cycles for the data.

Another application of Walsh-Fourier spectral analysis to time series data is given in [35]. This paper contains the analysis of the sleep state data which was briefly mentioned in the opening section of this thesis. In this study, the authors examine the EEG readings of twenty-four infants 24 to 36 hours after birth. There are two groups of twelve infants each: one group of infants whose mothers drank moderate amounts of alcoholic beverages during pregnancy and another group of infants whose mothers abstained from alcohol during pregnancy. Using Walsh-Fourier techniques, the researchers determine that infants exposed to alcohol have significantly different sleep-state patterns than the unexposed infants. These differences in the sleep-state cycling between the two groups reflect possible differences in the development of the central nervous system.

1.4.2 Nonlinear Time Series

One of the areas of ongoing research in time series analysis is the investigation of nonlinear threshold models. Walsh-Fourier analysis can also be used to examine a time series for changes in the mean in a much simpler manner than some time domain

approaches [5]). Brillinger begins by assuming that time is continuous and runs over the interval $[0, 1]$. Then, one divides the domain of observation into intervals of the form $I = (k2^{-p'}, (k+1)2^{-p'})$, where $p' > 0$ and $k = 0, \dots, 2^{p'} - 1$. Brillinger calls this type of interval a “dyadic subinterval”. Note that $n = 2^{p'}$ and $n < N$. p' is chosen according to the apparent number of shifts in the mean of the series. One can use several different values for p' and determine the value which best suits the data via the AIC or a similar criterion. Brillinger uses his technique on the series of annual volumes of Nile River discharge. He finds that dividing the data into four levels returns the smallest AIC.

Stoffer and Panchalingam [31] analyze a similar type of data. They simulate their data using a “clipped AR” model, which is given by

$$X_n = \begin{cases} 1, & \text{if } Z_n \geq u; \\ 0, & \text{if } Z_n < u \end{cases} \quad (1.12)$$

where Z_n is an unobservable strictly stationary, continuous valued time process and u is some fixed threshold level. Such a model may be feasible for several different applications. For instance, consider the case where a person has an allergic reaction to penicillin when the amount of the drug in his or her bloodstream exceeds a certain level. The clipped AR model might also apply in the case of a financial analyst who wishes to analyze how often the rate of return on a bond goes below a certain percentage. Stoffer and Panchalingam also consider the Walsh–Fourier spectra of a DARMA(1,1) process, with much the same results as obtained in section 1.3 of this thesis.

For an example with real data, the authors consider the diastolic blood pressure (DBP) of hypertensive patients. The patients participated in a study where they were asked to maintain a special diet for two months. After the first two months were complete, the participants were monitored for high blood pressure twice a day for two more months. They were monitored in the same manner after being on the diet for four months. The spectra changes from a relatively flat spectrum between two and four months to a spectrum which is suggestive of long runs of high readings and normal readings after four months. This behavior indicates a type of Markovian quality to the data: the DBP of a patient in the present is somewhat dependent on his or her DBP in the past.

From the many examples listed above, one can see that Walsh–Fourier spectral analysis is applicable in a variety of situations in a variety of disciplines. It is even useful for applications involving nonlinear time series. In addition, an analysis using

Walsh–Fourier techniques can produce meaningful results which often provide more insight into the particular problem being analyzed than Fourier spectral analysis. In chapter 5, several additional real data sets will be analyzed for significance of peaks in the Walsh–Fourier periodogram using techniques developed in chapters 2 and 4.

1.5 The Scope of the Thesis

Categorical time series appear in many different disciplines. It is necessary to develop methods to analyze such series. However, the spectral analysis of categorical and nonsinusoidal time series is a subject which is not well covered in the literature. Although there do exist papers on the theory and application of Walsh–Fourier functions to such data, not much has been explored in the way of statistical tests for Walsh–Fourier periodograms.

In this thesis, we wish to extend Fisher’s test for Fourier periodograms to Walsh–Fourier spectral analysis. We formulate this extension in chapter 2 and find the distribution of the test statistic. In chapter 3, we discuss the power of Fisher’s test by finding the simulated p-values for the test under the alternative hypothesis. The data which we use for the power study are simulated from several clipped AR(p) processes. We find that, in general, Fisher’s test does have poor power when more than one peak is present in the WFP. However, the test has reasonable power in the presence of large values in the WFP when the correlation structure of the underlying AR(p) model is greater than 0.8.

Since the test has low power in the presence of multiple peaks, we would like to find ways to increase that power. Chapter 4 explores four options for decreasing the probability of Type II error in the test. Each one of these options involves replacing the denominator of the test statistic for Fisher’s test with some type of robust statistic. We choose to analyze the properties of the test with the median and trimmed mean in the denominator of the test statistic. For each of these alternatives, we obtain distributions of the test statistic under the null hypothesis, and we compare the power of the respective tests via simulation. We also examine the issue of the appropriate amount of trimming for increasing the performance of Fisher’s test.

In chapter 5, we use the statistics developed in chapters 2 and 4 to analyze several different data sets. The data are representative of that encountered in the disciplines of medicine, computer science, geology, environmental science, and finance. For each example, we display and interpret the WFP. The peaks which are seen in the Walsh–

Fourier periodogram of each scenario are interpretable in the light of that particular data set. We also test the significance of the maximum peak. It is found that, for all of these data sets, except one, the maximum peak of the Walsh–Fourier periodogram is significant according to all versions of Fisher’s test. Chapter 6 explores some possibilities for future work. A list of the programs used to simulate data and perform certain calculations is given in the appendix.

Chapter 2

Fisher's Test for Walsh–Fourier Spectra

Since there are many time series data sets where Walsh–Fourier analysis is more appropriate than Fourier analysis, a test for significance of peaks in the Walsh–Fourier periodogram is needed. As mentioned in section 1.2, one of the most useful tests for significance of peaks in the Fourier periodogram is Fisher's test. The test statistic for Fisher's test is given in (1.1). For the Walsh–Fourier periodogram, we propose the following test statistic:

$$F_s = \frac{\max_j(d_N^2(\lambda_{j,N}))}{N^{-1} \sum_{j=0}^{N-1} (d_N^2(\lambda_{j,N}))}, \quad (2.1)$$

where $d_N^2(\lambda_{j,N})$ is given by (1.10). The Fourier analog of this test statistic is given in [6]. In this chapter, we obtain the distribution of this test statistic under two different models. The first model is a signal plus noise model given in [32]. The second model employs “dummy vectors” which represent each data point, much like the method of data representation used in [36]. We will show that the asymptotic distribution of (2.1) under the null hypothesis is more accurately approximated by simulation results which employ the second model.

2.1 Signal plus Noise Model for Categorical Time Series

Suppose that there exists a sequence $\{X_n\}$ which can be decomposed in the following manner:

$$X_n = \theta + S_n + \varepsilon_n \quad n = 0, \pm 1, \pm 2, \dots, \quad (2.2)$$

where $E[X_n] = \theta$, S_n is a zero mean, discrete-valued stationary signal which may depend on unknown parameters, and ε_n is a zero-mean, discrete-valued process which is uncorrelated with S_n . It is possible that there is some dependence between S_n and ε_n . We denote the true Walsh–Fourier spectrum of S_n and ε_n by $f_S(\lambda)$ and $f_\varepsilon(\lambda)$, respectively.

Now, consider a set of M sequences, $\{\lambda_{j_1,N}, \dots, \lambda_{j_M,N}\}$, “close to” the sequence λ , where $\lambda_{j_m,N} \oplus \lambda \rightarrow 0$ as $N \rightarrow \infty$, $m = 1, \dots, M$. Also, $|\lambda_{l,N} - \lambda_{k,N}| \geq N^{-1}$ for

$l = k$. We can regard these M sequences as points at which we sample the values of the Walsh–Fourier periodogram. Defining sequences in this manner gives us a way to smooth the periodogram other than windowing. It can be shown that

$$d_N(\lambda_{j_m, N}) = u_m + z_m \quad a. s. \quad (2.3)$$

for all $m = 1, \dots, M$, where $u_m \sim N(0, f_S(\lambda))$ and $z_m \sim N(0, f_\epsilon(\lambda))$ [32]. Then, from Cochran’s theorem and the fact that $E[d_N^2(\lambda_{j, N})] \rightarrow f(\lambda)$ if $\lambda_{j, N} \oplus \lambda \rightarrow 0$ as $N \rightarrow \infty$, we have

$$\begin{aligned} d_N^2(\lambda_{j_m, N}) &= (u_m + z_m)^2 \\ &= [f_S(\lambda) + f_\epsilon(\lambda)] \chi_1^2 \end{aligned} \quad (2.4)$$

almost surely, as $N \rightarrow \infty$ [32].

2.1.1 Asymptotic Distribution of the Maximum I

Under the null hypothesis that the data are generated by a white noise process, we consider the model (2.2) where $\theta = 0$ and there is no signal component. From (2.3) and (2.4), we have that the Walsh–Fourier periodogram ordinates are asymptotically from a χ_1^2 distribution. However, in order to find the distribution of (2.1), we must find the distribution of the maximum order statistic of the Walsh–Fourier periodogram ordinates. Using extreme–value theory, we can accomplish this goal.

Extreme–value distributions are asymptotic distributions for the minimum or maximum order statistic of a sequence of data. Gnedenko [13] showed that all non-asymptotic distributions are in the domain of attraction of one of three types of extreme–value distributions. Gumbel notes that the Gamma distribution, and thus the χ^2 distribution, belong to the *exponential type* [15]. This means that we want to choose constants a_N and b_N such that

$$\lim_{N \rightarrow \infty} P(Z_N < a_N + b_N x) = \exp(-e^{-x}), \quad -\infty < x < +\infty, \quad (2.5)$$

where $Z_N = \max(X_{(1)}, \dots, X_{(N)})$ [12]. a_N is given by

$$a_N = \inf \left\{ x : 1 - F(x) \leq \frac{1}{N} \right\}, \quad (2.6)$$

and b_N is given by

$$b_N = [1 - F(a_N)]^{-1} \int_{a_N}^{\omega(F)} [1 - F(t)] dt \quad (2.7)$$

where $\omega(F) = \sup\{x : F(x) < 1\}$ [12]. Since F is a Gamma distribution, we note that

$$\begin{aligned} 1 - F(x) &= 1 - [\Gamma(\alpha)\beta^\alpha]^{-1} \int_0^x y^{\alpha-1} e^{-y/\beta} dy \\ &= [\Gamma(\alpha)\beta^\alpha]^{-1} \int_x^\infty y^{\alpha-1} e^{-y/\beta} dy, \end{aligned}$$

which is the incomplete Gamma function. For the case of the Walsh–Fourier periodogram, recall that each ordinate is distributed as χ_1^2 .

Now we wish to determine the value of the parameters a_N and b_N . First of all, note that, by definition, the a_N are the quantiles of the χ_1^2 distribution. Therefore, we need only to calculate the sequence b_N . We are aided in this calculation by an asymptotic formula for the incomplete gamma function, which is found in [1]. This asymptotic formula is found by integrating the incomplete gamma function by parts. If we do this assuming a χ_1^2 distribution we obtain

$$2x^{-1/2}e^{-x/2} \left[1 - \frac{1}{x} + \frac{3}{x^2} - \frac{15}{x^3} + \cdots \right]. \quad (2.8)$$

To calculate b_N , we note that $[1 - F(a_N)]^{-1} = N$ and $\omega(F) = \infty$. Therefore,

$$b_N = \frac{N}{\sqrt{2\pi}} \int_{a_N}^\infty \int_x^\infty y^{-1/2} e^{-y/2} dy \, dx.$$

We can calculate the first integrand through its asymptotic expansion, given in (2.8). Then,

$$\begin{aligned} b_N &\approx \frac{2N}{\sqrt{2\pi}} \int_{a_N}^\infty e^{-x/2} x^{-1/2} dx \\ &\approx \frac{4N}{\sqrt{2\pi}} a_N^{-1/2} e^{-N/2} \end{aligned} \quad (2.9)$$

where we employ (2.8) once more to obtain (2.9). Now suppose there is some probability p for which we want to determine the value of x . Then, setting p equal to the right hand side of (2.5), we obtain that $x = -\log(-\log p)$. Therefore, (2.5) becomes

$$P\left(\frac{Z_N - a_N}{b_N} < -\log(-\log p)\right) = p,$$

with a_N equal to the p th quantile of F and b_N given by (2.9).

We now have the result that, as $N \rightarrow \infty$,

$$\frac{\max_j(d_N^2(\lambda_{j,N})) - a_N}{b_N} \xrightarrow{d} \exp(-e^{-x})$$

However, calculating the extreme-value distribution of the maximum order statistic of the Walsh-Fourier periodogram is only part of the solution to finding the extreme-value distribution of (2.1). We have found that the numerator of (2.1) converges in distribution to the extreme-value distribution of the exponential type. Now, we need to find the asymptotic distribution of the denominator. Note that the denominator is simply the arithmetic mean of the ordinates of the Walsh-Fourier periodogram. Therefore, it converges in probability to the mean of its asymptotic distribution. From (2.4), we know that $d_N(\lambda_{j,N})$ converges in distribution to a χ_1^2 random variable. By the weak law of large numbers, $N^{-1} \sum_{j=0}^{N-1} d_N^2(\lambda_{j,N}) \xrightarrow{P} 1$. Hence, by Slutsky's theorem,

$$\frac{\max_j(d_N^2(\lambda_{j,N})) - a_N}{b_N(N^{-1} \sum_{j=0}^{N-1} (d_N^2(\lambda_{j,N})))} \xrightarrow{d} X, \quad (2.10)$$

where $F_X(x) = \exp(-e^{-x})$.

2.1.2 Simulation Results I

In order to test the rate of convergence and to verify the correctness of (2.10), we perform several computer simulations in which we calculate the Walsh-Fourier periodogram under the null hypothesis of categorical white noise. The program which simulates this data is given in section A.3. A categorical white noise series is defined as a sequence $\{w_1, \dots, w_N\}$ in which each w_i takes on the integer values $1, \dots, C$ with a specified probability. We calculate the Walsh-Fourier periodogram of a categorical white noise series and store the maximum value and the mean of the periodogram for each simulation. Results are obtained for $N = 128$ and $C = 2, 3, 4$ and 5 . We then plot the quantiles of the extreme-value distribution given in (2.5) versus the quantiles obtained from the simulated maximum values, scaled by the appropriate a_N and b_N , and divided by a vector of simulated mean values (see Figure 2.1). Each simulation was run ten thousand times for each of the values of C . We also display a table comparing the upper quantiles of the simulated distribution to those of the extreme-value distribution.

The results of the simulation study are given in Table 2.1 and Figure 2.1. In Table 2.1, we display the values of the 90th, 95th, and 99th percentiles of the simulated data for each category. These can be compared with the values for the same quantiles of the extreme-value distribution, which are also found in the table. It is apparent that the extreme-value distribution is not a good approximation to the simulated distribution in the upper quantiles, except when $C = 5$. In fact, this is the only

Number of Categories	Quantile		
	90	95	99
$C = 2$	-7.672	-7.107	-5.909
$C = 3$	-1.697	-1.125	0.316
$C = 4$	0.813	1.463	2.948
$C = 5$	2.737	3.387	4.987
$C = 6$	3.426	4.171	5.903
$C = 7$	4.184	5.005	6.878
<i>Extreme Value</i>	2.531	2.968	4.590

Table 2.1 Comparison of Upper Quantiles of the Extreme Value Distribution with the Upper Quantiles of the Simulated Maximum of the WFP from the Signal plus Noise Model, $N = 128$.

time when the extreme-value distribution and the simulated distribution are close to each other. Studies for categorical white noise data with more than five categories show that the value of the quantiles for the simulated distribution keep increasing proportional to the number of categories. The number of categories obviously has an impact on the correct scaling of the simulated distribution; however, we see that C does not appear in (2.10). This fact suggests that we should look for a way to scale the simulated distribution so that the impact of the number of categories on the simulated distribution is taken into account.

The plots in Figure 2.1 support the results given in Table 2.1. Each plot is a visual comparison of the quantiles of the extreme-value distribution versus the quantiles of the simulated distribution. The straight line is the line $y = x$. Ideally, we would like the plot of the quantiles to lie along the straight line. Obviously, the approximation to the distribution of (2.1) using (2.10) is not very close, except for the case where $C = 5$. In each case, the discrepancy between the theoretical and simulated distributions is a function of the scaling of the extreme-value distribution, and not of the distribution itself. This fact is evident since the distributions coincide well with each other, yet the line the quantiles form does not coincide with the line $y = x$. As in Table 2.1, it appears that the number of categories should be considered when scaling the distribution.

Figure 2.1 Quantiles of Extreme-Value Distribution vs. the Maximum of the Walsh-Fourier periodogram for $N = 128$

2.2 Redefining the Walsh-Fourier Periodogram

In the previous section, we saw that the theoretical distribution for F_w was not well approximated by the experimental results. Since we wish to find an asymptotic approximation to the distribution of F_w for testing purposes, it is obvious that we need to find another approach. In this section, we argue that the reason the simulated distribution does not match the result of (2.10) is not a problem with the extreme-value distribution, but with the way in which the data are structured when implementing the Walsh-Fourier transform. Furthermore, we suggest a new definition of the WFP, based on a method of structuring the data given in [36], and present simulation results using this new version. These results show that the new definition provides a more accurate method of analyzing data via the Walsh-Fourier periodogram.

Recall the Walsh-Fourier periodogram defined by (1.10). When computing this version of the WFP, one multiplies a $1 \times N$ vector of data by the $N \times N$ matrix

of Walsh functions in order to obtain the Walsh transform of the data. Then the result of each multiplication is squared, returning a $1 \times N$ vector which is the Walsh-Fourier periodogram. There are two main problems with this method of coding the data. The first is that the value which one assigns to the occurrence of a particular category is arbitrary, yet different assignments of values produce different Walsh-Fourier periodograms. As an example, consider a sequence of categorical white noise, $\{w_1, w_2, \dots, w_N\}$, where w_i is either 1, 2, or 3 with probability $1/3$ for each $i = 0, \dots, N - 1$. The Walsh-Fourier transform of this data would be completely different from a series where we assign the values 4, 7, and 17 to the categories and allow each value to occur with equal probability. However, the assignment of numbers to each category is more or less arbitrary, even if the results of issuing a particular set of numbers to some data vary greatly from one assignment to another.

The second problem is the problem of interaction effects among the sequences. The data vector could be arranged in such a way so that, when multiplied by the Walsh transform matrix, some of the sequences cancel each other. For instance, a possible sequence of categorical white noise data for $N = 8$ and $C = 3$ is $\{1, 3, 1, 1, 2, 3, 2, 1\}$. When it is multiplied by $W(4, \lambda) = [1 \ 1 \ -1 \ -1 \ 1 \ -1 \ -1 \ 1]^T$, the resultant is 0. There could be information in this sequence at this sequency, yet it is lost due to the way in which the values of the categories were assigned. Because of these interaction effects and the problem of expeditiously assigning values to the categories, the signal plus noise model given in the previous section produces results which do not compare well with the extreme value distribution.

2.2.1 Asymptotic Distribution of the Maximum II

Now we redefine the Walsh-Fourier periodogram and find the distribution of its maximum. In [36], the authors suggest that the data be represented as a matrix whose dimension is the number of categories (C) by the sample size (N). In other words, the data matrix consists of column vectors $\{e_{1k}, \dots, e_{Nk}\}$, each of length C , where each e_{jk} vector, $j = 1, \dots, N$, contains a 1 in the k th row and zeroes elsewhere. Then, a C dimensional time series is defined by $Y_t = e_{tk}$ when category k occurs at time t . This representation solves the problem of assigning values to the categories of the time series for the case of categorical white noise. We make a slight modification to this representation. Instead of each e_{tk} consisting of 0's and 1's, we define e_{tk} as the

vector which contains $(C - 1)/C$ in the k th row and $-1/C$ in all other rows. With this modification, we are assured of a zero mean time series.

Since we have modified the way in which we describe categorical time series data, we also modify the definition of the Walsh–Fourier periodogram. Therefore, the Walsh–Fourier periodogram at sequency $\lambda_{j,N}$ is defined as

$$D_N^2(\lambda_{j,N}) = \left[\frac{C}{N} \mathbf{J} \{ \mathbf{R} \mathbf{W} \}^2 \right]_j \quad (2.11)$$

where $[\cdot]_j$ denotes the j th element of the $1 \times N$ vector that is the WFP. Furthermore $\mathbf{R}$ is the $C \times N$ matrix containing the e_{tk} vectors, $\mathbf{W} = [W(0, \lambda_{j,N}), W(1, \lambda_{j,N}), \dots, W(N - 1, \lambda_{j,N})]$, where $j = 0, \dots, N - 1$, and $\mathbf{J}$ is a $1 \times C$ vector of ones. Also, when squaring the product of the matrices $\mathbf{R}$ and $\mathbf{W}$, we square each element of the matrix.

Now we need to find the distribution of the WFP given by (2.11), under the assumption that the data are categorical white noise. Although it is not necessary, we will further simplify the derivation of the distribution by assuming that each category occurs with equal probability $1/C$, where C is the number of categories. We also assume that C is small with respect to N . There are N trials in this experiment and each trial is represented by a column in the matrix $\mathbf{R}$. Clearly, the columns of $\mathbf{R}$ form N independent multinomial experiments.

When we multiply the matrix $\mathbf{R}$ by $W(0, \lambda_{j,N})$, $j = 0, \dots, N - 1$, we obtain a $C \times 1$ vector whose elements can be interpreted as the number of times a particular category occurs out of the N experiments, shifted by the mean N/C . For the case where a row of $\mathbf{R}$ is multiplied by $W(i, \lambda_{j,N})$, where $i \neq 0$, the elements of the resulting $C \times 1$ vector are difficult to interpret. Consider the following case where $C = 3$ and $N = 4$. Suppose that $\mathbf{R} \mathbf{W}$ is given by

$$\begin{bmatrix} 2/3 & -1/3 & -1/3 & -1/3 \\ -1/3 & -1/3 & 2/3 & -1/3 \\ -1/3 & 2/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} -1/3 & 1 & 1 & 1 \\ -1/3 & -5/3 & -5/3 & 1 \\ 2/3 & 0 & 0 & -2 \end{bmatrix}.$$

When the mean (which is $4/3$ in this case, since $N = 4$ and $C = 3$) is added, we obtain the matrix

$$\mathbf{R} \mathbf{W}' = \begin{bmatrix} 1 & 7/3 & 7/3 & 7/3 \\ 1 & -1/3 & -1/3 & 7/3 \\ 2 & 4/3 & 4/3 & -2/3 \end{bmatrix}.$$

The first column tells us that the first and second categories appear once, and the third category appears twice. This is exactly what we see in the example matrix $\mathbf{R}$.

The other columns have a slightly different interpretation. The elements represent the number of times each category occurs at a high level (*i.e.* multiplied by 1) minus the number of times each category occurs at a low level. Therefore, a positive value in one of these columns would indicate that a particular category is not observed at the lower level of a particular Walsh function at a certain time. Consider the value 1 in the first row and second column of $\mathbf{RW}$ given previously. This value was obtained by multiplying the first row of $\mathbf{R}$ by $W(1, \lambda_{j,N})$. Two of the elements of the first row of $\mathbf{R}$ which indicate that a category is not observed at that time (a value of $-1/3$) are multiplied by -1 . A “ -1 ” in the Walsh function represents the lower of two levels. The product of a Walsh function at the lower level and the value indicating absence of a category at a particular time is positive, which increases the value of the corresponding element of $\mathbf{RW}$. Similarly, a negative value in $\mathbf{RW}$ indicates the presence of a category at a low level. The magnitude and sign of a particular element of $\mathbf{RW}$ would suggest how often a category is (or is not) observed at the lower levels.

In order to complete the calculation of $D_N^2(\lambda)$, we square each entry of the $C \times N$ matrix $\mathbf{RW}$. Returning to the previous example,

$$[\mathbf{RW}]^2 = \begin{bmatrix} -1/3 & 1 & 1 & 1 \\ -1/3 & -5/3 & -5/3 & 1 \\ 2/3 & 0 & 0 & -2 \end{bmatrix}^2 = \begin{bmatrix} 1/9 & 1 & 1 & 1 \\ 1/9 & 25/9 & 25/9 & 1 \\ 4/9 & 0 & 0 & 4 \end{bmatrix}.$$

Then, we sum each column to obtain the $1 \times N$ matrix of values which is the WFP. In the example, $D_N^2(\lambda) = [2/3 \ 34/9 \ 34/9 \ 6]$ This vector contains sums of the C observed values from $\mathbf{RW}$ minus their expected values, quantities squared. Dividing by N/C , the expected value, results in Pearson’s χ^2 statistic for each element of the vector. Thus, each element of the Walsh–Fourier periodogram has a χ_{C-1}^2 distribution. Recall from (2.4) that each ordinate of the Walsh–Fourier periodogram for the signal plus noise model is from a χ_1^2 distribution. Now we see that there is a basis for the apparent dependence of the fit of the simulated distribution to the extreme-value distribution on the number of categories seen in Table 2.1 and Figure 2.1.

Since the distribution of the Walsh–Fourier periodogram ordinates under the new definition is a χ_{C-1}^2 , we can easily determine the extreme-value distribution of

$$F_w = \frac{\max_j(D_N^2(\lambda_{j,N}))}{N^{-1} \sum_{j=0}^{N-1}(D_N^2(\lambda_{j,N}))} \quad (2.12)$$

from the previous work in section (2.1.1). The distribution of the maximum order statistic from any χ^2 distribution is given by (2.5). Now all that remains is to generalize a_N and b_N , given by (2.6) and (2.7), respectively, so that they can be calculated for a χ^2 with $C - 1$ degrees of freedom.

First, note that a_N is now the $1 - (1/N)$ th quantile of the χ^2_{C-1} distribution. To find b_N , we observe that the asymptotic approximation to the integral of the incomplete Gamma function, given in (2.8), can be written as

$$2x^{-(C-1)/2}e^{-x/2} \left[1 + \frac{C-3}{x} + \frac{(C-3)(C-5)}{x^2} + \dots \right].$$

Then, recalling (2.9), we find that b_N is given by

$$\frac{4N}{\Gamma(\frac{C-1}{2})2^{(C-1)/2}}e^{-a_N/2}a_N^{C-3}. \quad (2.13)$$

Therefore, the asymptotic distribution of the maximum order statistic from the Walsh-Fourier transform as defined in (2.12) is $\exp(-e^{-x})$, when the maximum is scaled by a_N and b_N as defined in this section.

Now, we concentrate on the denominator of the statistic given by (2.12). Recall that each ordinate of the WFP defined in this manner is independently and identically distributed as a χ^2_{C-1} random variable. Hence, by the weak law of large numbers, their mean converges in probability to $C - 1$. Employing Slutsky's theorem once more, we obtain

$$\frac{\max_j(D_N^2(\lambda_{j,N})) - a_N}{b_N(N^{-1} \sum_{j=0}^{N-1}(D_N^2(\lambda_{j,N})))} \xrightarrow{d} \frac{X}{C-1}, \quad (2.14)$$

where $F_X(x) = \exp(-e^{-x})$.

2.2.2 Simulation Results II

In order to test the accuracy of the extreme-value distribution given by (2.14), we perform several simulations for varying values of N and C . The computer code for this simulation is given in sections A.1 and A.10.

Figures 2.2, 2.3, and 2.4 are QQ-plots of the extreme-value distribution, given in (2.14), versus the simulated distribution of (2.12) under the null hypothesis. Each simulated maximum value is scaled by the appropriate constants a_N and b_N and divided by the mean of its WFP. All simulations were repeated 5,000 times.

In figure 2.2, we see the QQ-plots for $N = 32$ when $C = 2, 3, 4$ and 5 . Even with such a small sample size, the extreme-value distribution seems to be a good

Figure 2.2 QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N = 32$ and $C = 2, 3, 4$ and 5 .

approximation to the simulated distribution. The approximation is not as good in the upper tails as in the lower ones, but it is not unreasonable, either. Table 2.2 shows the numerical values for each of the 90, 95, and 99 quantiles of the extreme-value distribution, along with the same values for the simulated distribution, for all sample sizes and categories. Once again, these values are not extremely close together, yet they are within reasonable bounds of each other.

In order to obtain a measure of the closeness of the quantiles of the simulated distribution to the quantiles of the extreme-value distribution, we take the absolute error for each case. For the case where $N = 32$, the largest absolute error is 1.41 and it occurs when $C = 2$ at the 99th quantile. In general, the squared errors tend to become larger as the quantiles being estimated become more extreme. The smallest errors are exhibited for $C = 3$. There is an explanation for this behavior. When $C = 2$, the distribution from which we are taking the maximum is a χ_1^2 . Since this

Figure 2.3 QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N = 128$ and $C = 2, 3, 4$ and 5 .

distribution is highly skewed, we might expect poor performance of the simulated distribution. However, for $C = 3$, the distribution from which we take the maximum order statistic is a χ_2^2 . This distribution is also highly skewed, but the χ_2^2 and the exponential distribution are the same. The exponential distribution has the nice property that its maximum order statistic converges more quickly to its extreme-value distribution than any other of the exponential type [15]. Therefore, we would expect the case $C = 3$ to have better convergence properties than the other cases.

Figure 2.3 is a plot of the results of the simulation for all categories when $N = 128$. The QQ-plots in figure 2.3 show that the two distributions are quite close to each other, except for small fluctuations in the tails. The plots for $N = 128$ depict two distributions which are closer to each other than the plots for $N = 32$ depict. We can see that this is the case by looking at table 2.2 and calculating the errors for all categories. For $N = 128$, we find that the errors over all categories and all quantiles

are lower than those when $N = 32$. As was seen in the case where $N = 32$, the largest error occurs when $C = 2$ at the 99th quantile, and the smallest errors occur when $C = 3$.

Figure 2.4 QQplots of the scaled extreme-value distribution vs. the simulated distribution of F_w , $N=512$ and $C = 2, 3, 4$ and 5 .

Finally, we come to the case where $N = 512$, as plotted in figure 2.4. We can see that the theoretical and the simulated distributions are very close to each other at this large sample size, even for the case where $C = 2$. When looking at table 2.2, we see that the largest discrepancy between the distributions is, once again, for $C = 2$ at the 99th quantile. The fit is best for $C = 3$, as well. In general, the absolute errors for all categories at this sample size are much lower than those for the other sample sizes.

When investigating hypothesis tests, one is also interested in the size of the test. We determine the size of Fisher's test as given by (2.12) by counting the number of critical values of the simulated distribution which are greater than the corresponding

Number of Categories	Type of Distribution	Quantile		
		90	95	99
$C = 2$	<i>Ex Value</i>	2.249	2.968	4.590
	$N = 32$	1.402	2.213	3.177
	$N = 128$	1.856	2.395	3.557
	$N = 512$	1.915	2.530	3.846
$C = 3$	$N = 32$	2.081	2.466	3.841
	$N = 128$	2.167	2.853	4.398
	$N = 512$	2.238	2.964	4.673
$C = 4$	$N = 32$	2.455	3.118	4.612
	$N = 128$	2.476	3.239	5.006
	$N = 512$	2.533	3.281	4.999
$C = 5$	$N = 32$	2.818	3.556	5.229
	$N = 128$	2.845	3.654	5.574
	$N = 512$	2.812	3.704	5.709

Table 2.2 Comparison of Upper Quantiles of the Extreme Value Distribution with the Upper Quantiles of the Simulated Maximum of the WFP for all Categories and Sample Sizes.

quantiles of the extreme-value distribution (program A.10). This is done for theoretical sizes of 0.01, 0.05, and 0.10. The results are given in Table 2.3. The size of the test for $C = 3$ is almost exactly the same as the theoretical size. When the number of categories is greater than three, the size is larger than we would like.

Number of Categories	Theoretical Size		
	0.10	0.05	0.01
$C = 2$	0.057	0.030	0.003
$C = 3$	0.099	0.049	0.010
$C = 4$	0.127	0.067	0.015
$C = 5$	0.153	0.088	0.023

Table 2.3 Empirical Size of the Simulated F_w , $N = 512$.

2.3 Discussion

We have found the asymptotic distribution of Fisher's test for two types of models. The first model is the "signal plus noise" model introduced in [32]. We derived its asymptotic distribution, given by (2.10). While testing the fit of this distribution to some simulated data, it became evident that there was some dependence of the extreme-value distribution on the number of categories in the data. However, (2.10) does not account for dependence on the number of categories in the data. The second model, given for Fourier spectra in [31], uses a matrix of dummy variables to calculate the Walsh-Fourier periodogram. Its asymptotic distribution is stated in (2.14). This extreme-value distribution does take into account the number of categories in the data. Since it does, the fit of the asymptotic distribution to simulated data is much improved over that of (2.10). We also estimated the upper quantiles of both the simulated and the extreme-value distribution in order to compare them. In general, the quantiles were closest when $C = 3$. The size of the test was closest to the theoretical size of the test when $C = 3$, also.

Chapter 3

Performance of the Test Against Alternatives

In this chapter, we explore the accuracy of Fisher's test against alternative hypotheses. In other words, we discern how well the test developed in section 2.2.1 is able to determine the existence of significant peaks in the Walsh–Fourier spectra. First, we reintroduce the clipped AR(p) model, defined in (1.12). Then, we present results of various simulation experiments in which we determine the percentage of the time that the test will reject against different alternative hypotheses. The alternatives which we try are clipped AR(p) models with $p = 1$, $p = 2$, and various coefficients. We also present results for data which are specially suited for Walsh–Fourier spectral analysis: a time series of block pulses with various periodicities.

3.1 Power Against Non-sinusoidal Data

We mentioned in chapter 1 that the Walsh–Fourier transform is best utilized when we need to analyze the spectrum of square-wave or nonsinusoidal data. Now, we will see how well Fisher's test for the WFP performs when the input data are composed of square underlying waveforms. The program used to simulate this data is found in section A.6.

To investigate the p-values against Fisher's test for this type of data, we simulated 5,000 realizations of periodic but nonsinusoidal data, each with an easily discerned period. Then, we calculated the WFP of each realization and recorded the maximum. P-values for each maximum value were obtained by evaluating (2.14) with the simulated data from the non-sinusoidal process. Recall that we are testing the null hypothesis that the series is categorical white noise (CWN) against the alternative that the data are not CWN. Next, we evaluated the extreme-value distribution at the standardized values, counting the number of times these critical values were less than a given level of significance. We performed this experiment for data with 2, 3, 4, and 5 categories, and for sample sizes $N = 32, 128$, and 512.

According to the results of the simulation, all cases, except the case where $N = 32$ and $C = 3$, gave a rejection rate 100% at all levels of significance. In the case of the exception mentioned, the percent rejected was 0% for significance levels 0.01 and 0.05, but the rate was 100% when the level was 0.10. We would expect that the rejection rate would be either 100% or 0% for any sample size or number of categories. Each realization, with the same values for N and C , from which the maximum of the WFP is taken is designed to have exactly the same period. Therefore, each realization should have the same maximum value. If one of those values causes the test to reject, then the other ones will, also. Likewise, the test accepts the null hypothesis for a certain value, it will accept for the others.

Figure 3.1 Periodic Nonsinusoidal data and Walsh–Fourier periodograms for 3 Categories with $N = 32$ (top) and 4 Categories with $N = 128$ (bottom).

Figure 3.1 is a display of a realization from two of the nonsinusoidal processes for which we simulated the p-values of Fisher's test. Graph (a) of the figure is a plot of

a realization of the data where there are 3 categories and a sample size of 32. This is the only case in which the test never rejects. Graph (c) contains a realization of data with 4 categories with $N = 128$. In this case, the rejection rate is 100%.

Compare the Walsh–Fourier periodograms of the two data sets. In Figure 3.1(b), we see a WFP with more than one large peak. The largest peak is not even very large compared to the rest of the peaks. It is the occurrence of these other peaks that is causing the poor performance of the test. In the case where $C = 3$, but $N = 128$ (not pictured) spurious peaks still exist in the spectrum, but the largest peak has sufficiently higher magnitude than the others resulting in a higher rejection rate. Therefore, it is not only the presence of other peaks in the spectrum, but also the magnitude of the largest peak compared with the others that affects the power of Fisher’s test. Figure 3.1(d) has three very sharply defined peaks in the spectrum. Two peaks occur near sequency 64 and the other occurs at sequency 128. In spite of the presence of the other peaks in the spectrum, the test still performs well. The reason the test performs well for this data is that the peaks are the same height and there is no noise in the spectrum.

3.2 Power Against the Clipped AR(p) Model

In section 1.4.2, we introduced a “clipped AR” model, which was analyzed in [31]. The equation for the model is given in (1.12). Note that this model is defined as a binary model, but there is no reason why we cannot have clipped AR models with more than one threshold as well. In this section, we consider the rejection rates of models with both one and two thresholds for various orders of the AR processes. Clipped AR(p) models are simulated by the program given in section A.4.

In order to test how often Fisher’s test for Walsh–Fourier coefficients rejects when the data are from a clipped AR model, we first simulate the model. Then, we transform the simulated data so that they have the structure given in section 2.2.1. Next, the Walsh–Fourier periodogram is calculated and its maximum value is stored. This is done 5,000 times with $N = 512$. The values from the simulation are scaled using the values a_N and b_N given in section 2.2.1. The next step is to calculate the p-values from this distribution, and finally, to determine the percentage of those p-values which would signify a rejection of the null hypothesis for probabilities of Type I error given by 0.01, 0.05, and 0.10.

3.2.1 P-values for Clipped AR(1) Models

We first test a clipped AR(1) model, given by (1.12) with $Z_n = \gamma_1 Z_{n-1} + \epsilon_n$, where $\gamma_1 = \{\pm 0.01, \pm 0.15, \pm 0.35, \pm 0.45, \pm 0.55, \pm 0.65, \pm 0.75, \pm 0.85, \pm 0.95\}$ for different trials. In all trials, the threshold is 0. Various values of γ_1 are tried in order to ascertain if the amount of correlation in the AR(1) process affects the performance of the test. We find the WFP of the clipped processes via (2.12) and store the maximum and mean values for each run of the simulation. In all of the following simulations, $N = 512$. Before calculating the rejection rate for a given clipped AR(1) model, the results are standardized using the constants a_N and b_N derived in section 2.2.1.

Figure 3.2 Strength of correlation of clipped AR(1) process versus the percentage rejected at the 0.05 level for Fisher's test for Walsh-Fourier spectra.

Figure 3.2 is a graph of γ_1 versus the percentage rejected for each AR(1) process when the significance level of the test is 0.05. Note that there is a definite increase in the rejection rate as the correlation structure in the AR process becomes stronger,

whether the correlation is positive or negative. Table 3.1 lists the positive coefficients and their corresponding rejection rates. For low correlation in the AR(1) model

γ_1	.01	.15	.35	.45	.55	.65	.75	.85	.95
$N=32$	1.92	2.22	3.90	6.14	10.52	17.02	26.48	43.46	64.92
$N=128$	1.80	2.68	6.34	10.98	20.16	34.90	57.96	82.64	98.42
$N=512$	2.56	3.86	11.52	21.64	39.20	66.22	90.72	99.18	100

Table 3.1 Coefficients of γ_1 with corresponding rejection rates for the AR(1) model.

we are actually calculating the approximate size of the test in this simulation. Low correlation corresponds to the null hypothesis that the series is categorical white noise. The selected level of significance for the test is 0.05. We see that when $\gamma_1 = 0.01$ and 0.15, the test is conservative. For values of γ_1 larger than 0.15, we can no longer say that the clipped AR(1) data are categorical white noise. From that point on, Figure 3.2 is a graph of the power of the test. Obviously, Fisher's test does not reject the null hypothesis often enough, unless the correlation in the data is very strong. We can also see the asymptotic quality of the test since the test performs much better for sample size $N = 512$ than sample size $N = 32$.

In order to explain the relationship between the correlation structure of the data, the percent rejected for the test, and the Walsh-Fourier periodogram, we consider graphs for one realization of clipped AR(1) processes with different coefficients. Figures 3.3 and 3.4 show two such realizations. To have a clearer presentation of the graphs, we present them for $N = 128$ instead of $N = 512$. However, the features of the graphs at $N = 512$ are manifested in the graphs for the smaller sample size. Consider the clipped AR(1) model where $\gamma_1 = .35$, displayed in Figure 3.3. This value for the coefficient indicates some correlation in the AR(1) process, so we might expect to obtain a reasonable rejection rate. However, we see from figure 3.2 that the rejection rate is approximately 12%. Note that the Walsh-Fourier periodogram (Figure 3.3(b)) has many peaks which are all approximately the same height and that there is a large amount of noise in the spectrum. Recall that in section 1.2.3, we mention several papers which conclude that Fisher's test for Fourier spectra loses power when there are multiple peaks in the Fourier periodogram. Loss of power occurs for the same reasons for Fisher's test of the Walsh-Fourier periodogram.

Figure 3.3 Realization of data (a) and WFP (b) for clipped AR(1) data with $\gamma_1 = 0.35$.

If we examine Figure 3.2 we see that the simulated rejection rate is 100% when $\gamma_1 = .95$. Figure 3.4 gives us some insight into the reason for the large percentage of values rejected for this AR(1) coefficient. Again, it has to do with the number of peaks in the WFP. In this periodogram, there is one very large peak at sequency 128. We interpret this peak to mean that the data oscillate very quickly, crossing zero once every time unit. Note that there is noise in the spectrum, but it is very small compared with the peak at the last sequency. Therefore, Fisher's test is able to detect the significance of that peak quite easily.

From the analysis of the clipped AR(1) models, we see that Fisher's test performs best when the correlation is high. The higher the correlation, the further away from white noise the AR(1) data become. From Figures 3.3 and 3.4, we see that low positive or negative correlation in the AR(1) model translates into more randomness in the clipped version. Likewise, high positive correlation in the AR(1) model tends

Figure 3.4 Realization of data (a) and WFP (b) for clipped AR(1) data with $\gamma_1 = 0.95$.

to result in long strings of data values at one level, followed by long strings at the other level, etc. Frequent switching of the data between high and low levels indicates a high negative correlation. One serious problem with Fisher's test is that the correlation for the AR(1) model must be quite high in order to have a test with a reasonable rejection rate. We hope to show that the modifications made to the test statistic in chapter 4 result in a higher rejection rate with less correlation.

3.2.2 P-values for Clipped AR(2) Models

In addition to exploring the percent of values rejected from Fisher's test for the WFP of clipped AR(1) data, we also wish to explore the rejection rate for clipped AR(2) data. We seek to determine whether or not it is more difficult for the test to discern significant peaks in data with a more complicated correlation structure than the clipped AR(1) process. Again, zero is used as the threshold value.

Determining the effect of the correlation structure on the performance of Fisher's test for AR(2) data is more difficult than the problem presented in section 3.2.1. An AR(2) model is given by $Z_n = \gamma_1 Z_{n-1} + \gamma_2 Z_{n-2} + \epsilon_n$, where ϵ_n is Gaussian white noise. Now, we investigate the effect of certain combinations of values for γ_1 and γ_2 in addition to their marginal effects. The pairs of coefficients for which we simulate the AR(2) processes are $(\pm.01, \pm.02)$, $(\pm.15, \pm.5)$, $(\pm.23, \pm.95)$, $(\pm.45, \pm.5)$, $(\pm.5, \pm.01)$,

Figure 3.5 Perspective Plot of rejection rates for clipped AR(2) models (z-axis) with respect to γ_1 (x-axis) and γ_2 (y-axis), $N = 512$.

$(\pm.55, \pm.5)$, $(\pm.77, \pm.91)$, $(\pm.78, \pm.78)$, and $(\pm 0.84, \pm.17)$. Figure 3.5 is a perspective plot of the simulation results using clipped AR(2) models, and Table 3.2 shows specific rejection rates for selected pairs of AR(2) coefficients. All of the rejection rates given are based on a probability of Type I error of 0.05.

One of the most striking features of Figure 3.5 is that it is almost completely symmetric about $\gamma_1 = 0$. This fact indicates that the value of γ_2 influences the percentage of rejection more heavily than does the value of γ_1 . Numerical evidence of

γ_1, γ_2	0.01,0.02	0.15,0.5,	0.23,0.95	0.84,0.17	0.78,0.78
% Rejected	2.7	27.65	99.6	67.9	43.4

Table 3.2 Selected AR(2) coefficients and corresponding rejection rates, $N = 512$.

this is seen in Table 3.2. We can also see that small values for both coefficients in the model, which correspond to a white noise series, result in quite low rejection rates. Therefore, as in the clipped AR(1) case, the test is conservative. We also see that the power of Fisher's test for clipped AR(2) data behave in an analogous manner to the power for clipped AR(1) data.

When γ_1 is small and γ_2 is large, we see rather high power, but this is not necessarily so when both coefficients are large. However, when γ_2 is small and γ_1 is large, we see that the rejection rates are not as large as they are when the reverse is true. Figure 3.5 reveals that the magnitude of the value of γ_2 almost completely determines the percentage of test values rejected for the clipped AR(2) model. Since γ_2 is the coefficient which is more influential in determining the periodicity of the AR(2) process, we expect such a result.

In Figure 3.6 there are plots of the realizations of the data and the Walsh-Fourier periodograms for two clipped AR(2) processes. The process represented in plots (a) and (b) has coefficients $\gamma_1 = .23$ and $\gamma_2 = .95$, while the coefficients for the data and WFP represented in the plots (c) and (d), respectively, are $\gamma_1 = .84$ and $\gamma_2 = .17$. The plots from these two graphs explain what was seen in Figure 3.5. Recall that the percent of values rejected for the case where the AR(2) coefficients are .23 and .95 is almost 100%, even when the significance level is 0.05. We see that there is evidence of a periodic component in the data in Figure 3.6(a). This evidence is not as strong as the evidence seen in Figure 3.4(b) for the clipped AR(1) process, but, there is one main peak which is twice the height of any of the other peaks in the spectrum. This fact contributes to the high percent of rejection for the AR(2) data with coefficients .23 and .95.

Figure 3.6 also gives us insight into the reason that the rejection rate for the case where the coefficients are 0.84 and 0.17 is smaller than we expect. This rejection rate is 67.9%. The WFP of this data set shows the presence of two large peaks at the end of the spectrum, and another peak near the middle of the spectrum. Once again, we

Figure 3.6 Realization of data and Walsh-Fourier Periodograms for Clipped AR(2) processes with coefficients .23 and .95, (a) and (b), and .84 and .17, (c) and (d).

see that the presence of spurious peaks or harmonics in the Walsh-Fourier spectra is detrimental to the performance of Fisher's test.

3.3 Results for Multiple Thresholds

In the introductory section of this chapter, we mentioned that the clipped AR(p) model as defined in [31] is a binary model. However, there is no reason that we cannot have clipped AR(p) models with more than one threshold. In fact, such models could be used to describe many real situations. Suppose we measure the concentration of ambient ozone every hour for a month, for example. As a preliminary analysis, we break up the data into high, medium, and low readings to see if there is a pattern among those categories. We could select thresholds at which to clip the original time series data according to some specified criterion. Then, depending on

the correlation structure of the original data, we would have something like a clipped AR(p) model with two thresholds. One could also analyze high, average, and low cholesterol readings from a group of patients over a given time period in this manner. In this section, we consider the rejection rate for Fisher's test when the data are from a clipped AR(2) or AR(3) model with two thresholds.

Figure 3.7 Realizations of Data and WFP's for a Clipped AR(2) Process with Coefficients .23 and .95, (a) and (b), and a Clipped AR(3) Process with Coefficients .75, .15, .35, (c) and (d).

Figure 3.7 is a display of realizations of the data and the Walsh-Fourier periodograms of two processes. One of these processes is a clipped AR(2) process with thresholds -1 and 1 , and the other is a clipped AR(3) process with the same two thresholds. The clipped AR(2) process has coefficients $\gamma_1 = .23$ and $\gamma_2 = .95$. The realizations shown for this process are shown in Figure 3.7(a) and (b). In (c) and (d) of the figure, we have realizations from a clipped AR(3) process where $\gamma_1 = 0.75$, $\gamma_2 = 0.15$, and $\gamma_3 = 0.35$. We discuss the results for each of these processes in turn.

Recall that a clipped AR(2) process with coefficients .23 and .95, and only one threshold (which is at zero) has a rejection rate of nearly 100% when the level of significance is 0.05. However, when we clip the data at two places, we find that there is a dramatic drop in the percent rejected. At the 0.05 level, the percent rejected for the model in Figure 3.7(a) and (b) is 39.4%. In order to understand why there is such a drop in the rejection rate for this process, we examine the WFP spectra seen in Figure 3.7(b) and Figure 3.6(b). In the first plot, we see that there is a large peak at the zeroth sequency which is not present in the latter plot. Also, the WFP depicted in Figure 3.6(b) is flatter at the sequencies where a peak does not occur than the WFP given by Figure 3.7(b). There are three peaks in the spectrum of Figure 3.7(b). All of these factors mentioned contribute to the lower rejection rate found in the clipped AR(2) process seen in Figure 3.7(a) and (b).

For the clipped AR(3) model with two thresholds, seen in Figure 3.7(c) and (d), the percentage of p-values less than 0.05 is 93.25%. Figure 3.7(d) is the WFP of a realization from this data model. Note that there is a peak in the spectrum at sequency 126, approximately. This peak indicates that the data cross zero roughly once per unit time. If we look at Figure 3.7(c), we see that this seems to be correct. Another interesting feature of this periodogram is that it is fairly flat at all sequencies except the 126th, which contributes to the large rejection rate that we find for this particular model.

3.4 Discussion

In conclusion, we see that Fisher's test performs very well for nonsinusoidal data with a well-defined period. Results for the rejection rate of Fisher's test are somewhat more variable for the clipped AR(p) models. For the clipped AR(1) model with a threshold at zero, we see that models with high correlation are better suited for the test. The magnitude of γ_2 is what drives the power of Fisher's test for the clipped AR(2) models. For AR(2) models with more than one threshold, the test operates well when the first coefficient of the model is large.

The best indicator of whether or not Fisher's test performs well for any model is the number of spurious peaks in the periodogram, in general. The more peaks that appear and the larger the spurious peaks, the worse the test performs. In chapter 4, we introduce modifications to Fisher's test which improve the performance of the test in situations where multiple peaks are present.

Chapter 4

Modifications to Improve the Power of Fisher's Test

We studied the performance of (2.12) in the last chapter and found that there are cases when the probability of a Type II error is higher than it should be. In this chapter, we concentrate on methods to improve the power of Fisher's test for periodic components in the Walsh–Fourier spectra. We explore what happens to the results from chapters 2 and 3 when we replace the denominator of (2.12) by the median of the WFP components or the trimmed mean of the components. We also investigate the properties of the maximum of the WFP as a test statistic. First, we explore the respective distributional properties of (2.12) under all previously mentioned modifications. In section 4.4, we examine the size and power of Fisher's test based on these alleged improvements. A brief discussion of the effect of the amount of trimming on the size and power of Fisher's test is also included.

4.1 The Median

It is well known that the sample mean is not a robust estimate of the center of a data set. If the sample in question has one or two extreme outliers, then the sample mean can be heavily biased. In our case, we are calculating the mean of the WFP, which may or may not contain large values. Using a robust estimate of the center in the denominator of the test statistic might improve the power of the test.

4.1.1 Asymptotic Distribution of F_m

First, we consider using the sample median as an estimate of center for the WFP. In most instances, the median will not be greatly affected by the presence of peaks in the spectrum, while the mean is affected by the presence of peaks. Modifying (2.12) to incorporate the sample median, we obtain

$$F_m = \frac{\max_j(D_N^2(\lambda_{j,N}))}{\text{med}_j(D_N^2(\lambda_{j,N}))}. \quad (4.1)$$

In order to use this statistic to test for the significance of the maximum component from the Walsh–Fourier periodogram, we need to find its asymptotic distribution. This is a simple matter, since all we are doing is changing the denominator. Therefore, our derivation of the extreme-value distribution of (2.12) given in section 2.2.1 is valid for (4.1). We can also employ the same scaling parameters, a_N and b_N , that were used to scale the maximum value in section 2.2.1. Since the sample median is a strongly consistent estimator of the fiftieth percentile of a distribution, by Slutsky’s theorem

$$\frac{\max_j(D_N^2(\lambda_{j,N})) - a_N}{b_N[\text{med}_j(D_N^2(\lambda_{j,N}))]} \xrightarrow{d} \frac{1}{\mu_m} X, \quad (4.2)$$

where $F_X(x) = \exp(-e^{-x})$ and the population median is denoted by μ_m .

It is important that we find an estimate of the population median for testing purposes. We can either simulate this value or find it analytically. Recall from section 2.2.1 that each ordinate of $D_N^2(\lambda_{j,N})$ is asymptotically independent and identically distributed as a χ_{C-1}^2 variate. Therefore, μ_m must be the median of a sequence of χ_{C-1}^2 random variables. We can simulate the median using Splus, or use an approximation such as one of the approximations given in [19]. The one we choose for the ϵ th percentage point of a χ_{C-1}^2 distribution is given by

$$\chi_{C-1,\epsilon}^2 \approx (C-1) \left(U_\epsilon \sqrt{\frac{2}{9(C-1)}} + 1 - \frac{2}{9(C-1)} \right)^3, \quad (4.3)$$

where U_ϵ is the ϵ th quantile of the standard normal distribution [19]. In our case, $\epsilon = 0.5$. Therefore, in (4.2), we can approximate μ_m , the median of a χ_{C-1}^2 distribution, with (4.3).

4.1.2 Simulation Results for F_m

Now we wish to see how closely the upper quantiles of the simulated values of (4.1) approximate the quantiles from the extreme value distribution. Using QQ-plots and estimates of the upper quantiles, we compare the simulated results to the extreme-value distribution. First, we run simulations which calculate the value of the WFP, via (2.11), and its median for each of 5,000 runs. The sample size for all runs is $N = 512$. Then, we standardize the simulated maximums by the proper a_N and b_N , and multiply the simulated maximums by the quotient of the value of (4.3) and the simulated medians. The result of these operations is plotted on the y-axis of each plot. The extreme-value distribution is plotted on the x-axis of each plot. The results are given in Figure 4.1 and Table 4.1.

Figure 4.1 Quantiles of the Extreme Value Distribution for the maximum from the WFP scaled by the median vs. Quantiles of Simulated Distributions, $N = 512$.

Figure 4.1 is a display of the QQ-plots for the extreme-value distribution, given in (4.2), versus the simulated distribution for various numbers of categories. The line drawn on each graph is the line $y = x$. Except for the upper tail, the simulated distributions lie close to the line, which indicates that the two distributions approximate each other well. This is even true for the case when the data has 2 categories. Compare this result with that from Figure 2.4, which gives us the QQ-plots of the extreme-value distribution versus the simulated distribution of F_w . In general, the QQ-plots of F_m seem to be better than those for F_w , especially in the upper quantiles.

From Figure 4.1, we see that the simulated distribution is a good approximation of the extreme-value distribution. In order to discern how well the simulated distribution matches the extreme-value distribution in the upper tails, we estimate the 90th, 95th, and 99th quantiles of both distributions and display the results in Table

Number of Categories	Quantile		
	90	95	99
<i>Ex Value</i>	2.249	2.968	4.590
$C = 2$	1.980	2.707	4.275
$C = 3$	2.354	3.033	4.457
$C = 4$	2.551	3.370	5.185
$C = 5$	2.891	3.728	5.550

Table 4.1 Comparison of Upper Quantiles of the Extreme Value Distribution Scaled by the Median with the Upper Quantiles of the Simulated Maximum of the WFP, $N = 512$

4.1. Once again, we measure the closeness of the quantiles from the two distributions by the absolute difference between simulated and theoretical quantiles. The overall results in terms of error are basically the same as the results from Table 2.2. The smallest absolute errors occur when $C = 3$. The largest errors occur for all quantiles when $C = 2$. Also, the approximation of the quantiles becomes less accurate as the quantiles become more extreme.

Number of Categories	Theoretical Size		
	0.10	0.05	0.01
$C = 2$	0.085	0.041	0.006
$C = 3$	0.109	0.054	0.008
$C = 4$	0.127	0.073	0.018
$C = 5$	0.153	0.095	0.023

Table 4.2 Empirical Size of the Simulated F_m , $N = 512$.

Another important consideration for a hypothesis test is its probability of Type I error. Table 4.2 displays information on the size of F_m . The entries in the table were calculated for sizes of 0.01, 0.05, and 0.10 by counting the number of values of the simulated, properly scaled vectors which exceed the corresponding critical value of the extreme-value distribution. Once again we see evidence that the simulated distribution is closest to the extreme-value distribution when $C = 3$. Also, the test is conservative for $C = 2$ and $C = 3$, but not for the other categories. If we compare

the results with the results in Table 2.3, we find that the size of the test using F_m is much the same as the the size when using F_w .

4.2 The Trimmed Mean

Another robust measure of the center of the data is the trimmed mean. Since it is the large order statistics in the WFP that tend to unnecessarily inflate the mean, we do not use a symmetric trimmed mean. Rather, we trim the largest r order statistics from the WFP and calculate the mean with the remaining $N - r$ ordinates. This new statistic for Fisher's test is given by

$$F_r = \frac{\max_j(D_N^2(\lambda_{j,N}))}{(N - r)^{-1} \sum_{j=0}^{N-r-1} D_N^2(\lambda_{(j,N)})}, \quad (4.4)$$

where (j, N) denotes the j th order statistic of the WFP. In the sequel, the term "trimmed mean" will refer to the asymmetric trimmed mean defined above.

The logic behind using this measure for the denominator of Fisher's test statistic is much the same as the logic for the use of the median. However, the sample median tends to be unstable. This stability may be a factor in the performance. Also, by considering only one ordinate of the WFP in the case of the median, we are losing information from the other ordinates which may be important. Use of the trimmed mean allows us to consider as many ordinates of the periodogram in our analysis as possible, without including the large ones which inflate the denominator.

4.2.1 Asymptotic Distribution of F_r

It is necessary to compute the asymptotic distribution of F_r . We begin by noting that the numerator of (4.4) has the distribution which was derived in section 2.2.1. The denominator is quite a different matter. We want to find the constant, denoted μ_r , to which the denominator converges in probability. μ_r can be simulated by calculating the r th trimmed mean of the χ_{C-1}^2 distribution via Splus. However, we would like to find a theoretical value for μ_r . Using the approach of Chernoff, Gastwirth, and Johns, [7], and auxiliary theory given in Serfling, [28], it can be seen that the trimmed mean is asymptotically normal with mean μ_r and variance σ_r^2 . The variance approaches zero as $N \rightarrow \infty$. These facts imply convergence in probability of the trimmed mean to the desired constant.

The trimmed mean is a linear function of the order statistics, called an *L-estimate*. One form for L-estimates is given by

$$T_N = \frac{1}{N} \sum_{i=1}^N J\left(\frac{i}{N+1}\right) X_{Ni} + \sum_{j=1}^m a_j X_{N,[Np_j]}, \quad (4.5)$$

where X_{Ni} denotes the i th order statistic out of a sample size of N ; $J(u)$, $0 \leq u \leq 1$, is a generating function for weights; $0 < p_1 < \dots < p_m < 1$; and $a_1, \dots, a_m$ are nonzero constants. This formulation is a linear function of the ordered values plus a weight a_j given to particular percentiles if those percentiles are a part of the L-estimate. (For instance, in the case of the interquartile range, $J(u) = 0$, $m = 2$, $a_1 = a_2 = 0.5$, $p_1 = 0.25$, and $p_2 = 0.75$.) The trimmed mean is given by (4.5) with $J(u) = \frac{1}{1-\alpha}$, $0 < u < 1 - \alpha$ and zero otherwise, and $m = 0$. Here and in the sequel, $\alpha = \frac{r}{N}$. The sum in (4.5) goes from $i = 1$ to $N - [N\alpha]$ for the trimmed mean.

Chernoff, Gastwirth, and Johns consider the general formulation

$$T'_N = \frac{1}{N} \sum_{i=1}^N c_{Ni} h(X_{Ni}),$$

where $h(\cdot)$ is some measurable function [28]. For the trimmed mean, $h(X_{Ni}) = X_{Ni}$. Under certain conditions stated below, the special case of T'_N given by (4.5) is $AN(\mu_r, N^{-1}\sigma_r^2)$, where

$$\mu_r = \int_0^1 J(u)H(u)du + \sum_{j=1}^m a_j H(p_j), \quad (4.6)$$

and $H(\cdot) = h \circ F^{-1}$ [28]. The symbol “ $\circ$ ” denotes the convolution of its operands. Since h is the identity function in the case of the trimmed mean, we need to consider only the inverse cumulative distribution function when calculating μ_r . We will not concern ourselves with the value of σ_r^2 other than to note that it does not depend upon N . Therefore, $N^{-1}\sigma_r^2 \rightarrow 0$ as $N \rightarrow \infty$ which implies that $T_N \xrightarrow{P} \mu_r$ [28], p. 53.

Now we state the “certain conditions”, mentioned previously, under which the trimmed mean, given by

$$T_N = \frac{1}{N} \sum_{i=1}^{N-[N\alpha]} \frac{1}{1-\alpha} X_{Ni}, \quad (4.7)$$

is asymptotically normal. The conditions are quoted directly from [7].

Condition 4.1 $H(\cdot)$ is continuous on $(0, 1)$ and satisfies a first order Lipschitz condition in every interval bounded away from 0 and 1. H' exists and is continuous except on a set S of Jordan content 0. (We shall replace H' by 0 wherever H' is not defined.)

Condition 4.2 The Riemann integrals

$$B_1 = \int_0^1 J(u)H'(u)[u(1-u)]^{1/2}du$$

and

$$B_2 = \int_\delta^{1-\delta} J(u)du, \quad 0 < \delta < \frac{1}{2}$$

converge absolutely.

Condition 4.3 There exists a δ_0 , $0 < \delta_0 < 1$, such that

- (a) either $c_{jN} = 0$ for $j \leq N\delta_0$ for all sufficiently large N , or for each $K > 0$ there exists a finite M such that if $0 < u_1, u_2 < \delta_0$ and $K^{-1} < u_1/u_2 < K$ then $M^{-1} < H'(u_1)/H'(u_2) < M$, and
- (b) either $c_{jN} = 0$ for $j \geq N(1 - \delta_0)$ for all sufficiently large N , or for each $K > 0$ there exists a finite M such that if $1 - \delta_0 < u_1, u_2 < 1$ and $K^{-1} < (1 - u_1)/(1 - u_2) < K$ then $M^{-1} < H'(u_1)/H'(u_2) < M$.

Condition 4.4 σ_r^2 is bounded away from zero.

Since $H(u) = F^{-1}(u)$, which is the inverse CDF of a χ_{C-1}^2 distribution, the function $H(u)$ is well-behaved. Clearly, Condition 4.1 is satisfied. The integral B_2 is given by $\frac{1-2\delta-\alpha}{1-\alpha}$, which is absolutely convergent. We see that B_1 is absolutely convergent after performing integration by parts. The second half of part (a) of Condition 4.3 places restrictions on $H'(u)$ so that the tails of that function are not too heavy. In the case of the trimmed mean, $H(u) = F^{-1}(u)$, which, for this setting, is the quantile function of a χ_{C-1}^2 distribution. This is not a heavy-tailed distribution, so part (a) is satisfied. The first half of part (b) is satisfied when $\delta_0 = \frac{r}{N}$. σ_r^2 is a constant not dependent on N , therefore it is bounded away from zero and Condition 4.4 is satisfied.

Let $K_1(t) = \int_0^t J(u)du$. By Lemma A in chapter 8 of [28], $\int_0^1 J(u)F^{-1}(u)du = \int_{-\infty}^{\infty} x dK_1[F(x)]$, when the LHS is finite. Then, (4.6) is given by

$$\mu_r = \int_0^1 J(u)F^{-1}(u)du$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} x dK_1[F(x)] \\
&= \frac{1}{1-\alpha} \int_0^{F^{-1}(1-\alpha)} x f(x) dx.
\end{aligned}$$

Letting $f(x)$ be the density function of a χ_{C-1}^2 distribution and completing the integration, it can be seen that

$$\begin{aligned}
\mu_r &= \frac{C-1}{1-\alpha} \int_0^{G_{C-1}^{-1}(1-\alpha)} \frac{1}{\Gamma(\frac{C+1}{2}) 2^{C+1/2}} x^{C+1/2-1} e^{-x/2} dx \\
&= C-1 \cdot \frac{G_{C+1}[G_{C-1}^{-1}(1-\alpha)]}{1-\alpha},
\end{aligned}$$

where $G_i(\cdot)$ is the CDF of a χ^2 random variable with i degrees of freedom. A similar derivation of this result is also given in [29].

Hence, applying the previous calculations and Slutsky's theorem,

$$\frac{\max_j(D_N^2(\lambda_{j,N})) - a_N}{b_N(N-r)^{-1} \sum_{j=0}^{N-r-1} D_n^2(\lambda_{(j,N)})} \xrightarrow{d} \frac{X(1-\alpha)}{(C-1)G_{C+1}[G_{C-1}^{-1}(1-\alpha)]}, \quad (4.8)$$

where $F_X(x) = \exp(-e^{-x})$.

4.2.2 Simulation Results for F_r

In this section, we explore the relationship between data simulated from F_r and data calculated by (4.8). We begin by examining QQ-plots of the extreme-value distribution versus the simulated distribution for $r = 50$. Then, we compare the upper quantiles of both distributions for different amounts of trimming. The size of the test for various C and r is also displayed and discussed. We conclude that the extreme-value distribution given by (4.8) is a good approximation for the distribution of (4.4).

Figure 4.2 is a series of QQ-plots of the extreme-value distribution of (4.4) versus the simulated distribution of the WFP for various numbers of categories, sample size $N = 512$, and $r = 50$. The distribution of the simulated values in Figure 4.2 matches the extreme-value distribution as well as the simulated values of the statistics F_w and F_m match their respective extreme-value results.

Instead of comparing the values of the upper quantiles from the extreme-value distribution to those of the simulated values via a table, it seems appropriate to display the absolute differences between these two distributions for various values of

Figure 4.2 Quantiles of the Extreme Value Distribution for the maximum from the WFP Scaled by the Trimmed Mean vs. Quantiles of Simulated Distributions, $N = 512$ and $r = 50$.

r . Figure 4.3 is such a display. The values of r investigated were 5, 15, 50, 100, 200, and 256. The absolute differences displayed in this figure are calculated for the 95th percentile. From this graph, it is apparent that the amount of trimming of the test statistic does not have as much effect on how well the simulated distribution approximates the extreme-value distribution as does the number of categories. For each category, the corresponding line on the graph is almost completely horizontal. There does seem to be some effect due to trimming when the amount of trimming is small ($r < 50$). Also, note that the absolute errors for $C = 3$ are much smaller than the absolute errors for the other categories.

Related to the error between the distributions is the size of the test. Table 4.3 presents information on the probability of Type I error for Fisher's test using (4.4). Again, we simulate the size for various values of the trimmed mean. The same pattern

Figure 4.3 Absolute Error Between 95th Quantile of the Extreme Value Distribution and 95th Quantile of F_r for $r = 5, 15, 50, 100, 200, 256$.

that was seen in Tables 2.3 and Table 4.2 occurs here, also. The test is conservative for $C = 2$, exact for $C = 3$, and non-conservative for $C > 3$. The values in Table 4.3 also reinforce what was seen in Figure 4.3. It is not so much the amount of trimming that affects the size of the test as it is the number of categories in the data.

In this section we explored another modification of the test statistic for Fisher's test. This modification, given by (4.4), uses a trimmed mean in the denominator of the test statistic to reduce the influence of large values of the spectra. Figure 4.3 presents little evidence that large amounts of trimming have an adverse effect on the fit of a simulated distribution of F_r to its asymptotic distribution. The same is true of the effect of trimming on the size of the test, as seen in Table 4.3. These results raise the issue of the "optimal" amount of trimming for a given WFP. Currently, the best advice for trimming large values from the WFP is to trim "3 or 4 times" as many ordinates as it is believed that there are cycles in the data [4]. From Table 4.3, we see that too much trimming only results in a slight increase in size. We will explore a more concrete methods of determining an "optimal" amount of trimming in the future. In section 4.4, we compare the performance of the statistic F_r when the data

are not categorical white noise. We use various amounts of trimming for F_r in our comparison in order to determine the effect of trimming on the power of the test.

Number of Categories	Amount of Trimming	Theoretical Size		
		0.10	0.05	0.01
2	5	0.062	0.023	0.004
	15	0.077	0.034	0.005
	50	0.068	0.025	0.004
	100	0.068	0.036	0.005
	200	0.068	0.032	0.004
	256	0.065	0.030	0.004
3	5	0.089	0.039	0.008
	15	0.104	0.050	0.009
	50	0.097	0.046	0.009
	100	0.097	0.047	0.009
	200	0.095	0.046	0.009
	256	0.094	0.045	0.009
4	5	0.122	0.066	0.013
	15	0.130	0.068	0.015
	50	0.123	0.065	0.013
	100	0.119	0.067	0.017
	200	0.119	0.068	0.016
	256	0.118	0.067	0.015
5	5	0.142	0.078	0.020
	15	0.146	0.081	0.021
	50	0.149	0.086	0.022
	100	0.142	0.083	0.022
	200	0.143	0.082	0.023
	256	0.141	0.081	0.022

Table 4.3 Empirical Size of the Simulated F_r for Various Values of r , $N = 512$.

4.3 The Maximum Order Statistic

The Fourier periodogram is a graph of the squared amplitudes of the data. It is necessary to scale the maximum order statistic of the Fourier periodogram since its values increase as the variation in the original time series increases. The same reasoning stands when the signal plus noise model for the Walsh–Fourier periodogram is used. However, the approach derived in section 2.2 of this thesis does not result in a periodogram which is the “squared amplitude” of the data. Therefore, it seems that scaling the maximum in order to obtain a suitable test for significance of the maximum component would not be necessary. The properties of the test statistic

$$F_d = \max_j(D_N(\lambda_{j,N})) \quad (4.9)$$

are investigated in this section. From the previous sections, it is easy to see that

$$\frac{F_d - a_N}{b_N} \xrightarrow{d} X \quad (4.10)$$

where $F_X(x) = \exp(-e^{-x})$, and a_N and b_N are given by (2.6) and (2.13), respectively.

Figure 4.4 QQ Plots for F_d where $N = 512$ and $C = 2, 3, 4$, and 5 .

Figure 4.4 allows us to see how closely the simulated distribution of F_d matches the theoretical distribution given in (4.10) for $N = 512$. Clearly, the two types of

distributions are quite close to each other for all categories. These QQ-plots are similar to all the other QQ-plots in that the distributions are closest to each other when $C = 3$. In addition, the simulated distribution has smaller values than the extreme-value distribution for $C = 2$. The values of the simulated distribution are larger than those of the extreme-value distribution for $C = 4$ and $C = 5$.

Number of Categories	Theoretical Size		
	0.10	0.05	0.01
$C = 2$	0.054	0.026	0.003
$C = 3$	0.096	0.047	0.009
$C = 4$	0.130	0.068	0.015
$C = 5$	0.149	0.082	0.022

Table 4.4 Empirical Size of the Simulated F_d , $N = 512$.

Table 4.4 is a display of the size of Fisher's test when the test statistic F_d is used. Note that the sizes for this test are smaller than those of the other tests, although the test is still not conservative for $C = 4$ and $C = 5$. In terms of probability of a Type I error, F_d performs as well as (if not slightly better than) F_w , F_m , and F_r . Its performance in terms of power will be examined in the next section.

4.4 Performance of the Modified Statistics

The reason for investigating the modifications to Fisher's test statistic is to improve the performance of (2.12) under the alternative hypothesis. We hope that one of the variations we have introduced will have better power than that of F_w . In this section, the performance of F_m , F_r , and F_d are investigated using pulse data, clipped AR(1) data, and clipped AR(2) data.

For the data which are generated by block pulses occurring at regular intervals, the power studies revealed much the same for the cases of F_m , F_r , and F_d as they did for F_w . Pulse data with sample sizes 32, 128, and 512 were generated with categories 2, 3, 4, and 5. The test rejected the null hypothesis that the data were white noise for all combinations of categories and sample sizes.

Figure 4.5 is a plot of the value of γ_1 , the AR(1) coefficient, versus the percent of values rejected from the clipped AR(1) model for F_m , F_r , and F_d . Values of r ranging from $r = 50$ to $r = 256$ were used in order to ascertain the effect of various

Figure 4.5 Coefficients of γ_1 versus Rejection Rate at the 0.05 Level of Significance for Modified Statistics.

amounts of trimming on the power of the test. We have also plotted the rejection rates for (2.12) for comparison. The legend in the bottom right-hand corner of the graph shows the symbol on the graph corresponding to the plotting character for each of the modifications. The legend is labeled according to the denominator of the particular statistic being examined. For instance, the curve for (4.9), which is Fisher's test using the unscaled maximum, is denoted with octagons. Since it was found that the power for the various tests is a symmetric function of the correlation in the clipped AR(1) model, we plot only the positive values.

There is little difference in the power of the test statistics plotted in Figure 4.5. In addition, the power when the correlation in the model is near zero (corresponding to the size of the test) is approximately the same for all modifications of Fisher's test presented in this plot. Of the statistics plotted, F_m yields the highest power for the clipped AR(1) model. F_r , where $r = 256$, has the next highest. The power decreases

Statistic	Coefficients of AR(1) Model								
	.01	.15	.35	.45	.55	.65	.75	.85	.95
F_n	2.95	4.00	11.90	23.35	41.35	65.70	91.20	99.40	100
F_w	2.56	3.86	11.52	21.64	39.20	66.22	90.72	99.18	100
F_{15}	2.76	4.28	12.84	25.38	45.12	71.14	92.96	99.84	100
F_{50}	3.68	4.66	14.06	27.10	47.82	73.66	94.52	99.86	100
F_{100}	2.92	4.44	12.94	24.98	46.94	74.38	95.68	99.82	100
F_{200}	2.90	4.28	13.76	26.14	48.90	76.92	95.82	99.94	100
F_{256}	2.72	4.14	13.44	25.50	48.18	76.54	95.68	99.94	100
F_m	4.50	5.40	15.80	31.20	51.65	78.80	95.90	99.90	100

Table 4.5 Coefficients of γ_1 with Corresponding Rejection Rates for F_w , F_m , and F_r , $N = 512$.

slightly with the amount of trimming in the WFP, with F_w , which is essentially F_r with $r = 0$, having the least power. F_d has marginally better power than F_w for small values of γ_1 ; the two statistics have the same power for larger values of the AR(1) coefficient. We present the numerical values for the power at the various values of γ_1 in Table 4.5.

As in Table 3.1, we can approximate the size of the test by examining the value of the test statistic when $\gamma_1 = 0.01$. In this case, all statistics presented in the table represent conservative tests. This is expected since $C = 2$ for this particular clipped AR(1) model. The test statistics F_w , F_d , and F_r all have sizes which are close together. F_m , although it has the largest size, is also the least conservative test. In addition, F_m has the most power against the alternative hypothesis. The slightly larger size of Fisher's test using F_m may be explained by a "power-size trade-off".

Table 4.6 is a display of the rejection rates resulting from the simulation of the power of the test for clipped AR(2) data. The coefficients for γ_1 and γ_2 listed below are the same as those for which the power is reported in Table 3.2. In this case, F_r has high power for all amounts of trimming. The power of F_r is higher than that of F_w , F_m , and F_d for all values of r and the combinations of γ_1 and γ_2 shown in the table. Although the results are not shown here, this is the case for all combinations of γ_1 and γ_2 displayed in Figure 3.5.

The clipped AR(2) process examined in the first column of Table 4.6 is a process in which the correlation is minimal. We would not expect the power to be large for any test statistic. We see that this is the case. In fact, this process is close to a white

Statistic	Coefficients of AR(2) Model, (γ_1, γ_2)				
	(0.01,0.02)	(0.15,0.5)	(0.23,0.95)	(0.84,0.17)	(0.78,0.78)
F_d	4.20	30.86	99.72	72.48	49.06
F_w	4.10	32.85	99.85	73.35	50.40
F_5	4.32	70.40	100	100	100
F_{50}	4.42	75.78	100	100	100
F_{100}	3.24	73.22	100	100	100
F_{200}	3.28	74.22	100	100	100
F_{256}	3.16	74.00	100	100	100
F_m	5.34	40.42	100	86.04	61.18

Table 4.6 Coefficients of γ_1 and γ_2 with Corresponding Rejection Rates for F_w , F_m , and F_r , $N = 512$.

noise process, which means that the entries in the first column correspond to the size of the test (0.05). All test statistics given in this table have sizes which are close to the true size of the test. Since there is more structure in the clipped AR(2) models, there is an increase in size and power for all tests, also.

For the data given by the coefficients $\gamma_1 = 0.15$ and $\gamma_2 = 0.5$, there is a large increase in power after trimming only five ordinates from the periodogram. From Figure 4.6(b), it is clear that the WFP has approximately five relatively large peaks which could inflate the value of the mean. When these peaks are eliminated, an increase in the power of Fisher's test results. This same phenomenon occurs in the cases where $\gamma_1 = 0.84, 0.78$ and $\gamma_2 = 0.17, 0.78$, respectively. In all of these cases, the low power of F_m , as compared to F_r , is probably due to instability in the value of the median. For $\gamma_1 = 0.23$ and $\gamma_2 = 0.95$, we expect high power because the second coefficient of the model is so large. Therefore, the results in Table 4.6 are not surprising for this particular case.

4.5 Discussion

In this chapter, we have examined modifications to Fisher's test for the significance of the maximum ordinate from the Walsh-Fourier periodogram. The alternatives we have considered are given by (4.1), (4.4), and (4.9). All modifications fit their extreme-value distributions quite well; as well as or better than the distribution of (2.12) approximates its asymptotic distribution. The main reason for considering

Figure 4.6 Data (a) and Walsh-Fourier periodogram (b) for a Clipped AR(2) with $\gamma_1 = 0.15$ and $\gamma_2 = 0.5$.

such modifications was to lessen the influence of large peaks on the denominator (or of the denominator altogether), thereby decreasing the probability of a Type II error. The case where the test statistic was the maximum ordinate of the WFP resulted in size and power comparable to that of F_w . When the denominator was changed, as for F_m and F_r , size and power both increased.

In addition to considering size and power of the test, we also must consider issues of stability and robustness when choosing the denominator for Fisher's test. The median is certainly the most robust of the alternatives we have examined, yet it is also the most unstable. The median also has another disadvantage: by using the median in the denominator of Fisher's test statistic, we lose information from the smaller peaks. Although there is a point where that information is spurious, some of it may be quite valuable. The trimmed mean is fairly robust and stable. With the trimmed mean, there is also a side issue of the amount of trimming which is

optimal for a particular data set. The optimal amount of trimming should eliminate large peaks which are inflating the variance while keeping relevant information in the periodogram. The mean is certainly the most stable statistic of the three alternatives, and it uses all of information in the periodogram. Nevertheless, it is the least robust.

The main goal of this chapter was to develop a test with better power. Robustness and stability, while they may affect the power, are only secondary considerations to it. From Figure 4.5, we see that (4.1) has the most power against the clipped AR(1) alternative for every value of γ_1 . However, (4.1) also has the largest size of any of the test statistics under the clipped AR(1) alternative. These sizes are still less than (or slightly greater than) the size of the test. As we increase the value of r in (4.4), the power of the test becomes larger, which causes the size to grow accordingly. The increase in power corresponding to larger values of r is more dramatic when the underlying data are from a clipped AR(2) process. It should be noted that the increase in size is small even for large amounts of trimming. In the next chapter, we investigate the power and size of F_w , F_m , F_r , and F_d using real data.

Chapter 5

Data Analysis

In this chapter, we use the theory developed in the previous chapters to analyze various data sets. The first of these data sets is the sleep state data found in [35]. The next data set is the number of jobs running on the computer in the Rice University Statistics Department over a period of almost 2 days. The purpose in this analysis is to determine whether or not there is a particular time of the day when the load on the machine is very high. We also analyze a series of geomagnetic reversals in the polarity of the earth over the last 40 million years, again looking for patterns in the data [26]. The analysis of data pertaining to the level of carbon dioxide over Mauna Loa is also discussed. Finally, we explore the patterns of the true rate of return of AT&T bonds.

All of these series are analyzed with Walsh–Fourier techniques. First, we give some background on each data set, including a graph of the data and their Walsh–Fourier periodogram. Then, a brief interpretation of the periodogram is given. We also look at the results of Fisher’s test, as defined in (2.12), for each series. In addition, we consider the test results for each data set when the statistics F_d , F_m and F_r , for three values of r , are used. All tests are performed at the 0.05 level. For each data set, the three values of r for which the statistic F_r is calculated represent roughly 10%, 20%, and 50% trimming, respectively, from the data.

5.1 Analysis of the Sleep State Data

In this section we wish to apply Fisher’s test for periodicities in the Walsh–Fourier spectrum to a real data set involving the sleep states of newborn infants. The data are taken from a study in which the sleep states of twenty-four neonates were recorded for 119–122 minutes approximately 24 to 36 hours after birth. Twelve of the infants had mothers who abstained from the use of alcohol during pregnancy, while the other twelve had mothers who used moderate amounts of alcohol during pregnancy. The neurologists who conducted this study were interested in the cycle of sleep states for

each group of neonates. They also wanted to know whether or not the cycles of sleep states were different for the exposed and unexposed groups. For more information on the study, see [35].

Figure 5.1 (a): Sleep state record of an unexposed infant. (b): Walsh-Fourier periodogram of the same record.

For our purposes, we use data for the EEG readings of one of the infants and compute the Walsh-Fourier transform. The data were recorded in terms of six categories corresponding to six sleep states. The sleep states are labeled

- state 1 quiet sleep - trace alternate,
- state 2 quiet sleep - high voltage,
- state 3 indeterminate sleep,
- state 4 active sleep - low voltage,

- state 5 active sleep - mixed, and
- state 6 awake.

The sleep-state record shown in Figure 5.1(a) has 122 data points. In order to calculate its Walsh-Fourier periodogram, according to (2.12), we pad the series with its mean until we obtain 128 data points. Then, we take the Walsh transform of a 6×128 matrix (denoted by $\mathbf{R}$) of dummy variables whose values correspond to a particular sleep state. If the infant is in sleep state j at time t , then the t^{th} column vector of the matrix $\mathbf{R}$ has the value $5/6$ in the j^{th} row and the value $-1/6$ in the other rows.

Figure 5.1(a) shows the sleep state record of one of the unexposed infants. Note that there seems to be a cycle of approximately 45 minutes in the data. The existence of this cycle is confirmed by the Walsh-Fourier periodogram, shown in Figure 5.1(b). This plot, and all subsequent plots of Walsh-Fourier periodograms in this chapter, is actually a plot of sequency against $10 \log_{10}(x + 1)$, where x represents the values of the WFP. One of the distinguishing features of Figure 5.1(b) is the large peak at the third sequency (starting the counting from zero). This peak indicates that the data cross zero every $128/3 = 42.67$ minutes, which translates to a “period” in the data of roughly 45 minutes. There are two secondary peaks in the spectrum at the zeroth sequency and the twelfth sequency. The presence of these peaks and the noise in the tail of the spectrum may result in a slight loss of power in the test.

	F_d	F_w	F_m	F_{15}	F_{30}	F_{64}
<i>Test Statistic</i>	38.61	7.72	20.59	20.55	27.98	46.56
<i>P-value</i>	0	0.0004	1×10^{-9}	1×10^{-9}	0	0
<i>Empirical Size</i>	0.127	0.103	0.112	0.108	0.108	0.105

Table 5.1 Test Statistics, P-values, and Empirical Sizes for Modifications of Fisher’s test on the Sleep State Data.

Now we wish to test the significance of the peak seen at the third sequency. We will do this for all four versions of the Fisher’s test and compare the results. Table 5.1 gives the values of the different test statistics calculated for the sleep state data. Each test statistic is scaled by the appropriate a_N and b_N as given in section 2.2.1. We also calculate the p-values based on the theoretical asymptotic distribution of the test statistic and the empirical sizes of all tests. Clearly, the test rejects for all tests.

From Table 5.1, it is clear that the peak at the third sequency in the Walsh–Fourier periodogram is significant. The p -values for all reported tests are much smaller than the level of significance which we have chosen for the test. Notice that, as more large values are trimmed from the periodogram, the test statistic becomes larger and the p -value becomes smaller, as was seen in section 4.2. However, the size of the test becomes larger, also. In fact, all of the sizes are at least twice as large as the theoretical size of the test (0.05). Since it is obvious that the null hypothesis should be rejected for the sleep state data, we would choose the test with the smallest size. For this data, that test is F_w .

5.2 Analysis of the Job Load Data

The next data set we analyze represents the number of jobs running simultaneously on the server for the statistics department at Rice University. “Stat0”, as the server is called, is used by both professors and students in the department. Jobs which run on Stat0 range from computationally intensive high-dimensional density estimation routines, to simulations, to simple regression. Each of these jobs occupies a certain amount of memory on the computer. We think it would be an interesting exercise to discover any existing pattern in the number of jobs running simultaneously on this server.

Every ten seconds, Stat0 updates a file which contains the number of jobs running during that time. Another file, which is read from the first file, contains the average number of jobs for each five minute increment. We have extracted $N = 512$ data points from that file, which correspond to approximately forty-two hours of data. Each five minute average is then rounded to the nearest integer in order to obtain an estimate of the total number of jobs within each five minute period. Figure 5.2(a) is a graph of the job load data. Within the data themselves, there does not seem to be any fixed pattern, except for the occurrence of spikes at intervals of 100. However, a “periodic” pattern is not immediately apparent.

The Walsh–Fourier periodogram for the job load data is given in Figure 5.2(b). Here, we see a huge spike at the zeroth sequency and very low levels of noise at the other sequencies. A spike at the zeroth sequency obviously cannot be interpreted in the same manner as peaks at the other sequencies can be. For the other sequencies, we usually divide the sample size by the value of the sequency at which the peak occurs to obtain the average number of zero-crossings per unit time. In this case,

Figure 5.2 (a): Job Load data, (b):
Walsh-Fourier periodogram of the Job Load Data.

the usual method would result in an undefined average number of crossings. Instead, we have to consider the spike at the zeroth sequency in terms of the definition of the Walsh-Fourier periodogram itself.

The zeroth sequency is obtained when the first row of the data matrix $\mathbf{R}$ is multiplied by $W(0, \lambda)$. This result is then squared and added to the square of the result of the multiplication of the second row of $\mathbf{R}$ with $W(0, \lambda)$, and so on, up to the C th row. The value at the zeroth sequency tells us the number of times that a particular category occurred over all sequencies. We do not know which category, however. Therefore, we can say that a large value would correspond to one category occurring more often than any other category. If we examine Figure 5.2(a), we can see that this is the case for the job load data. The first category, corresponding to only one job per unit time, occurs the most often in the data. Therefore, if this peak is statis-

tically significant, we can say that Stat0 was occupied with only one job during this particular 42 hour time period.

	F_d	F_w	F_m	F_{50}	F_{100}	F_{256}
<i>Test Statistic</i>	657.56	219.18	2805	2125	2853	5676
<i>P-value</i>	0	0	0	0	0	0
<i>Empirical Size</i>	0.069	0.067	0.073	0.062	0.068	0.067

Table 5.2 Test Statistics, P-values, and Empirical Sizes for the Job Load Data.

In order to make judgements about the significance of the occurrence of the first category, we perform Fisher’s test for Walsh–Fourier spectra. The results of this test are given in Table 5.2. The test statistics for this data are all extremely large. This is because the maximum value of the WFP is 1,251; its mean is 3; its median is 0.234; and its trimmed mean for $r = 1$ is 0.556. Clearly, this large peak has a considerable affect on the value of the test statistics. It seems reasonable to remove this peak when calculating the test statistic because of its influence. When the peak is removed, the maximum recalculated, and the test is performed again, the result is that the null hypothesis is accepted for all test statistics in Table 5.2.

5.3 Analysis of the Polarity of the Earth

It has been said that “Everything on the earth operates in cycles” [26]. Geologists and geophysicists spend a great deal of time and energy analyzing the patterns within the earth’s processes. One of the patterns which they analyze is that of geomagnetic reversals in the polarity of the earth. Over a period of many millions of years, the magnetic center of the earth shifts from the north to the south and back again. Geoscientists call the north pole position “normal” polarity and the south pole position “reversed” polarity. With these shifts in polarity come corresponding changes in weather patterns. It is important to know how often the earth’s polarity switches positions so that we can predict the next shift and prepare for the other changes which are effects of that shift.

There are a variety of ways to study patterns in the earth’s processes. These methods include spectroscopy, autocorrelation techniques, and Fourier transform methods.

While these techniques are quite appropriate to ascertain effects from sinusoidal variations in the data, they are not recommended for use in the analysis of nonsinusoidal variations in the earth's processes. Frequently, the use of techniques appropriate for sinusoidal processes for the analysis of nonsinusoidal ones results in loss of power and the observation of “corrupted, unstable peaks” [26]. Therefore, it is necessary to use a method of pattern analysis which produces accurate, reliable results for nonsinusoidal data. The Walsh–Fourier technique of spectral analysis is one such method.

Figure 5.3 Walsh–Fourier Transform of Geomagnetic Reversals over 40 million years.

Figure 5.3(a) is a graph of the geomagnetic reversals over 7.7 million years. These data are taken from a larger data set which covers 40 million years [26]. The larger data set is divided into approximately equal shorter segments of length $N = 256$ in order to combat nonstationarity in the data. We assume that the shorter segments, which each cover a time period of 7.7 million years, are stationary. In this data set, the value 1 represents normal polarity and the value -1 represents reversed polarity. The

data seem to follow two patterns. One of these patterns is a period of approximately 3 million years (or 100 units of time). The other, which is found within the first, is a time of 0.6 million years (20 units) at normal polarity, followed by a time of 0.06 million years at reverse polarity, followed by another period of 0.6 million years at normal polarity. This second pattern occurs three times during the time period covered by the data.

The Walsh-Fourier periodogram for the geomagnetic reversal data is given in Figure 5.3(b). The maximum peak occurs at the eighth sequency. A secondary peak occurs at sequency fifteen. The primary peak corresponds to thirty-two zero crossings per unit time (or one zero crossing per 0.96 million years). If we look at the plot of the data, we can see that it is approximately thirty-two time units from the first wide valley to the next wide valley. The secondary peak corresponds to seventeen crossings per unit time (or one crossing every half-million years). From the plot, we ascertain that it is close to seventeen units of time from the first wide valley to the first sharp valley, and from that sharp valley to the next wide valley.

	F_d	F_w	F_m	F_{25}	F_{50}	F_{128}
<i>Test Statistic</i>	12.57	12.57	50.27	43.98	69.04	209.25
<i>P-value</i>	3×10^{-6}	3×10^{-6}	0	0	0	0
<i>Empirical Size</i>	0.033	0.034	0.033	0.032	0.031	0.026

Table 5.3 Test Statistics, P-values, and Empirical Sizes for the Geomagnetic Reversal Data.

Now we consider the results of Fisher's test for the data. Table 5.3 lists the p-values of the theoretical asymptotic distribution and values of the test statistic for the various forms of Fisher's test. Again, the test rejects for all versions of the test statistic. Both the p-values and the sizes for each test are small. Also, the sizes are all approximately equal. It is not clear which test, if any, is the "best" test for this data, yet it is clear that there is a significant peak of 0.96 million years in the data.

Note that, in this case, the values of F_d and F_w are the same (to two decimal places). However, with the sleep-state data and the job load data, these values were quite different. This can be explained by the number of categories present in the data. For $C = 2$, these two statistics should be very similar. This was seen in the clipped AR(p) data of the previous chapter, also. However, the more categories there

are in a data set, the more dissimilar these statistics will become. For larger numbers of categories, F_d will result in a larger test statistic, which will produce a test with more power.

5.4 Analysis of Carbon Dioxide Levels above Mauna Loa

Another area where Walsh–Fourier spectral analysis can be applied is in the area of environmental science. Suppose we wish to measure the amount of carbon dioxide in the air above Mauna Loa, an active volcano on the Big Island of Hawaii. Scientists are interested in knowing the levels of CO_2 at high elevations in order to measure the presence of the “Greenhouse Effect”.

Figure 5.4 Walsh–Fourier periodogram of the monthly CO_2 levels above Mauna Loa, 1979–1990

These data are taken from the Splus data set entitled “co2”. The data were gathered every month from January 1958 until December 1990. We choose to concentrate

on the years 1979–1990. This gives us 132 data points, of which we analyze a series of $N = 128$ data points. The original data are continuous with a linear and seasonal trend. We performed first differences in order to remove the linear trend. Next we differenced the data at lag twelve in order to remove the seasonal trend. The data were categorized into three levels, corresponding to high, medium, and low levels of carbon dioxide in the air, according to the first and third quartiles of the data. If a measurement is below the first quartile, we assign it a value of 1. If a measurement is between the quartiles, it is assigned a value of 2. A value of 3 is given to those measurements in the upper quartile of the data. One may complain that such a categorization reduces the amount of information contained in the data. This is true; however, we shall see that the categorization leads to the same results as have been found with Fourier analysis.

Figure 5.4 contains the graph of the data (a) and of the Walsh–Fourier periodogram (b) of this data. Looking at the data, we see that the series seems to be switching back and forth with no discernible pattern. The Walsh–Fourier periodogram, shown in Figure 5.4(b), contains many peaks. The largest peak is at the zeroth sequency. As mentioned in section 5.2, this should indicate that one category occurs more often than any other category in the data set. From (a), we can see that there does not seem to be one category occurring more frequently than the other two.

	F_d	F_w	F_m	F_{15}	F_{30}	F_{64}
<i>Test Statistic</i>	2.42	1.21	1.66	1.65	2.08	3.58
<i>P-value</i>	0.085	0.257	0.174	0.174	0.117	0.027
<i>Empirical Size</i>	0.053	0.040	0.051	0.049	0.051	0.045

Table 5.4 Test Statistics, P-values, and Empirical Sizes for the Mauna Loa Data.

Table 5.4 contains the results of Fisher’s test for the carbon dioxide data. We see that the peak at the zeroth sequency is significant at the 0.05 level for F_{64} and at the 0.10 level for F_d . All of the empirical sizes seem to match the theoretical size of 0.05 quite well. Since the peaks in the WFP of this data have mostly the same height, F_w would probably be the best choice for this data set. The test statistics F_d and F_{64} probably allow too much trimming in this case, which explains their small p-values.

5.5 Analysis of AT&T Bond Rates

The final series we analyze is that of the daily bond yields of an AT&T bond whose coupon rate is listed as 8.25%. This data set is taken from an Splus data set which reports the daily yields of 6 such bonds and their coupon rates for the period from April 1975 to December 1975. We choose $N = 128$ days from the data set with which to perform our analysis. The data are divided into two categories, according to whether or not the yield at maturity of the bond was higher or lower rate than its coupon rate. If the yield was lower, that data point was assigned a 1; otherwise, it was assigned a 2. In this case, we wish to know if the amount which the bond returns follows any pattern over time.

Figure 5.5 Walsh-Fourier periodogram of daily bond yields from April 1975 to December 1975

Figure 5.5(a) is a plot of the data themselves. We see three distinct segments of the data. The bond returns a higher rate than its coupon rate for approximately fifty

days, then oscillates between the two levels for another forty days, then returns a higher rate for the remaining days. The three segments are almost equal in length.

As for the WFP, seen in Figure 5.5(b), we see a peak of the zeroth sequency and at the second sequency. Recall from section 5.2 that a peak at the zeroth sequency means that one value in the data set occurs more than any other. We can see that the value 2 occurs much more often than the other value in this data set. The peak at the second sequency corresponds to a zero-crossing every 64 time units. This period is longer than we expected for this data set. A set of “twin peaks” occurs at sequencies 4 and 6, also. All of these secondary peaks could be a result of the noise in the data, however. Since Fisher’s test is specifically formulated to test the significance of the maximum order statistic, testing secondary peaks would require a modification of the test. We suggest some approaches to such tests in the next chapter.

	F_d	F_w	F_m	F_{15}	F_{30}	F_{64}
<i>Test Statistic</i>	14.97	14.97	31.82	15.77	130.42	294.84
<i>P-value</i>	3×10^{-7}	3×10^{-7}	0	1×10^{-7}	0	0
<i>Empirical Size</i>	0.034	0.021	0.036	0.028	0.024	0.018

Table 5.5 Test Statistics, P-values, and Empirical Sizes for the Bond Rate Data.

From the values given in Table 5.5, it is obvious that Fisher’s test is significant at the 0.05 level for all modifications for the bond yield data. Therefore, we can be reasonably confident that this particular bond tends to yield higher than its coupon rate (for this time period). All of the tests are conservative for this data. This is expected, since simulation studies in previous chapters have shown that all the tests presented in this thesis are conservative for $C = 2$.

5.6 Discussion

In this chapter, we have analyzed five different data sets which arise out of various disciplines. The fact that Walsh–Fourier analysis can be applied in so many different areas attests to its utility. In addition to its usefulness, we found that spectral analysis using the Walsh transform produces interesting and interpretable results.

In these five cases, Fisher’s test performed the same, in terms of rejecting the null hypothesis, no matter what modification of it was used. All of the tests showed

that the maximum peak in each periodogram was statistically significant at the 0.05 level, except in the case of the CO₂ data. However, there were differences in the performances of the test with respect to size when the number of categories was large. This fact suggests that trimming (or not trimming) the denominator of Fisher's test statistic becomes a more important issue as the number of categories increases. Also, the size of the test, for all test statistics presented here, becomes larger proportional to the number of categories. F_d tends to have the most dramatic increase in size, although for as few as six categories, the size for all tests can become unacceptably large.

Another issue which arises from the analysis of these five data sets is that of the peak at the zeroth sequency. Although the peak did not occur in all of the Walsh-Fourier periodograms, it had a significant effect on the outcome of the test in the cases of the job load data. In the case of the Mauna Loa data, the peak was not large enough to have any effect, and in the case of the bond rate data, the absence of this peak would not have affected the outcome. These observations lead to the question of whether or not the peak at the zeroth sequency should be included when performing Fisher's test for the WFP. This peak is easy to interpret, yet it represents information which can be gleaned from the data quite easily.

Chapter 6

Future Work

The current direction of future work for this project falls into two categories. One of the topics of interest is testing for the significance of peaks in Walsh–Fourier spectra other than the maximum peak. The other topic is the comparison of multiple spectra via tests for the significance of peaks in Walsh–Fourier cross spectra. These two areas are briefly addressed in this chapter.

6.1 Tests for Multiple Peaks

The test statistic (2.12) gives us a means of determining the significance of the maximum peak of the periodogram. Suppose another peak is present and we wish to know whether or not it is significant. Instead of taking the maximum periodogram ordinate and dividing its value by either the mean, median, or trimmed mean of all the ordinates, we could use the value of the second order statistic in the numerator of (2.12). Then, we would need to find the distribution of that statistic under the null hypothesis. This asymptotic distribution would be given by the extreme-value distribution of the second largest order statistic from a χ^2_{C-1} distribution.

However, as seen in chapter 3, it is important to take the presence of other peak in the spectrum into account. By testing the second largest order statistic, we are assuming that a larger one exists. If the second one is significant, then the first one should be, also. Therefore, it would make more sense to develop a test statistic based upon the values of both the largest and second largest order statistics. Then, we would need to find the asymptotic distribution of the sum of the largest and second largest order statistics from a χ^2_{C-1} distribution in order to perform the test.

Another possibility would be to modify Siegel’s approach for testing for compound frequencies in the Fourier periodogram [30]. This test is described in section 1.2.3. Siegel’s test considers the hypothesis $H_0 : R_1 = R_2 = \cdots = R_n = 0$ against the hypothesis $H_1 : R_j > 0, R_k > 0, \text{ all other } R_i = 0$, where R_j represents the squared amplitude of the Fourier periodogram at frequency j . This particular alternative hy-

pothesis test for the presence of two significant frequencies. It can easily be extended to cover more. The test statistic, denoted T_λ , is given in (1.6). Each order statistic which corresponds to a peak to be tested in the periodogram contributes to the value of T_λ . The distribution under the null hypothesis for (1.6) is a non-central χ_0^2 distribution. A similar test for Walsh–Fourier ordinates could be developed.

6.2 Tests for Comparing Multiple Spectra

Another topic of interest is the comparison of two or more periodograms through their cross spectra. Stoffer performs two tests to compare the values between spectra in [34]. One of these test is an ANOVA test for “equality” of two Walsh–Fourier spectra. The second test is a Walsh–Fourier analysis of coherency. He applies both types of tests to the sleep–state data discussed in section 5.1. The ANOVA test compares WFP between the two groups of infants, while the coherency is a sequency dependent measure of the correlation between two time series within the same group.

Two ideas for testing the differences in two (or more) Walsh–Fourier spectra come to mind. The first is based on the measure of coherency. The sample coherency between two Walsh–Fourier periodograms $d_A(\lambda_{j,N})$ and $d_B(\lambda_{j,N})$ is given by

$$\hat{K}_{AB}(\lambda) = \frac{d_{AB}(\lambda_{j,N})}{[d_{AA}(\lambda_{j,N})d_{BB}(\lambda_{j,N})]^{1/2}}$$

where $d_{AA}(\lambda_{j,N})$ and $d_{BB}(\lambda_{j,N})$ are obtained by smoothing the Walsh periodograms $d_A^2(\lambda_{j,N})$ and $d_B^2(\lambda_{j,N})$ in a neighborhood of λ . The cross-spectral term is obtained by smoothing the product $d_A(\lambda_{j,N})d_B(\lambda_{j,N})$ in a neighborhood of the sequency of interest [34]. Stoffer simply plots the coherency between sleep–state and total number of body movements for the unexposed infant versus the coherency between the same two spectra for the exposed infants. He does not offer a statistical test to discern the significance of any differences in the coherencies. Such a test might be useful to develop.

The second direction in testing for differences in two or more Walsh–Fourier spectra is testing for significance of peaks in the cross spectra. Walsh–Fourier phase and amplitude spectra can also be defined; however, due to the variance of Walsh–Fourier spectra under cyclical shifts, there are no natural interpretations for such spectra as in the Fourier case [33]. The cross spectra, on the other hand, can be interpreted in much the same way as the cross spectra for the Fourier case. An estimate for

the cross spectra of two Walsh–Fourier periodograms was given in the previous paragraph. Using much the same techniques as were seen in chapter 2, we can find the asymptotic distribution of Fisher’s test for cross spectra.

Appendix A

Computer Code

A.1 Walshmax.c : Computes the Walsh-Fourier transform of categorical white noise data

```
#include <math.h>
#include <stdio.h>
#include <string.h>
#include <malloc.h>

/***** Program to Obtain Walsh-Fourier Spectra *****/

unsigned seed;

/***** Input Variables *****/
/* N - the sample size */
/* C - number of categories in the data */
/* trim - number of large order statistics to exclude */
/*      (can be zero for untrimmed mean) */
/* seed - input for the random number generator */
/* ntimes - number of simulations to run */
/* filename - name of file to which results are written */
/*****/

main(argc,argv) /* Input order: N, C, trim, seed, ntimes, filename */
{
    int argc;
    char *argv[];
    void hadamard(),spectra(),urand();
    int number,*newhmat,N,C,i,j,m,ntimes,len,trim;
    double *flam,*ucat,*maxorder,maxofxvec,*meanvec,mean;
```

```

char *filename;
FILE *fp;

N = atoi(argv[1]);
C = atoi(argv[2]);
trim = atoi(argv[3]);
srandom((seed = atoi(argv[4])));
ntimes=atoi(argv[5]);

len = strlen(argv[6]);
strncpy(filename,argv[6],len);

/***** Allocating memory to vectors and matrices *****/

ucat = (double *)calloc((unsigned) C*N,sizeof(double));
newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
flam = (double *)calloc((unsigned) N,sizeof(double));
maxorder = (double *)calloc((unsigned) ntimes,sizeof(double));
meanvec = (double *)calloc((unsigned) ntimes,sizeof(double));
filename = (char *) calloc((unsigned) len,sizeof(double));

maxofxvec = 0.0;
mean = 0.0;

hadamard(N,newhmat);

for (i=0; i < ntimes;i++)
{
    spectra(N,C,newhmat,flam,ucat,&maxofxvec,&mean);
    maxorder[i] = maxofxvec;
    meanvec[i] = mean;
}

/***** Writing the results to a file *****/
fp = fopen(filename,"w");
for(i=0;i <=(N-1);i++)
    fprintf(fp,"%3.7lf\n",flam[i]);

```

```

/* for (i = 0;i <= (ntimes-1);i++)
    fprintf(fp,"%3.7lf %3.7lf\n", maxorder[i],meanvec[i]);*/
fprintf(fp,"\n\n");
fclose(fp);
}

/*****
/* urand - simulates categorical white noise data in
/*      matrix form
*****/

void urand(N,C,ucat)
    int N,C;
    double *ucat;
{
    long int ia;
    int i,j;
    double out,left,right,DIV;

    for (i=0;i<=(N-1);i++)
        for (j=0; j<=(C-1);j++)
            ucat[i+(j*N)] = (-1.0/((double)C));

    DIV = pow(2.0,31.0)-1;
    for (i = 0;i <= (N-1);i++)
    {
        out = random()/DIV;

        j = 0;
        left = 0.0;
        right = (1.0/((double)C));
        while (!(out > left && out <= right))
        {
            left += 1.0/((double)C);
            right += 1.0/((double)C);
            j+=1;
        }
    }
}

```

```

        ucat[i+(j*N)] = (((double)C)- 1.0)/((double)C);
    }
}

/*****
/* hadamard - computes the NxN hadamard matrix to be
/*      used in calculating the Walsh-Fourier transform */
*****/

void hadamard(N,newhmat)
    int N,*newhmat;
{
    int j,u,v,uv,i,cnt,*ncount,*hmat;

    /***** Initialization of Variables *****/
    ncount = (int *)calloc((unsigned) N,sizeof(int));
    hmat = (int *)calloc((unsigned) N*N,sizeof(int));

    /***** Creating Hadamard Matrix (hmat)*****/
    for (u = 0;u < N; u++)
        for (v = 0;v < N;v++)
            {
                cnt = 0;
                for (uv=u&v; uv ; uv >>= 1 ) {
                    cnt += uv & 1;
                }
                *(hmat+(N*u)+v) = ((cnt & 1) ? -1 : 1);
            }

    /***** Rearranging Hadamard matrix in sequency order *****/
    /* ncount - vector of number of zero crossings in each row of hmat*/

    for (i = 0;i < (N-1);i++)
        for (j = 0;j < N;j++)
            if (*(hmat+i+(j*N)) != (*(hmat+(i+1)+(j*N)))) ncount[j] += 1;

    /*** newhmat - hadamard matrix in sequency order *****/

```

```

for (i = 0; i < N; i++)
    for (j = 0; j < N; j++)
        newhmat[N * ncount[i] + j] = hmat[N * i + j];

free((char *)hmat);
free((char *)ncount);
}

/*****
/* spectra - calculates the Walsh-Fourier periodogram (flam) */
/* of the categorical white noise data (ucat) */
*****/

void spectra(N, C, trim, newhmat, flam, ucat, maxofxvec, mean)
    int N, C, *newhmat, trim;
    double *flam, *mean, *maxofxvec, *ucat;
{
    void urand();
    int i, j, k;
    double summ, *flres;

    flres = (double *)calloc((unsigned) (C * N), sizeof(double));

    urand(N, C, ucat, seed);

    /***** Calculating the Walsh spectrum *****/
    for (i = 0; i <= (C - 1); i++)
    {
        for (j = 0; j <= (N - 1); j++)
        {
            for (k = 0; k <= (N - 1); k++)
                flres[j + (i * N)] += ucat[k + (i * N)] * ((double) newhmat[j + (k * N)]);
        }
    }
    for (i = 0; i <= (C - 1); i++)
    {
        for (j = 0; j <= (N - 1); j++)

```

```

flam[j] += (flres[j+(i*N)] * flres[j+(i*N)]);
    }
    for (i=0;i<=(N-1);i++)
        flam[i] /= ((double)N)/((double)C);

/***** Finding the maximum and mean of flam *****/

*maxofxvec = flam[0];
for (j=1;j <= (N-1);j++)
{
    if (flam[j] > *maxofxvec) *maxofxvec=flam[j];
    else *maxofxvec=*maxofxvec;
}

/*****
    This part of the program can be changed to find the median
    or trimmed mean. Since it is the only part of the program
    which changes when calculating modifications of Fisher's test,
    we simply include the code for those parts which change,
    and for the sorting program used.
    *****/

/***** Mean - for calculating F_w *****/
summ = 0.0;
*mean = 0.0;
for(j=0;j <= (N-1);j++)
    summ += flam[j];
*mean = summ/((double)N);

/***** Median - for calculating F_m *****/
*mean = 0.0;
sort(N,flam-1);
NN = N/2;
*mean = (flam[NN]+flam[NN-1])/2;

/***** Trimmed mean - for calculating F_r *****/
summ = 0.0;

```

```

*mean = 0.0;
sort(N,flam-1);
for (i=0; i <= (N-trim-1);i++)
    summ += flam[i];
*mean = summ/((double)N-(double)trim);

free((char *)flres);
}

```

A.2 Sort.c: program for sorting the elements of a vector

```

/*****
/* sort - sorts a vector in ascending order */
/* n - length of vector to sort          */
/* ra - vector to sort                    */
*****/
void sort(n,ra)
    int n;
    double ra[];
{
    int l,j,ir,i;
    double rra;

    l=(n >> 1)+1;
    ir=n;
    for (;;)
    {
        if (l > 1)
rra=ra[--l];
        else
{
            rra=ra[ir];
            ra[ir]=ra[l];
            if (--ir == 1)
            {
                ra[l]=rra;

```

```

        return;
    }
}

    i=1;
    j=1 << 1;
    while (j <= ir)
{
    if (j < ir && ra[j] < ra[j+1]) ++j;
    if (rra < ra[j])
    {
        ra[i]=ra[j];
        j += (i=j);
    }
    else j=ir+1;
}
    ra[i]=rra;
}
}

```

A.3 Walshcat.c: calculates WFP of categorical white noise data via the “signal plus noise” model

```

#include <math.h>
#include <stdio.h>
#include <string.h>
#include <malloc.h>

/***** Input Variables *****/
/* N - sample size of the WFP */
/* seed - input for the random number generator */
/* ntimes - number of simulations to run */
/* filename - name of file to which results are written */
/*****/

main(argc,argv) /*N, C, seed, ntimes, filename */
    int argc;
    char *argv[];

```



```

{
    void hadamard(),spectra();
    int number,*newhmat,N,C,i,j,m,ntimes,len;
    double *flam,urand(),*noise,*maxorder,maxofxvec,*meanvec,mean;
    char *filename;
    FILE *fp;

    N = atoi(argv[1]);
    C = atoi(argv[2]);
    srandom((seed = atoi(argv[3])));
    ntimes=atoi(argv[4]);

    maxofxvec = 0.0;
    mean = 0.0;

    len = strlen(argv[5]);
    filename = (char *) calloc((unsigned) len,sizeof(double));
    strncpy(filename,argv[5],len);

    newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
    flam = (double *)calloc((unsigned) N,sizeof(double));
    noise = (double *)calloc((unsigned) N,sizeof(double));
    maxorder = (double *)calloc((unsigned) ntimes,sizeof(double));
    meanvec = (double *)calloc((unsigned) ntimes,sizeof(double));

    hadamard(N,newhmat);

    for (i=0; i < ntimes;i++)
    {
        spectra(N,C,newhmat,noise,flam,&maxofxvec,&mean);
        maxorder[i] = maxofxvec;
        meanvec[i] = mean;
    }

    fp = fopen(filename,"w");
    for (i = 0;i < ntimes;i++)
        fprintf(fp,"%3.7lf %3.7lf\n", maxorder[i],meanvec[i]);
}

```

```

    fprintf(fp, "\n");
    fclose(fp);
}

/** Program to Obtain Walsh-Fourier Spectra **/

unsigned seed;

/*****
/* urand - simulates a vector of categorical white noise data */
*****/

double urand(C)
    int C;
{
    int ucat;
    double out, DIV, left, right;
    DIV = pow(2.0, 31.0) - 1;
    out = (double) random() / DIV;

    ucat = 0;
    left = 0.0;
    right = (1.0 / ((double) C));
    while (!(out > left && out <= right))
    {
        left += 1.0 / ((double) C);
        right += 1.0 / ((double) C);
        ucat += 1;
    }
    return (ucat);
}

/*****
/* spectra - calculated WFP of data from the signal */
/*      plus noise model                               */
*****/

```

```

void spectra(N,C,newhmat,noise,flam,maxofxvec,mean)
    int N,C,*newhmat;
    double *noise,*flam,*mean,*maxofxvec;
{
    int i,j,m;
    double summ,xbar;
    summ=0.0;
    xbar = 0.0;

    for(m=0;m < N;m++)
        noise[m] = urand(C);

    for(j=0;j < N;j++)
        summ += noise[j];
    xbar = summ/((double)N);

    for(j=0;j< N;j++)
        noise[j] = noise[j]-xbar;

    /* Calculating the Walsh spectrum */
    for (j = 0;j < N;j++)
    {
        flam[j] = 0.0;
        for (m = 0;m < N;m++)
            flam[j] += (1/sqrt((double)(N)))*
                        (noise[m] * newhmat[j + m*N]);
        flam[j] *= flam[j];
    }

    /***** Finding the maximum and the mean of flam *****/
    *maxofxvec = flam[0];
    for (j=1;j < N;j++)
    {
        if (flam[j] > *maxofxvec) *maxofxvec=flam[j];
        else *maxofxvec=*maxofxvec;
    }
    summ = 0.0;

```

```

*mean = 0.0;
for(j=0;j < N;j++)
    summ += flam[j];
*mean = summ/((double)N);
}

```

A.4 Walshclipar1.c: calculates the WFP for clipped AR(1) data

We only present the program to calculate the Walsh–Fourier periodogram from clipped AR(1) data. Clipped AR(2) or AR(3) data can be simulated with slight modifications to the “spectra” subroutine.

```

#include <math.h>
#include <stdio.h>
#include <string.h>
#include <malloc.h>

/***** Input Variables *****/
/* N - the sample size */
/* thr - the level at which the AR series is clipped */
/* coef - the coefficient of Zn(1) for the AR(1) model */
/* seed - input for the random number generator */
/* ntimes - number of simulations to run */
/* filename - name of file to which results are written */
/*****

/***** Main Program *****/

main(argc,argv) /*N, thr, coef, seed, ntimes, filename */
    int argc;
    char *argv[];
{
    void hadamard(),spectra(),boxmuller();
    int number,*newhmat,N,thr,i,j,m,ntimes,len,sequen,*seqvec;
    double *flam,*noise,*maxorder,maxofxvec,*meanvec,mean,coef,DIV;
    char *filename;
    FILE *fp;

```

```

N = atoi(argv[1]);
thr = atoi(argv[2]);
coef = atof(argv[3]);
srandom((seed =atoi(argv[4])));
ntimes = atoi(argv[5]);

maxofxvec = 0.0;
mean = 0.0;
DIV = pow(2.0,31.0)-1.0;
sequen = 0;

len = strlen(argv[6]);
filename =(char *) calloc((unsigned) len,sizeof(double));
strncpy(filename,argv[6],len);

newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
flam = (double *)calloc((unsigned) N,sizeof(double));
noise = (double *)calloc((unsigned) N,sizeof(double));
maxorder = (double *)calloc((unsigned) ntimes,sizeof(double));
meanvec = (double *)calloc((unsigned) ntimes,sizeof(double));
seqvec = (int *)calloc((unsigned) ntimes,sizeof(int));

hadamard(N,newhmat);

for (i=0; i < ntimes;i++)
{
    spectra(N,thr,coef,newhmat,noise,flam,
            &maxofxvec,&mean,&DIV,&sequen);
    maxorder[i] = maxofxvec;
    meanvec[i] = mean;
}

fp = fopen(filename,"w");
for (i = 0;i <=(ntimes-1);i++)
    fprintf(fp,"%3.7lf %3.7lf\n",maxorder[i],meanvec[i]);
fprintf(fp,"\n");

```

```

    fclose(fp);
    exit(0);
}

/***** Auxillary Programs *****/

unsigned seed;

/*****/
/* boxmuller - simulates standard normal random variables */
/* via the Box-Muller method */
/*****/

void boxmuller(x,y,DIV)
    double *x,*y,*DIV;
{
    int i,NN;
    double u[2];
    for (i=0;i<2;i++) u[i]=0.0;

    for (i=0;i< 2;i++) u[i] = (double)(random()/(*DIV));
    *x = sqrt((-2*log(u[0])))*cos(2*PI*u[1]);
    *y = sqrt((-2*log(u[0])))*sin(2*PI*u[1]);
}

/*****/
/* spectra - simulates clipped AR(1) data and */
/* calculates its WFP */
/*****/

void spectra(N,thr,coef,newhmat,noise,flam,
             maxofxvec,mean,DIV,sequen)
    int N,*newhmat,thr,*sequen;
    double *noise,*flam,*maxofxvec,*mean,coef,*DIV;
{
    int i,j,j1,NN;
    double x,y,summ,xbar,*Zn,*Xn,*flres;

```

```

summ = 0.0;
xbar = 0.0;
NN = N/2;

Zn = (double *)calloc((unsigned) N,sizeof(double));
Xn = (double *)calloc((unsigned) (2*N),sizeof(double));
flres = (double *)calloc((unsigned) (N*N),sizeof(double));

for (i=0;i <= (N-1);i++)
    for (j=0;j <= 1;j++)
        Xn[i + (j*N)] = -.5;

for (j=0;j <= (NN-1);j++)
{
    x=0.0;
    y=0.0;
    boxmuller(&x,&y,DIV);
    noise[j] = x;
    noise[NN+j] = y;
}

for (i=0;i <= (N-1);i++)
{
    if (i == 0) Zn[i] = 1.0;
    else Zn[i] = -coef*Zn[i-1] + noise[i];
    if (Zn[i] >= thr) j=1;
    else j=0;
    Xn[i+(j*N)] = .5;
}

/* Calculating the Walsh spectrum */
for (i = 0;i <= 1;i++)
{
    for (j = 0;j <= (N-1);j++)
{
    for (j1 = 0;j1<=(N-1);j1++)

```

```

        flres[j+(i*N)] += Xn[j1+(i*N)] *
                        ((double)newhmat[j+(j1*N)]);
    }
}

for (j=0;j<=(N-1);j++)
{
    flres[j+(i*N)] *= flres[j+(i*N)];
    flam[j] += flres[j+(i*N)];
}

for (i=0;i<=(N-1);i++)
    flam[i] /= ((double)N)/(2.0);

*maxofxvec = flam[0];
for (j=1;j < N;j++)
{
    if (flam[j] > *maxofxvec)
{
    *maxofxvec=flam[j];
    *sequen = j;
}
    else *maxofxvec=*maxofxvec;
}

/*****
    This part of the program can be modified to calculate
    the median or trimmed mean. The code for these
    modifications is found in the program walshmax.c
    *****/

summ=0.0;
for (j=0;j<=(N-1);j++)
    summ += flam[j];
*mean = summ/((double)N);

free((char *)Zn);
free((char *)Xn);

```



```

    free((char *)flres);
}

```

A.5 Walshreal.c: calculates the WFP of data read from a file

```

#include <math.h>
#include <stdio.h>
#include <string.h>
#include <malloc.h>

/***** Input Variables *****/
/* N - the sample size */
/* C - number of categories in the data */
/* filename1 - name of file from which data is read */
/* filename2 - name of file to which results are written */
/*****

/***** Main Program *****/
main(argc,argv) /* N, C, filename1, filename2 */
    int argc;
    char *argv[];
{
    void hadamard(),spectra();
    int number,*newhmat,N,C,i,j,m,len,length;
    double *flam,*data;
    char *filename1,*filename2;
    FILE *fp,*fpin;

    N = atoi(argv[1]);
    C = atoi(argv[2]);
    length = strlen(argv[3]);
    len = strlen(argv[4]);
    filename1 = (char *) calloc((unsigned) length,sizeof(double));
    filename2 = (char *) calloc((unsigned) len,sizeof(double));

```

```

    strncpy(filename1,argv[3],length);
    strncpy(filename2,argv[4],len);

/***** data- represents the data vector from the file*****/
    data = (double *)calloc((unsigned) N,sizeof(double));
    newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
    flam = (double *)calloc((unsigned) N,sizeof(double));

    if ((fpin=fopen(filename1,"r")) == NULL)
        printf("Sorry, can't open %s.",filename1);
    else
    {
        fpin = fopen(filename1,"r");
        for (i=0;i <= (N-1);i++)
            fscanf(fpin,"%lf ", &data[i]);
        fclose(fpin);
    }

    hadamard(N,newhmat);
    spectra(N,C,newhmat,flam,data);

    fp = fopen(filename2,"w");
    for (i = 0;i <= (N-1);i++)
        fprintf(fp,"%3.7lf\n",flam[i]);
    fclose(fp);
}

/*****
/* spectra - coerces input data into a matrix of dummy */
/*          variables and calculates its WFP          */
*****/
void spectra(N,C,newhmat,flam,data)
    int N,C,*newhmat;
    double *flam,*data;
{
    int i,j,k;
    double summ,*flres,*ucat;

```

```

flres = (double *)calloc((unsigned) (C*N),sizeof(double));
ucat = (double *)calloc((unsigned) (C*N),sizeof(double));

/* Transforming the data from a vector to a matrix */
for (i=0;i <= (N-1);i++)
{
    for (j=0;j <= (C-1);j++)
{
    if (data[i] == j+1)
        ucat[i +(j*N)] = (((double)(C-1))/((double)C));
    else ucat[i + (j*N)] = (-1.0)/((double)C);
}
}

/* Calculating the Walsh spectrum */
for (i = 0;i <= (C-1);i++)
{
    for (j = 0;j <= (N-1);j++)
{
    for (k = 0;k <= (N-1);k++)
        flres[j+(i*N)] += ucat[k+(i*N)] * ((double)newhmat[j+(k*N)]);
}
}

for (i=0;i<=(C-1);i++)
{
    for (j=0;j<=(N-1);j++)
flam[j] += (flres[j+(i*N)] * flres[j+(i*N)]);
}
for (i=0;i<=(N-1);i++)
    flam[i] /= ((double)N)/((double)C);

free((char *)ucat);
free((char *)flres);
}

```

A.6 Walshpulse.c: simulates square-wave data

```

#include <math.h>
#include <stdio.h>
#include <string.h>
#include <malloc.h>

/***** Input Variables *****/
/* N - the sample size */
/* C - number of categories in the data */
/* ntimes - number of simulations to run */
/* filename - name of file to which results are written */
/*****

/***** Main Program *****/
main(argc,argv) /* N, C, ntimes, filename */
{
    int argc;
    char *argv[];

    void hadamard(),spectra(),urand();
    int number,*newhmat,N,C,i,j,m,ntimes,len;
    double *flam,*ucat,*maxorder,maxofxvec,*meanvec,mean;
    char *filename;
    FILE *fp;

    N = atoi(argv[1]);
    C = atoi(argv[2]);
    ntimes=atoi(argv[3]);
    maxofxvec = 0.0;
    mean = 0.0;

    len = strlen(argv[4]);
    filename = (char *) calloc((unsigned) len,sizeof(double));
    strncpy(filename,argv[4],len);

    ucat = (double *)calloc((unsigned) C*N,sizeof(double));
    newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
    flam = (double *)calloc((unsigned) N,sizeof(double));

```

```

maxorder = (double *)calloc((unsigned) ntimes,sizeof(double));
meanvec = (double *)calloc((unsigned) ntimes,sizeof(double));

/** Spectra and hadamard programs can be found in walshmax.c **/
hadamard(N,newhmat);
for (i=0; i < ntimes;i++)
{
    spectra(N,C,newhmat,flam,ucat,&maxofxvec,&mean);
    maxorder[i] = maxofxvec;
    meanvec[i] = mean;
}

fp = fopen(filename,"w");
for (i = 0;i <= (ntimes-1);i++)
    fprintf(fp,"%3.7lf  %3.7lf\n", maxorder[i],meanvec[i]);
fprintf(fp,"\n");
fclose(fp);
}

/*****/
/* urand - simulates square-wave data */
/*****/
void urand(N,C,ucat)
    int N,C;
    double *ucat;
{
    int i,j;

    for (i=0;i<=(N-1);i++)
        for (j=0; j<=(C-1);j++)
            ucat[i+(j*N)] = (-1.0/((double)C));

    for (i = 0;i <= (C-1);i++)
        for (j = i;j <= (N-1);j+=C)
            ucat[j+(i*N)] = ((double)(C-1))/((double)C);
}

```

A.7 Walshdarma.c: calculates the logiacal covariance function of a DARMA(1,1) process

```

/***** Logicalcov *****/
/* N - the sample size */
/* rho - the correlation in the process (0 <= rho < 1) */
/* delta - 0 <= delta < 1 */
/* filename - name of file to which results are written */
/*****

void logicalcov(N,delta,newhmat,flam,DIV,maxofxvec,mean,sequen)
    int N,*newhmat,*sequen;
    double delta,*flam,*DIV,*maxofxvec,*mean;
{
    int j,k,m,dsum;
    double *tau,*temptau,rho,rhosq,hold,summ;

    rho = 0.0;
    rhosq = 0.0;
    hold = 0.0;
    temptau = (double *)calloc((unsigned) N,sizeof(double));
    tau = (double *)calloc((unsigned) N,sizeof(double));
    rho = random(seed)/(*DIV);
    rhosq = 1.0-rho;
    rhosq *= rhosq;
    hold = rho + rhosq*delta*(1.0-delta);
    hold *= hold;

    for (j = 0;j <= (N-1);j++)
    {
        temptau[j] = 0.0;
        for (k = 0;k <= (N-1);k++)
        {
            dsum = 0;
            dsum = (j~k) - k;
            if (dsum < 0)

```

```

        dsum = abs(dsum);
        temptau[j] += pow(rho,(double)dsum);
    }

    tau[j] = (1.0/(double)N)*(1.0/rho)*hold*temptau[j];
}
free((char *) temptau);
free((char *) tau);
}

main(argc,argv)    /* N, delta, seed, ntimes, filename */
    int argc;
    char *argv[];
{
    void hadamard(),logicalcov();
    int number,*newhmat,N,i,j,m,ntimes,len,sequen,*seqvec;
    double delta,DIV,*flam,*maxorder,maxofxvec,*meanvec,mean;
    char *filename;
    FILE *fp;

    N = atoi(argv[1]);
    delta = atof(argv[2]);
    seed = atof(argv[3]);
    ntimes = atoi(argv[4]);

    maxofxvec = 0.0;
    mean = 0.0;
    DIV = pow(2.0,31.0)-1;
    sequen = 0;

    len = strlen(argv[5]);
    filename = (char *) calloc((unsigned) len,sizeof(double));
    strncpy(filename,argv[5],len);

    newhmat = (int *)calloc((unsigned) N*N,sizeof(int));
    flam = (double *)calloc((unsigned) N,sizeof(double));

```

```

maxorder = (double *)calloc((unsigned) ntimes,sizeof(double));
meanvec = (double *)calloc((unsigned) ntimes,sizeof(double));
sequec = (int *)calloc((unsigned) ntimes,sizeof(int));

hadamard(N,newhmat);
for (i=0; i < ntimes; i++)
{
    logicalcov(N,delta,newhmat,flam,&DIV,&maxofxvec,&mean,&sequen);
    sequec[i] = sequen;
    maxorder[i] = maxofxvec;
    meanvec[i] = mean;
}

fp = fopen(filename,"w");
for (i = 0;i < N;i++)
    printf("%3.5lf ",tau[i]);
    printf("\n\n");
fclose(fp);
}

```

A.8 Pvalue.S: programs to calculate the power and size of Fisher's test

```

#####
# pvalmean - calculates the power (or size) of Fisher's #
#           test statistic given simulated data         #
#####

# maxvec - a vector of maxima from the WFP of non-white
#           (or white) noise data
# meanvec - vector of means (medians, trimmed means) of
#           the WFP of non-white (white) noise data
# denom - quantity to which the denominator of Fisher's
# test converges
# crit - the critical value of an extreme-value distribution
# N - sample size for the WFP
# C - number of categories in the data

```



```

pvalmean <- function(maxvec,meanvec,denom,crit,N,C)
{
  cnt1 <- 0
  stan <- standvals(maxvec,N,C)*(denom/meanvec)
  len <- length(stan)
  for (i in 1:len)
  {
    if (stan[i] > crit)
    {cnt1 <- cnt1+1}
  }
  per <- 100 * (cnt1/len)
  return(per)
}

```

A.9 Darma.S: Splus programs simulating DARMA and DAR models

```

#####
# rdarma - simulates a darma process of      #
#   length N with parameters beta and rho #
#####

```

```

rdarma <- function(N,beta=0.7,rho=0.4)
{
  Zn <- rep(0,N)
  Yn1 <- rpois(N+100,1)
  Zn[1] <- mean(Yn1)
  Yn <- Yn1[1:N+100]
  Xn <- rep(0,N)
  for (i in 2:N){
    pick1 <- runif(1)
    if (pick1 <= rho) Zn[i] <- Zn[i-1]
    else if (pick1 > rho) Zn[i] <- Yn[i]
  }
  for (i in 3:N){

```

```

    pick <- runif(2)
    if (pick[1] <= beta) Xn[i] <- Yn[i]
    else if (pick[1] > beta && pick[2] <= rho) Xn[i] <- Zn[i-2]
    else if (pick[1] > beta && pick[2] > rho) Xn[i] <- Yn[i-1]
  }
  return(Xn)
}

#####
# dar.corr - calculates the autocorrelation function      #
#   of a DAR(1) process with length n and coefficient p #
# dar.spec - calculates the spectrum of a DAR(1) process #
# dar.dat - simulates DAR(1) data                       #
#####

dar.corr <- function(p,n)
{
  out <- rep(0,n)
  for (j in 1:n) {out[j] <- p^j}
  return(out)
}

dar.spec <- function(p,n)
{
  w <- seq(-pi,pi,length=n)
  num <- 1-p^2
  denom <- 1+p^2-2*p*cos(w)
  spectrum <- (1/(2*pi))*(num/denom)
  return(spectrum)
}

dar.dat <- function(p,n)
{
  an <- rep(0,n)
  yn1 <- rpois((2*n),1)
  an[1] <- 1
  yn <- yn1[n:(2*n)]

```

```

for(i in 2:n)
{
  pick <- runif(1)
  #browser()
  if (pick <= p) an[i] <- an[i-1]
  else if (pick > p) an[i] <- yn[i]
}
return(an)
}

```

A.10 Other programs: Splus functions to accomplish various small calculations

```

#p - probabilities at which to calculate the quantiles
#N - sample size of the WFP (must be a power of 2)
#x - data vector (vector of maximum order statistics, most likely)
#C - number of categories for the noise series
#r - amount of trimming

```

```

#####
# standvals - standardizes the x vector by subtracting      #
#   asubn from each x and dividing the result by bsubn. #
#####

```

```

standvals <- function(x,N,C)
{
  asubn <- qchisq(1-(1/N),(C-1))
  numer <- x-asubn
  alpha <- (C-1)/2
  product <- gamma(alpha)*2^(alpha)
  bsubn <- ((4*N)/product)*exp((-asubn/2))*asubn^((alpha-1))
  standard <- numer/bsubn
  return(standard)
}

```

```

#####

```

```

# extremeval - calculates critical values for the extreme #
#   value distribution, given a vector of probabilities #
#####
extremeval <- function(p)
{
  out <- -log(-log(p))
  return(out)
}

#####
# medapp - approximates the 'ep'th percentage point of #
#   a chi-square distribution with v degrees of freedom #
#####
medapp <- function(C, ep=0.5)
{
  v <- C-1
  uni <- qnorm(ep)*(sqrt(2/(9*v)))
  y <- 1- (2/(9*v))
  out <- v*(uni + y)^3
  return(out)
}

#####
# truetrim - calculates the value to which the trimmed #
#   mean converges for a chi-square random variable #
#####
truetrim <- function(C,r,N)
{
  alp <- 1-r/N
  x <- qchisq(alp,C-1)
  u <- pchisq(x,C+1)
  denom <- (C-1)*(u/alp)
  return(denom)
}

```

Bibliography

- [1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions*, Dover Publications, New York, 1970.
- [2] K. G. Beauchamp. *Applications of Walsh and Related Functions*. Academic Press, London, 1984.
- [3] Laurel Beckett and Persi Diaconis. "Spectral Analysis for Discrete Longitudinal Data." Technical Report, Harvard University, 1991.
- [4] Erik Bølviken. "New Tests of Significance in Periodogram Analysis." *Scandinavian Journal of Statistics*, 10:1-9, 1983.
- [5] David R. Brillinger. "Comment: Walsh–Fourier Analysis and Its Statistical Applications." *Journal of the American Statistical Association*, 86:481-482, 1991.
- [6] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, second edition, 1991.
- [7] Hermann Chernoff, Joseph L. Gastwirth, and M. V. Johns. "Asymptotic Distributions of Linear Combinations of Functions of Order Statistics." *The Annals of Mathematical Statistics*, 38:52-72, 1967.
- [8] Ludwig Fahrmeir and Heinz Kaufmann. "Regression Models for Non-Stationary Categorical Time Series." *Journal of Time Series Analysis*, 8: 147-160, 1987.
- [9] William Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, New York, 1957.
- [10] R. A. Fisher. "Tests for Significance in Harmonic Analysis." *Proceedings of the Royal Society of London, Series A*, 125:54-59, 1929.
- [11] Halina Frydman and Burton Singer. "Assessing Qualitative Features of Longitudinal Data," in *Longitudinal Analysis of Labor Market Data*, J. Heckman and B. Singer, editors. Cambridge University Press, Cambridge, 1985.
- [12] Janos Galambos. *The Asymptotic Theory of Extreme Order Statistics*. John Wiley & Sons, New York, 1978.

- [13] B. V. Gnedenko. "On the Limiting Distribution of the Maximum Term in a Random Series," in *Breakthroughs in Statistics*, N. L. Johnson and S. Kotz, editors. Springer-Verlag, New York, 1991.
- [14] Larry Golstein and Louis Gordon. "Asymptotics for the Maximum of Half-Normal Plots." Technical Report, University of Southern California, 1994.
- [15] E. J. Gumbel. "Statistical Theory of Extreme Values," in *Contributions to Order Statistics*, Ahmed E. Sarhan and Bernard G. Greenburg, editors. John Wiley & Sons, New York, 1962.
- [16] Patricia A. Jacobs and Peter A. W. Lewis. "Discrete Time Series Generated by Mixtures I: Correlational Runs and Properties." *Journal of the Royal Statistical Society: Series B*, 40:94-105, 1978.
- [17] Patricia A. Jacobs and Peter A. W. Lewis. "Discrete Time Series Generated by Mixtures II: Asymptotic Properties." *Journal of the Royal Statistical Society: Series B*, 40:222-228, 1978.
- [18] Patricia A. Jacobs and Peter A.W. Lewis. "Stationary Discrete Autoregressive-Moving Average Time Series Generated by Mixtures." *Journal of Time Series Analysis*, 4:19-36, 1983.
- [19] Norman L. Johnson and Samuel Kotz. *Distributions in Statistics: Continuous Univariate Distributions I*. Houghton Mifflin Company, Boston, 1970.
- [20] R. Kohn. "On the Spectral Decomposition of Stationary Time Series Using Walsh Functions I." *Advances in Applied Probability*, 12:183-199, 1980.
- [21] R. Kohn. "On the Spectral Decomposition of Stationary Time Series Using Walsh Functions II." *Advances in Applied Probability*, 12:462-474, 1980.
- [22] J. L. LaBreque, D. V. Kent, and S. C. Cande. "Revised Magnetic Polarity Time Scale for Late Cretaceous and Cenozoic Time." *Geology*, 5:300-335, 1977.
- [23] William Lowrie and Walter Alvarez. "One Million Years of Geomagnetic Polarity History." *Geology*, 9:392-397, 1981.
- [24] Pedro A. Morettin. "Walsh Spectral Analysis." *SIAM Review* 23: 279-291, 1981.
- [25] Pedro A. Morettin. "Comment: Walsh-Fourier Analysis and Its Statistical Applications." *Journal of the American Statistical Association*, 86:482-483, 1991.

- [26] J. G. Negi, R. K. Tiwari, and K. N. N. Rao. "Comparison of Walsh and Fourier Spectroscopy of Geomagnetic Reversals and Nonsinusoidal Paleoclimatic Time Series." *IEEE Transactions on Geoscience and Remote Sensing*, **31**:127-135, 1993.
- [27] B. T. Polyak and Y. A. Schreider. "The Application of Walsh Functions in Approximate Calculations." *Voprosy Teorii Matematicheskikh Mashin*, II, 1962.
- [28] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, 1980.
- [29] Galen R. Shorack and Jon A. Wellner. *Empirical Processes with Applications to Statistics*. John Wiley & Sons, New York, 1988.
- [30] Andrew F. Siegel. "Testing for Periodicity in Time Series." *Journal of the American Statistical Association*, **75**:345-347, 1980.
- [31] D. S. Stoffer and T. Panchalingham. "A Walsh-Fourier Approach to the Analysis of Binary Time Series," in *Time Series and Econometric Modelling*, I. B. MacNeill and G. J. Umphrey, editors. D. Reidel Publishing Company, Boston, 1987.
- [32] David S. Stoffer. "Walsh-Fourier Analysis of Discrete-Valued Time Series." *Journal of Time Series Analysis*, **8**: 449-467, 1987.
- [33] David S. Stoffer. "Multivariate Walsh-Fourier Analysis." *Journal of Time Series Analysis*, **8**:449-467, 1990.
- [34] David S. Stoffer. "Walsh-Fourier Analysis and Its Statistical Applications." *Journal of the American Statistical Association*, **86**: 461-485, 1991.
- [35] David S. Stoffer, Mark S. Scher, Gale A. Richardson, Nancy L. Day, and Patricia A. Coble. "A Walsh-Fourier Analysis of the Effects of Moderate Maternal Alcohol Consumption on Neonatal Sleep-State Cycling." *Journal of the American Statistical Association*, **83**: 954-963, 1988.
- [36] David S. Stoffer, David E. Tyler, Andrew J. McDougall, and Gabriel Schachtel. "Spectral Analysis of DNA Sequences." *Bulletin of the ISI*, 1993.