WALSH–FOURIER ANALYSIS OF DISCRETE-VALUED TIME SERIES

By David S. Stoffer

University of Pittsburgh

Abstract. An approach to the analyses of discrete-valued time series is discussed. The analyses are accomplished in the spectral domain using the Walsh–Fourier transform which is based on Walsh functions. This approach will enable an investigator of discrete systems to analyse the data in terms of square waveforms and sequency rather than sine waves and frequency.

We develop a general signal-plus-noise type model for discrete-valued time series in which Walsh–Fourier spectral analysis is of interest. We consider the problems of detecting whether a common signal exists in repeated measures on discrete-valued time series and in discrete-valued processes collected in an experimental design. We show that these models may depend on unknown regression parameters and we develop consistent estimates of these parameters based on the finite Walsh–Fourier transform. Applications to certain Markov models are given; however, the methods presented also apply to non-Markov cases.

Keywords. Discrete-valued time series; Walsh–Fourier spectrum; Walsh–Fourier transform; discrete signal-plus-noise models; regression and analysis of power; Markov chains.

1. INTRODUCTION

Implicit in the Fourier (trigonometric) analysis of time series is one of two extreme assumptions about the process: (i) the very long stretch of the time series is the only time series we want to consider and consists of the superposition of not too many sinusoidal terms of substantially different frequencies; and (ii) the time series is to be regarded as a realization of an ergodic Gaussian process; it is one of many possible time series and the analyses are directed toward the properties of the ensemble of the series, not toward those of a specific realization (cf. Brillinger and Tukey, 1982).

There are, however, many physical situations in which time series are either positive or discrete and are patently non-normal, so that the analyses cannot be handled by transforming the data and applying Gaussian techniques (see, for example, Lewis, 1980, p. 154). Similarly there are processes, such as those which take values in a discrete finite set, which can neither be thought of as Gaussian, nor as the superpositions of well-separated sinusoids. Models for discrete-valued time series which have an ARMA structure are considered in Jacobs and Lewis (1978a, b, 1983) and in Lewis (1980). For the case of continuous-valued non-normal time series, it is perhaps still reasonable, in appropriate cases, to do spectral analysis via trigonometric methods. However, in the cases in which time series take values in a discrete (and possibly finite) set, it makes little statistical sense to correlate the data with sines and cosines. As an alternative, we suggest that the spectral analysis of discrete-valued time series be accomplished in the

0143-9782/87/04 0449–19 \$02.50/0
JOURNAL OF TIME SERIES ANALYSIS Vol. 8, No. 4

$$H_W(3)=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}$$

FIGURE 1. Sequency-ordered discrete Walsh functions $W(n, m/N)$, $n, m = 0, 1, \ldots, 7$, for a sample of size $N = 2^3$ as the rows of a Hadamard matrix.

'sequency' domain via the Walsh–Fourier transform (cf. Ahmed and Rao, 1975; Kohn, 1980a; Morettin, 1981). This seems to be a natural alternative to the usual Fourier analysis, since the Walsh–Fourier transform is based on the 'square-wave' Walsh functions. This approach would enable investigators to analyse discrete-valued time series (which we may think of as square waveforms) in terms of square waves and sequency (switches per unit time) rather than sine waves and frequency. As empirically demonstrated in Beauchamp (1975, chapter 5):

> These examples indicate clearly the respective roles of Walsh and Fourier spectral analysis for discontinuous and smooth-varying signals respectively. Where the signal is derived from a sinusoidally-based waveform ... then Fourier analysis is relevant. Where the signal contains sharp discontinuities and a limited number of levels ... then Walsh analysis is appropriate.

The Walsh functions, which are defined via the Rademacher functions (cf. Ahmed and Rao, 1975; Kohn, 1980a, or Morettin, 1981), form a complete orthonormal sequence on $[0, 1)$ and take on only two values, $+1$ and -1 (or 'on' and 'off'). They are ordered by the number of zero-crossings (or switches) which is called *sequency*. Let $W(n, \lambda)$, $n = 0, 1, 2, \ldots, 0 \leqslant \lambda < 1$, denote the nth sequency-ordered Walsh function, then $W(n, \cdot)$ makes n zero-crossings in $[0, 1)$. The first eight discrete, sequency-ordered Walsh functions $W(n, m/N)$, $m, n = 0, 1, \ldots, 7$, corresponding to a sample of length $N = 2^3$ are shown in Figure 1 in an 8×8 symmetric matrix called the Walsh-ordered Hadamard matrix, $H_W(3)$ (see Appendix A for details). We note that other orderings exist, for example, Paley order and Hadamard order are often used (cf. Ahmed and Rao, 1975); however, sequency or Walsh ordering is more natural in that it is comparable to the frequency ordering of sines and cosines. We shall discuss methods of generating the discrete Walsh functions in Appendix A.

Walsh spectral analysis has been used for several purposes, primarily in the engineering sciences, such as speech processing, word recognition, image coding and transmission, filtering and multiplexing. It has also been used to describe biological and medical systems such as monitoring EEG and ECG signals (see, for example, the *Proceedings on the Applications of Walsh Functions*; Ahmed and Rao, 1975; Beauchamp, 1975; Harmuth, 1972; to mention a few). Applications of Walsh functions in the statistics literature are rather scarce and we mention two.

Ott and Kronmal (1976) use the Walsh transform in classification and prediction problems for strictly stationary binary time series. Stoffer and Panchalingam (1987) analyse simulated and real binary time series in the sequency domain.

At present, there are two modes of development of Walsh spectral analysis in the literature. The first mode is termed *Walsh spectral analysis* and is developed via the concept of *dyadic stationarity*. That is, it is based on processes $\{X(n); n = 0, 1, 2, \ldots\}$ for which $\text{cov}(X(n), X(n \oplus m)) = B(m)$ is a function only of the dyadic distance between n and $n \oplus m$, where $n \oplus m$ denotes the dyadic addition of n and m (cf. Morettin, 1974b, 1981, for definitions, discussions and references). In this mode, one has in mind that the process of interest is the superposition of not too many Walsh functions of substantially different sequencies, that is,

$$X(n) = \sum_{k=1}^{K} Z(k)W(n, \lambda_k)$$

where $Z(1), \ldots, Z(K)$ are uncorrelated random variables with mean zero and variance σ_k^2, $k = 1, \ldots, K$, with $\lambda_1, \ldots, \lambda_K$ constants, $\lambda_i \neq \lambda_j$, $i \neq j$. The other mode of development is termed *Walsh–Fourier spectral analysis* and is based on *real-time stationarity*. Theoretical results concerning the statistical application of Walsh–Fourier spectral analysis are relatively recent and to the best of our knowledge, are limited to the works of Kohn (1980a, b), Morettin (1974a, 1981, 1983) and Stoffer (1985a).

One is warned in the examination of non-mathematical literature concerning this subject to keep in mind that the two different modes exist, although the particular mode is not always apparent. This matter can be quite confusing since the results are considerably different and results from one mode do not typically apply to the other. An excellent review of the two different approaches is given in Morettin (1981). We believe that although dyadic time has some theoretical appeal in the Walsh spectral domain, owing to its strange behaviour (see, for example, the discussions in Robinson, 1972; Beauchamp, 1975), it is of little practical use. We therefore concentrate on real-time stationary processes.

A brief account of the existing Walsh–Fourier theory, as well as some new results and necessary tools to be used in the sequel, are given in the next section. In Section 3 we present a general signal-plus-noise type model for discrete-valued time series in which Walsh–Fourier spectral analysis is of interest. Next we consider the problem of detecting whether a common signal exists in repeated measurements on discrete-valued time series, and mention that this method extends to discrete-valued time series collected in an experimental design. We then consider the analysis of discrete signal-plus-noise models in a regression setting in which the signal is observable but the process depends on regression parameters. Our main goal is to estimate, via Walsh–Fourier spectral methods, the regression parameters and the noise spectrum. In all cases we give concrete examples by showing that certain Markov chains satisfy our models; however, the methods presented here also apply in non-Markov cases. For completeness, we provide a discussion in Appendix A on simple and economical methods for generating the discrete sequency-ordered Walsh functions and hence the finite Walsh–Fourier transform.

2. PRELIMINARIES

In this section we give definitions, establish some new results, and give a brief summary of the existing Walsh–Fourier theory for stationary time series which we use in the sequel. In particular, we concentrate on those properties which we can directly apply to discrete-valued time series. Hence, for example, we do not consider properties of processes which are generalized linear processes; such processes can be handled by the theory given in Kohn (1980a, b), Morettin (1983) and Stoffer (1985a). At present, our discussion will be for univariate time series; the multivariate versions follow in an obvious way and we mention them briefly at the end of this section.

Let $X(0), X(1), \ldots, X(N-1)$ by a sample of length $N = 2^p$, $p > 0$ integer, from a weakly stationary time series $\{X(n), n = 0, \pm 1, \pm 2, \ldots\}$ with absolutely summable autocovariance function $\gamma(h) = \text{cov}(X(n), X(n+h))$, $h = 0, \pm 1, \pm 2, \ldots$. We assume for now that the constant mean value of $X(n)$ is zero. Let $W(n, \lambda)$ be the nth Walsh function in sequency order, and let

$$d_N(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} X(n)W(n, \lambda), \qquad 0 \leqslant \lambda < 1 \tag{2.1}$$

be the finite (or discrete) Walsh–Fourier transform of the data. The logical covariance of $X(n)$ (cf. Robinson, 1972; Kohn, 1980a) is defined to be

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

where by $j \oplus k$ we mean the dyadic addition of j and k. It can then be shown (cf. Kohn, 1980a) that the variance of $d_N(\lambda)$ is given by

$$\text{var}\{d_N(\lambda)\} = \sum_{j=0}^{N-1} \tau(j)W(j, \lambda). \tag{2.2}$$

Taking the limit $(N \to \infty)$ in (2.2) we have that $\text{var}\{d_N(\lambda)\} \to f(\lambda)$, where

$$f(\lambda) = \sum_{j=0}^{\infty} \tau(j)W(j, \lambda), \qquad 0 \leqslant \lambda < 1 \tag{2.3}$$

is called the Walsh–Fourier spectral density of $X(n)$. We note that $f(\lambda)$ exists since the absolute summability of $\gamma(h)$ implies the absolute summability of $\tau(j)$. Specifically, Kohn (1980a, lemma 3) shows that if

$$\lim_{n \to \infty} \sum_{|j| < 2^n} (1 - |j|/2^n)|\gamma(j)| < \infty \tag{2.4}$$

then $\sum_{j=0}^{\infty} |\tau(j)| < \infty$ and $f(\lambda)$ is well-defined.

If $X(0), X(1), \ldots, X(N-1)$ is a sample of length $N = 2^p$, the finite transform (2.1) is calculated for $\lambda_N = m/N$, $m = 0, 1, \ldots, N-1$. Since the discrete Walsh functions are symmetric in their arguments for $N = 2^p$, that is,

$$W(n, m/N) = W(m, n/N) \qquad (m, n = 0, 1, \ldots, N-1) \tag{2.5}$$

the value of λ_N in the finite Walsh–Fourier transform corresponds to sequency. As with the usual Fourier analysis, if the mean of the series is unknown, the only sequency of the form $\lambda_N = m/N$ for which the transform cannot be evaluated is at the zero $(m = 0)$ sequency. To see this, let $\theta = EX(n)$, all n, and note that for $m = 0, 1, \ldots, N-1$,

$$N^{-1} \sum_{n=0}^{N-1} W(n, m/N) = \delta_0^m \tag{2.6}$$

where δ is the Kronecker delta (see Kohn, 1980a, lemma 1). It is clear from (2.6) that the mean-centred transform will be the uncentred transform except at $m = 0$, and in particular

$$E\{d_N(m/N)\} = N^{-1/2} \sum_{n=0}^{N-1} \theta W(n, m/N) = N^{1/2}\theta\delta_0^m \qquad m = 0, 1, \ldots, N-1.$$

Kohn (1980a, corollary 3) gives the following useful results on the convergence of the second moment of the finite Walsh–Fourier transform under condition (2.4). Let λ_N be dyadically rational (that is, its binary representation is finite). If $\lambda_N \oplus \lambda \to 0$ as $N = 2^p \to \infty$, then

$$E\{d_N^2(\lambda_N)\} \to f(\lambda). \tag{2.7}$$

In general, the asymptotic covariance of the Walsh–Fourier transform at two distinct sequencies is not zero (cf. Kohn, 1980a, theorem 3). However, if $\lambda_{1,N}$ and $\lambda_{2,N}$ are dyadically rational and $|\lambda_{1,N} - \lambda_{2,N}| \geqslant N^{-1}$ with $\lambda_{i,N} \oplus \lambda \to 0$, $i = 1, 2$ as $N = 2^p \to \infty$, then

$$E\{d_N(\lambda_{1,N})\, d_N(\lambda_{2,N})\} \to 0.$$

Various authors have established central limit theorems for the finite Walsh–Fourier transform under a wide range of conditions (cf. Kohn, 1980a; Morettin, 1983; Stoffer, 1985a). We state three versions which are applicable to discrete-valued time series. The first version (Assumption 2.1) follows its trigonometric counterpart given in Hannan (1973) and can be found in Kohn (1980a, theorem 4). The second version (Assumption 2.2) follows its trigonometric counterpart based on the existence of higher moments given in Brillinger (1975) and can be found in Morettin (1983, theorem 1). We remark that the above two versions exist side-by-side and that neither is included in the other (cf. Morettin, 1983). The third version (Assumption 2.3), although not vital to the developments which follow, is included to enrich the application possibilities. This version establishes a central limit theorem for discrete-valued second-order stationary processes which satisfy a type of finite dependence property. A proof is provided in Appendix B.

ASSUMPTION 2.1. *$X(n)$ is strictly stationary with zero mean. Let $\mathscr{F}_n$ be the σ-field generated by $\{X(j), j \leqslant n\}$; put*

$$\alpha_j = [E\{E(X(n)\,|\,\mathscr{F}_{n-j}) - E(X(n)\,|\,\mathscr{F}_{n-j-1})\}^2]^{1/2}, \qquad j \geqslant 0$$

$\mathscr{F}_{-\infty}$ is trivial and $\sum_{j=0}^{\infty} \alpha_j < \infty$.

ASSUMPTION 2.2. *$X(n)$ is strictly stationary with zero mean and finite moments. Let $C_r(j_1, \ldots, j_r) = \operatorname{cum}\{X(j_1), \ldots, X(j_r)\}$ be the* rth *cumulant of $X(n), j_1, \ldots, j_r = 0, \pm 1, \pm 2, \ldots$;*

$$\sum_{j_1=0}^{\infty} \cdots \sum_{j_{r-1}=0}^{\infty} |C_r(j_1, \ldots, j_{r-1})| < \infty.$$

ASSUMPTION 2.3. *$X(n)$ is second-order stationary with zero mean and covariance function $\gamma(k)$; $\sup_n E\{|X(n)|^{2+\delta}\} < \infty$ for some $\delta > 0$. Let $\mathscr{F}_n$ be the σ-field generated by $\{X(j), j \leqslant n\}$. There exists a positive integer κ such that*
(i) *$E\{X(n)\,|\,\mathscr{F}_{n-\kappa}\} = 0$ a.s.; and*
(ii) *$E\{X(n)X(n+k)\,|\,\mathscr{F}_{n-\kappa}\} = \gamma(k)$ a.s. for $k = 0, 1, \ldots, \kappa - 1$.*

If condition (2.4) and either Assumption 2.1, 2.2 or 2.3 hold, then $d_N(\lambda)$ converges in distribution to a normal variate with mean zero and variance $f(\lambda)$ given by (2.3). We note, for example, that discrete-valued time series which are based on mixtures, such as the discrete MA processes described in Jacobs and Lewis (1978a, b, 1983), Lewis (1980), and the geometric processes discussed in Langberg and Stoffer (1987), satisfy the dependence properties of Assumption 2.3.

In order to be able to consistently estimate the Walsh–Fourier spectrum, we need asymptotic results for smoothing the Walsh–Fourier periodogram, $I_N(\lambda_N) = d_N^2(\lambda_N)$. We now state the following results in a theorem for use in the sequel. The theorem combines the three previous central limit theorems with a theorem given in Kohn (1980a, theorem 4).

THEOREM 2.1. *Let condition (2.4) and either Assumption 2.1, 2.2 or 2.3 hold. Let $\lambda_{j,N} = j/N$, $1 \leqslant j \leqslant N-1$, and suppose for $\{\lambda_{j(1),N}, \ldots, \lambda_{j(M),N}\}$, $\lambda_{j(m),N} \oplus \lambda \to 0$ as $N \to \infty$, $m = 1, \ldots, M$ and $|\lambda_{j(l),N} - \lambda_{j(k),N}| \geqslant N^{-1}$ for $l \neq k$, $l, k = 1, \ldots, M$. Then $\boldsymbol{d}_N \xrightarrow{\mathscr{D}} N(\boldsymbol{0}, \Delta)$, where $\boldsymbol{d}_N = (d_N(\lambda_{j(1),N}), \ldots, d_N(\lambda_{j(M),N}))'$ and Δ is an $M \times M$ diagonal matrix with $f(\lambda)$ along the diagonal. Also*

$$\boldsymbol{d}_N' \boldsymbol{d}_N \xrightarrow{\mathscr{D}} f(\lambda)\chi_M^2 \tag{2.8}$$

so that $M^{-1}\boldsymbol{d}_N'\boldsymbol{d}_N$ is an estimate of $f(\lambda)$ having variance $2f^2(\lambda)/M$.

If we let $M \to \infty$ as $N \to \infty$ with $M/N \to 0$ in Theorem 2.1, the smoothed periodogram $M^{-1}\boldsymbol{d}_N'\boldsymbol{d}_N$ is a mean square consistent estimate of the Walsh–Fourier spectrum $f(\lambda)$, $0 < \lambda < 1$.

Results obtained for the univariate finite Walsh–Fourier transform carry over, in an obvious way, to the vector case, say $\boldsymbol{X}(n) = (X_1(n), \ldots, X_r(n))'$, except that (cf. Kohn, 1980b, section 3):

(i) The logical covariance will now be $r \times r$ matrices given by

$$\tau(j) = N^{-1} \sum_{k=0}^{N-1} \{\Gamma(j \oplus k - k) + \Gamma'(j \oplus k - k)\} \tag{2.9}$$

where $\Gamma(h)$ is the $r \times r$ autocovariance matrix of $\boldsymbol{X}(n) = (X_1(n), \ldots, X_r(n))'$.

(ii) The Walsh–Fourier spectrum $f(\lambda)$ is an $r \times r$ real positive semidefinite matrix.

(iii)
$$\operatorname{cov}\{I_{ij}(\lambda) I_{lm}(\lambda)\} \rightarrow f_{il}(\lambda) f_{jm}(\lambda) + f_{im}(\lambda) f_{lj}(\lambda) \tag{2.10}$$

as $N \rightarrow \infty$ where $I_{ij}(\lambda)$ is the (i, j)th element of the $r \times r$ periodogram matrix $I_N(\lambda) = \boldsymbol{d}_N^X(\lambda) \boldsymbol{d}_N^{X'}(\lambda)$ with

$$\boldsymbol{d}_N^X(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} \boldsymbol{X}(n) W(n, \lambda).$$

3. SIGNAL-PLUS-NOISE MODELS FOR DISCRETE-VALUED TIME SERIES

In this section we discuss models for discrete systems in which Walsh–Fourier analysis is desirable. We consider a discrete version of the signal-plus-noise models used for sinusoidal and for Gaussian processes. In general, write the $r \times 1$ vector, discrete-valued time series as

$$\boldsymbol{X}(n) = \boldsymbol{S}(n) + \boldsymbol{\varepsilon}(n) \qquad (n = 0, \pm 1, \pm 2, \ldots) \tag{3.1}$$

where $\boldsymbol{S}(n)$ is a random stationary discrete signal which possibly depends on unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, and $\boldsymbol{\varepsilon}(n)$ is a zero-mean discrete-valued process (possibly white noise) which is *uncorrelated* with $\boldsymbol{S}(n)$. We note that the support of $\boldsymbol{X}(n)$, $\boldsymbol{S}(n)$ and $\boldsymbol{\varepsilon}(n)$ need not necessarily be the same and that there may be some *dependence* structure between $\boldsymbol{S}(n)$ and $\boldsymbol{\varepsilon}(n)$.

For a specific example of such a process, consider a macro model on a finite state space (cf. Basawa and Prakasa Rao, 1980). Let $X_j(n)$, $j = 1, \ldots, r$ denote the number of individuals in state j at time n. In particular, $X_j(n)$ is the aggregate over several independent chains evolving simultaneously. Let θ_j, $j = 1, \ldots, r$, be the probability of being in state j at any given time, and let Q denote the total number of individuals under consideration. Denote the $r \times 1$ vector by $\boldsymbol{X}(n) = (X_1(n), \ldots, X_r(n))'$ and suppose that $\boldsymbol{X}(n)$ is Markov with transition probabilities p_{ij}, $1 \leqslant i, j \leqslant r$. Then, for $n = 1, 2, \ldots$,

$$EX_j(n) = \sum_{i=1}^{r} EX_i(n-1) p_{ij}, \qquad 1 \leqslant j \leqslant r$$

from which we obtain the signal-plus-noise model:

$$X_j(n) = \sum_{i=1}^{r} X_i(n-1) p_{ij} + \varepsilon_j(n), \qquad 1 \leqslant j \leqslant r \tag{3.2}$$

where $\boldsymbol{\varepsilon}(n) = (\varepsilon_1(n), \ldots, \varepsilon_r(n))'$, $n = 0, 1, 2, \ldots$, is a zero-mean multinomial-type white noise process and is uncorrelated with the random signal

$$\boldsymbol{S}(n) = (S_1(n), \ldots, S_r(n))', \qquad S_j(n) = \sum_{i=1}^{r} X_i(n-1)p_{ij}, \qquad 1 \leqslant j \leqslant r.$$

It is easy to check the fact that $\boldsymbol{\varepsilon}(n)$ and $\boldsymbol{S}(n)$ are uncorrelated.

For the signal-plus-noise model presented in this section, Walsh–Fourier analysis would be useful for detecting whether a discrete signal exists in a given system and, if so, to determine the cyclic behaviour, in terms of sequency, of the signal. Moreover, for discrete systems in which the signal is observable but the process depends on unknown parameters, Walsh–Fourier methods can be used to consistently estimate the parameters as well as to consistently estimate the error spectrum. We discuss these types of analyses in the following subsections.

3.1. *Detecting a common signal*

Consider now the problem of detecting whether a common discrete-valued signal exists in Q replications of a discrete-valued time series $\{X_q(n)\}$, $q = 1, \ldots, Q$ which are of the signal-plus-noise form (3.1). Many of our techniques follow those of their trigonometric counterparts developed by Brillinger (1973; 1975, section 7.9; 1980).

As an example, suppose we wish to analyse Q independent queueing situations hypothesized to be similar. Let $X_q(n)$, $q = 1, \ldots, Q$, be the number of individuals in queue q at time n and suppose that $EX_q(n) = \theta_q$ for all n. Let $S(n)$ be a zero-mean stationary discrete-valued signal hypothesized to be common to all queues, such as a rate of change in the number of individuals in a queue, and let $\varepsilon_q(n)$, $q = 1, \ldots, Q$ be stationary zero-mean discrete-valued processes with common Walsh–Fourier spectrum. Whether a common signal exists among the Q queueing situations is expressed in whether the signal, or equivalently its Walsh–Fourier spectrum is identically zero.

In general, we suppose that the discrete system is of the form $X_q(n)$, $q = 1, \ldots, Q$; $n = 0, 1, \ldots, N-1$, $N = 2^p$, $p > 0$ integer, and can be modelled as

$$X_q(n) = \theta_q + S(n) + \varepsilon_q(n) \tag{3.3}$$

where θ_q are constants, $S(n)$ is a realization of a stationary discrete-valued time series with mean zero, and $\varepsilon_q(n)$, $q = 1, \ldots, Q$, are independent realizations of a zero-mean discrete-valued stationary time series which are uncorrelated with $S(n)$. Let $\gamma_{SS}(h)$ and $\gamma_{\varepsilon\varepsilon}(h)$, $h = 0, \pm 1, \pm 2, \ldots$, denote the autocovariance functions of $S(n)$ and $\varepsilon_q(n)$, $1 \leqslant q \leqslant Q$, respectively. We assume that $\gamma_{SS}(h)$ and $\gamma_{\varepsilon\varepsilon}(h)$ satisfy condition (2.4) and that every linear combination of $S(n)$ and $\varepsilon_q(n)$, $1 \leqslant q \leqslant Q$, satisfies one of Assumptions 2.1, 2.2 or 2.3. Denote the logical covariances of $S(n)$ and $\varepsilon_q(n)$, $1 \leqslant q \leqslant Q$ by $\tau_{SS}(j)$ and $\tau_{\varepsilon\varepsilon}(j)$, respectively, and the respective Walsh–Fourier spectrum of $S(n)$ and $\varepsilon_q(n)$, $1 \leqslant q \leqslant Q$, by $f_{SS}(\lambda)$ and $f_{\varepsilon\varepsilon}(\lambda)$, $0 \leqslant \lambda < 1$. The following conditions hold:

(1) $EX_q(n) = \theta_q$;

(2) $\tau_{X_q X_q}(j) = \tau_{SS}(j) + \tau_{\varepsilon\varepsilon}(j)$;
(3) $f_{X_q X_q}(\lambda) = f_{SS}(\lambda) + f_{\varepsilon\varepsilon}(\lambda)$;
(4) $\tau_{X_q X_l}(j) = \tau_{SS}(j), \qquad q \neq l$;
(5) $f_{X_q X_l}(\lambda) = f_{SS}(\lambda), \qquad q \neq l$.

Before we proceed with the analyses we need the following lemma.

LEMMA 3.1. *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ such that for all n, $EX_n^2 < \infty$ and $EY_n^2 < \infty$. If $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$, $EX_n^2 \to EX^2$, and $EY_n^2 \to EY^2$ as $n \to \infty$, then $EX_n Y_n \to EXY$.*

PROOF. By Chung (1974, theorem 4.5.4) the conditions $EX_n^2 < \infty$, $X_n \xrightarrow{p} X$, and $EX_n^2 \to EX^2$ imply the uniform integrability of $\{|X_n|^2\}$. Similarly, $\{|Y_n|^2\}$ is uniformly integrable. By the Cauchy–Swartz inequality,

$$\int_A |X_n Y_n| \, d\mathscr{P} \leqslant \left[\int_A |X_n|^2 d\mathscr{P}\right]^{1/2} \left[\int_A |Y_n|^2 d\mathscr{P}\right]^{1/2}$$

for any set $A \in \mathscr{F}$ and hence by Chung (1974, theorem 4.5.3), the uniform integrability of $\{|X_n|^2\}$ and $\{|Y_n|^2\}$ implies the uniform integrability of $\{|X_n Y_n|\}$. Clearly $X_n Y_n \xrightarrow{p} XY$ and $EX_n Y_n < \infty$, all n; hence by Chung (1974, theorem 4.5.4), $EX_n Y_n \to EXY$. ■

The particular analyses of this section are based on the following theorem.

THEOREM 3.1. *Let $\{X_q(n);\ q = 1, \ldots, Q;\ n = 0, \ldots, N-1\}$ satisfy the model conditions (3.3). Let $\lambda_{j(m),N}$, $m = 1, \ldots, M$ be defined as in Theorem 2.1. Then the finite Walsh–Fourier transform*

$$d_{N,q}(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=0}^{N-1} X_q(n) W(n, \lambda_{j(m),N}) \qquad 1 \leqslant q \leqslant Q$$

has the representation

$$d_{N,q}(\lambda_{j(m),N}) = u_m + z_{q,m} \quad \text{a.s.} \tag{3.4}$$

as $N \to \infty$ where the u_m are independent $N(0, f_{SS}(\lambda))$ variates and the $z_{q,m}$ are independent $N(0, f_{\varepsilon\varepsilon}(\lambda))$ variates. Moreover, u_m and $z_{q,m}$ are mutually independent, $1 \leqslant m \leqslant M$, $1 \leqslant q \leqslant Q$.

PROOF. Taking the finite Walsh–Fourier transform of (3.3) we obtain

$$d_{N,q}(\lambda_{j(m),N}) = d_N^S(\lambda_{j(m),N}) + d_{N,q}^{\varepsilon}(\lambda_{j(m),N}), \qquad 1 \leqslant q \leqslant Q$$

where $d_N^S(\cdot)$ and $d_{N,q}^{\varepsilon}(\cdot)$ are the Walsh–Fourier transforms of $S(n)$ and $\varepsilon_q(n)$, respectively. The constant term drops out in view of (2.6) and the fact that $\lambda_{j(m),N} \neq 0$. Invoking Skorokhod's Representation Theorem (Skorokhod, 1956) and Theorem 2.1, $d_N^S(\lambda_{j(m),N})$ and $d_{N,q}^{\varepsilon}(\lambda_{j(m),N})$ have the representations $d_N^S(\lambda_{j(m),N}) \to u_m$ a.s. and $d_{N,q}^{\varepsilon}(\lambda_{j(m),N}) \to z_{q,m}$ a.s. as $N \to \infty$, where the u_m are independent $N(0, f_{SS}(\lambda))$ variates and the $z_{q,m}$ are independent $N(0, f_{\varepsilon\varepsilon}(\lambda))$ variates. It remains to show that u_m and $z_{q,m}$ are mutually independent and, since they are jointly normal, it suffices to show that they are uncorrelated. Since $S(n)$ and $\varepsilon_q(n)$ are uncorrelated we have that $E\{d_N^S(\lambda_{j(m),N})d_{N,q}^{\varepsilon}(\lambda_{j(m),N})\} = 0$ for all N, $m = 1, \ldots, M$, and $q = 1, \ldots, Q$. Since $\gamma_{SS}(h)$ and $\gamma_{\varepsilon\varepsilon}(h)$ satisfy condition (2.4) we have by (2.7) $E\{d_N^S(\lambda_{j(m),N})^2\} \to f_{SS}(\lambda)$ and $E\{d_{N,q}^{\varepsilon}(\lambda_{j(m),N})^2\} \to f_{\varepsilon\varepsilon}(\lambda)$ as $N \to \infty$, $1 \leqslant m \leqslant M$, $1 \leqslant q \leqslant Q$. Thus by Lemma 3.1 we have $E\{d_N^S(\lambda_{j(m),N})d_{N,q}^{\varepsilon}(\lambda_{j(m),N})\} \to E\{u_m z_{q,m}\}$ as $N \to \infty$ and hence $E\{u_m z_{q,m}\} = 0$, $m = 1, \ldots, M$, $q = 1, \ldots, Q$. ■

Using the representation (3.4), we may proceed with the problem of determining whether a common discrete-valued signal exists. Our analysis follows that of the analysis of random effects models (cf. Scheffé, 1959). Consider the quantities

$$d_{N,\cdot}(\lambda_{j(m),N}) = Q^{-1}\sum_{q=1}^{Q} d_{N,q}(\lambda_{j(m),N}), \qquad 1 \leqslant m \leqslant M,$$

$$\sum_{m=1}^{M} d_{N,\cdot}^2(\lambda_{j(m),N}), \tag{3.5}$$

$$\sum_{q=1}^{Q}\sum_{m=1}^{M} [d_{N,q}(\lambda_{j(m),N}) - d_{N,\cdot}(\lambda_{j(m),N})]^2. \tag{3.6}$$

By Cochran's Theorem and in view of the representation (3.4), the quantities (3.5) and (3.6) have the almost sure representations

$$\sum_{m=1}^{M} (u_m + z_{\cdot,m})^2 = [f_{SS}(\lambda) + f_{\varepsilon\varepsilon}(\lambda)/Q]\chi_M^2 \quad \text{a.s.} \tag{3.7}$$

and

$$\sum_{q=1}^{Q}\sum_{m=1}^{M} (z_{q,m} - z_{\cdot,m})^2 = f_{\varepsilon\varepsilon}(\lambda)\chi_{M(Q-1)}^2 \quad \text{a.s.} \tag{3.8}$$

respectively, as $N \to \infty$, where the χ^2 variates in (3.7) and (3.8) are independent. The fact that the χ^2 variates are independent follows from the fact that $\operatorname{cov}(z_{q,m} - z_{\cdot,m}, z_{\cdot,m}) = 0$.

The test of the null hypothesis that there is no common signal, $S(n) \equiv 0$ or equivalently $f_{SS}(\lambda) \equiv 0$, may be examined for particular sequencies λ by comparing

$$Q\sum_{m=1}^{M} d_{N,\cdot}^2(\lambda_{j(m),N})/M \Big/ \sum_{q=1}^{Q}\sum_{m=1}^{M} [d_{N,q}(\lambda_{j(m),N}) - d_{N,\cdot}(\lambda_{j(m),N})]^2/M(Q-1) \tag{3.9}$$

with an $F_{M,M(Q-1)}$ distribution. In view of (3.7) and (3.8), (3.9) has the representation

$$\frac{[f_{\varepsilon\varepsilon}(\lambda) + Qf_{SS}(\lambda)]}{f_{\varepsilon\varepsilon}(\lambda)} F_{M,M(Q-1)} \quad \text{a.s.}$$

as $N \to \infty$.

Using the representations (3.7) and (3.8), it is easy to find an asymptotically ($N \to \infty$) unbiased estimate of the Walsh–Fourier spectrum $f_{SS}(\lambda)$ of the unobserved signal, say $\hat{f}_{SS}(\lambda)$, $0 < \lambda < 1$. Denoting the numerator of (3.9) by $Q\hat{f}_{\bar{X}\bar{X}}(\lambda)$ and the denominator of (3.9) by $\hat{f}_{\varepsilon\varepsilon}(\lambda)$, we see that the desired quantity is

$$\hat{f}_{SS}(\lambda) = \hat{f}_{\bar{X}\bar{X}}(\lambda) - Q^{-1}\hat{f}_{\varepsilon\varepsilon}(\lambda). \tag{3.10}$$

The techniques of this section may easily be extended to more complicated designs. For example, a balanced one-way random effects design in which observations are discrete-valued time series is discussed in Stoffer (1985b, pp. 19–22); see also Brillinger (1980).

3.2. *Estimating regression parameters when the discrete signal is observable*

In this section we concentrate on the discrete signal-plus-noise model where the signal is observable but the vector process of interest, $\boldsymbol{X}(n)$, $n = 0, 1, \ldots, N-1$, depends on unknown regression parameters. At the end of this section, we apply the results of this section to estimating the transition matrix in a Markov model. We suppose that we may write the model (3.1) in the $r \times 1$ vector form

$$\boldsymbol{X}(n) = \boldsymbol{\theta} + \beta \boldsymbol{S}(n) + \boldsymbol{\varepsilon}(n) \tag{3.11}$$

where $\boldsymbol{\theta}$ is an $r \times 1$ vector of constants, $\boldsymbol{S}(n)$ is a $q \times 1$ vector, observable discrete-valued signal, β is an $r \times q$ matrix of regression parameters, and $\boldsymbol{\varepsilon}(n)$ is a discrete-valued zero-mean white noise process which is uncorrelated with the signal $\boldsymbol{S}(n)$. Let $\Gamma_{SS}(h)$ denote the autocovariance matrix of $\boldsymbol{S}(n)$, where $\sum_{-\infty}^{\infty} \| \Gamma_{SS}(h) \| < \infty$. Further, we assume that every linear combination of the components of $\boldsymbol{\varepsilon}(n)$ satisfy one of Assumptions 2.1, 2.2 or 2.3. Let $\tau_{SS}(j)$ and $\tau_{\varepsilon\varepsilon}(j)$ be the logical covariance matrices of $\boldsymbol{S}(n)$ and $\boldsymbol{\varepsilon}(n)$, respectively (cf. 2.9) and note that $\tau_{\varepsilon\varepsilon}(j) = 0_r$ for $j \neq 0$, where 0_r is the $r \times r$ matrix of zeros. Let $f_{SS}(\lambda) = \{f_{S_a, S_b}(\lambda)\}$, $1 \leqslant a, b \leqslant q$, be the $q \times q$ positive definite Walsh–Fourier spectrum of $\boldsymbol{S}(n)$ and let $f_{\varepsilon\varepsilon}(\lambda) = \{f_{\varepsilon_a, \varepsilon_b}(\lambda)\}$, $1 \leqslant a, b \leqslant r$, be the $r \times r$ Walsh–Fourier spectrum of $\boldsymbol{\varepsilon}(n)$.

Taking transforms in (3.11) and following the previous section, we have the representation

$$\boldsymbol{d}_N^X(\lambda_{j(m), N}) = \beta \boldsymbol{d}_N^S(\lambda_{j(m), N}) + \boldsymbol{z}_m + \boldsymbol{0}_{\text{a.s.}}(1) \qquad (m = 1, \ldots, M) \tag{3.12}$$

where $\boldsymbol{0}_{\text{a.s.}}(1)$ denotes a vector variate tending to the zero vector almost surely as $N \to \infty$, $\boldsymbol{z}_m$ are independent multivariate $N(\boldsymbol{0}, f_{\varepsilon\varepsilon}(\lambda))$ variates;

$$\boldsymbol{d}_N^X(\lambda_{j(m), N}) = N^{-1/2} \sum_{n=0}^{N-1} \boldsymbol{X}(n) W(n, \lambda_{j(m), N})$$

and

$$\boldsymbol{d}_N^S(\lambda_{j(m), N}) = N^{-1/2} \sum_{n=0}^{N-1} \boldsymbol{S}(n) W(n, \lambda_{j(m), N})$$

denote the $r \times 1$ and $q \times 1$ finite Walsh–Fourier transforms of the vector time series $\boldsymbol{X}(n)$ and $\boldsymbol{S}(n)$, respectively.

If we assume that $M \geqslant q + r$, we may consider the asymptotic problem of estimating the parameter matrix β and the error spectrum in the multivariate analysis setting (see Anderson, 1958, chapter 8). Similarly, tests of hypotheses about the elements of β may be carried out by MANOVA techniques with an appropriate partitioning of that matrix (cf. Anderson, 1958, section 8.3). Following least-squares theory in view of (3.12), the estimate of β as a function of sequency is given by

$$\hat{\beta}_{M,N}(\lambda) = \left[\sum_{m=1}^{M} \boldsymbol{d}_N^X(\lambda_{j(m),N})\boldsymbol{d}_N^{S'}(\lambda_{j(m),N})\right]\left[\sum_{m=1}^{M} \boldsymbol{d}_N^S(\lambda_{j(m),N})\boldsymbol{d}_N^{S'}(\lambda_{j(m),N})\right]^{-1} \tag{3.13}$$

and the corresponding estimate of the error spectrum $f_{\varepsilon\varepsilon}(\lambda)$ is

$$\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} [\boldsymbol{d}_N^X(\lambda_{j(m),N}) - \hat{\beta}_{M,N}(\lambda)\boldsymbol{d}_N^S(\lambda_{j(m),N})] \times [\boldsymbol{d}_N^X(\lambda_{j(m),N}) - \hat{\beta}_{M,N}(\lambda)\boldsymbol{d}_N^S(\lambda_{j(m),N})]'. \tag{3.14}$$

Before discussing the consistency of the estimates $\hat{\beta}_{M,N}(\lambda)$ and $\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda)$, we state the following results which follow from the model assumptions and the results of Section 2. First write

$$f_{S_a,S_b}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} d_{N,a}^S(\lambda_{j(m),N})d_{N,b}^S(\lambda_{j(m),N}), \qquad 1 \leqslant a, b \leqslant q \tag{3.15a}$$

$$f_{\varepsilon_a,S_b}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} d_{N,a}^\varepsilon(\lambda_{j(m),N})d_{N,b}^S(\lambda_{j(m),N}), \quad 1 \leqslant a \leqslant r, \quad 1 \leqslant b \leqslant q \tag{3.15b}$$

and

$$f_{\varepsilon_a,\varepsilon_b}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} d_{N,a}^\varepsilon(\lambda_{j(m),N})d_{N,b}^\varepsilon(\lambda_{j(m),N}), \qquad 1 \leqslant a, b \leqslant r \tag{3.15c}$$

where $d_{N,a}^S(\cdot)$ and $d_{N,a}^\varepsilon(\cdot)$ denote the finite Walsh–Fourier transform of the ath element of $\boldsymbol{S}(n)$ and $\boldsymbol{\varepsilon}(n)$, respectively. Then as $M, N \to \infty$ with $M/N \to 0$ we have

(1) $Ef_{S_a,S_b}^{M,N}(\lambda) \to f_{S_a,S_b}(\lambda)$;
(2) $Ef_{\varepsilon_a,\varepsilon_b}^{M,N}(\lambda) \to f_{\varepsilon_a,\varepsilon_b}(\lambda)$;
(3) $Ef_{\varepsilon_a,S_b}^{M,N}(\lambda) \equiv 0$;
(4) $M\ \mathrm{var}\{f_{S_a,S_b}^{M,N}(\lambda)\} \to 2[f_{S_a,S_b}(\lambda)]^2$;
(5) $M\ \mathrm{var}\{f_{\varepsilon_a,\varepsilon_b}^{M,N}(\lambda)\} \to 2[f_{\varepsilon_a,\varepsilon_b}(\lambda)]^2$;
(6) $M\ \mathrm{var}\{f_{\varepsilon_a,S_b}^{M,N}(\lambda)\} \to f_{\varepsilon_a,\varepsilon_a}(\lambda)f_{S_b,S_b}(\lambda)$.

It follows from the above that as $M, N \to \infty$ with $M/N \to 0$, $f_{S_a,S_b}^{M,N}(\lambda) \to f_{S_a,S_b}(\lambda)$, $f_{\varepsilon_a,\varepsilon_b}^{M,N}(\lambda) \to f_{\varepsilon_a,\varepsilon_b}(\lambda)$, and $f_{\varepsilon_a,S_b}^{M,N}(\lambda) \to 0$ in mean square and hence in probability.

We now show the consistency of the estimates (3.13) and (3.14).

THEOREM 3.2. *Let the model assumptions (3.11) be satisfied. Then as $M, N \to \infty$ with $M/N \to 0$ the estimates $\hat{\beta}_{M,N}(\lambda)$ given in (3.13) and $\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda)$ given in (3.14) are consistent in probability for β and $f_{\varepsilon\varepsilon}(\lambda)$, $0 < \lambda < 1$, respectively.*

PROOF. Using the definitions of (3.15), write

$$\hat{\beta}_{M,N}(\lambda) = \beta + f_{\varepsilon S}^{M,N}(\lambda)[f_{SS}^{M,N}(\lambda)]^{-1} \tag{3.16}$$

where $f_{\varepsilon S}^{M,N}(\lambda) = \{f_{\varepsilon_a, S_b}^{M,N}(\lambda)\}$ and $f_{SS}^{M,N}(\lambda) = \{f_{S_a, S_b}^{M,N}(\lambda)\}$ are $r \times q$ and $q \times q$ matrices, respectively. By previous results we know that as $M, N \to \infty$ with $M/N \to 0$, $f_{SS}^{M,N}(\lambda) \xrightarrow{p} f_{SS}(\lambda)$ which is positive definite, and $f_{\varepsilon S}^{M,N}(\lambda) \xrightarrow{p} 0_{r\times q}$ where $0_{r\times q}$ is an $r \times q$ matrix of zeros. In view of (3.16) we see that $\hat{\beta}_{M,N}(\lambda)$ is consistent for β.

To show the consistency of $\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda)$ given by (3.14), first note that we may write the $r \times r$ matrix $f_{\varepsilon\varepsilon}^{M,N}(\lambda) = \{f_{\varepsilon_a, \varepsilon_b}^{M,N}(\lambda)\}$ whose elements are given by (3.15c) as

$$f_{\varepsilon\varepsilon}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} [\boldsymbol{d}_N^X(\lambda_{j(m),N}) - \beta \boldsymbol{d}_N^S(\lambda_{j(m),N})][\boldsymbol{d}_N^X(\lambda_{j(m),N}) - \beta \boldsymbol{d}_N^S(\lambda_{j(m),N})]'.$$

Since $f_{\varepsilon\varepsilon}^{M,N}(\lambda)$ is consistent for $f_{\varepsilon\varepsilon}(\lambda)$, it suffices to show that $\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda) - f_{\varepsilon\varepsilon}^{M,N}(\lambda) \to 0_r$ in probability as $M, N \to \infty$ with $M/N \to 0$. To see this, write

$$\begin{aligned}\hat{f}_{\varepsilon\varepsilon}^{M,N}(\lambda) - f_{\varepsilon\varepsilon}^{M,N}(\lambda) &= M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^X(\lambda_{j(m),N}) \boldsymbol{d}_N^{S\prime}(\lambda_{j(m),N})[\beta - \hat{\beta}_{M,N}(\lambda)]' \\ &\quad + [\beta - \hat{\beta}_{M,N}(\lambda)] M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^S(\lambda_{j(m),N}) \boldsymbol{d}_N^{X\prime}(\lambda_{j(m),N}) \\ &\quad + [\hat{\beta}_{M,N}(\lambda) - \beta] M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^S(\lambda_{j(m),N}) \boldsymbol{d}_N^{S\prime}(\lambda_{j(m),N}) \beta' \\ &\quad + \hat{\beta}_{M,N}(\lambda) M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^S(\lambda_{j(m),N}) \boldsymbol{d}_N^{S\prime}(\lambda_{j(m),N})[\hat{\beta}_{M,N}(\lambda) - \beta].\end{aligned} \tag{3.17}$$

We consider the first term of (3.17), the other cases will follow by similar methods. Expanding $\boldsymbol{d}_N^X(\lambda_{j(m),N})$, we have

$$\begin{aligned}M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^X(\lambda_{j(m),N}) \boldsymbol{d}_N^{S\prime}(\lambda_{j(m),N})[\beta - \hat{\beta}_{M,N}(\lambda)]' &= \beta f_{SS}^{M,N}(\lambda)[\beta - \hat{\beta}_{M,N}(\lambda)]' \\ &\quad + f_{\varepsilon S}^{M,N}(\lambda)[\beta - \hat{\beta}_{M,N}(\lambda)]' \to 0_r\end{aligned}$$

in probability as $M, N \to \infty$ with $M/N \to 0$ by (3.16) and the convergence results for $f_{SS}^{M,N}(\lambda)$ and $f_{\varepsilon S}^{M,N}(\lambda)$. ■

We close this section with an application to Markov chains. In particular, consider the macro model given by (3.2) and suppose that we are interested in estimating the transition probabilities p_{ij}. We shall assume that detailed transitions are not available, otherwise, the p_{ij} may be estimated by MLE techniques, see Basawa and Prakasa Rao (1980, section 2.1). Recall that the model is of the form

$$X_l(n) = \sum_{k=1}^{r} X_k(n-1) p_{kl} + \varepsilon_l(n), \qquad 1 \leqslant l \leqslant r \tag{3.18}$$

where $X_l(n)$ is the (aggregate) number of individuals in state l (over several independent Markov chains) at time n and $\varepsilon_l(n)$ is zero-mean white noise which is uncorrelated with $X_k(n-1)$ for $k = 1, \ldots, r$, and r is the number of states. We suppose that data is available in the form $\{X_l(n),\ l = 1, \ldots, r;\ n = 0, \ldots, N-1\}$ and that the total number of individuals in the system, viz. $\sum_{l=1}^{r} X_l(n)$, is fixed. To avoid singularities, we remove $X_r(n)$ from the model (3.18). Let T be the total number of individuals in the system, then for $l = 1, \ldots, r-1$,

$$X_l(n) = T\left(1 - \sum_{j=1}^{r-1} p_{jl}\right) + \sum_{j=1}^{r-1}\left[X_j(n-1) + \sum_{k=1}^{r-1} X_k(n-1)\right]p_{jl} + \varepsilon_l(n). \quad (3.19)$$

Now, put $\theta_l = T(1 - \sum_{j=1}^{r-1} p_{jl})$, $S_j(n) = X_j(n-1) + \sum_{k=1}^{r-1} X_k(n-1)$ with $S_j(0) \equiv 0$, and let $\boldsymbol{X}(n) = (X_1(n), \ldots, X_{r-1}(n))'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{r-1})'$, $\boldsymbol{S}(n) = (S_1(n), \ldots, S_{r-1}(n))'$ and $\boldsymbol{\varepsilon}(n) = (\varepsilon_1(n), \ldots, \varepsilon_{r-1}(n))'$. We have already seen that $\boldsymbol{\varepsilon}(n)$ and $\boldsymbol{S}(n)$ are uncorrelated. In view of (3.19), the model (3.18) may now be represented in the form of (3.11), that is,

$$\boldsymbol{X}(n) = \boldsymbol{\theta} + P\boldsymbol{S}(n) + \boldsymbol{\varepsilon}(n), \qquad n \geqslant 1$$

where P is the $(r-1) \times (r-1)$ matrix of unknown transition probabilities

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{r-1,1} \\ p_{12} & p_{22} & \cdots & p_{r-1,2} \\ \vdots & \vdots & & \vdots \\ p_{1,r-1} & p_{2,r-1} & \cdots & p_{r-1,r-1} \end{bmatrix}.$$

If we choose M sequencies, $\lambda_{j(m),N}$, $m = 1, \ldots, M$, $\lambda_{j(m),N} \neq 0$ in such a way that $M \geqslant 2(r-1)$ we obtain the consistent estimate of the transition matrix

$$\hat{P}(\lambda) = f_{XS}^{M,N}(\lambda)[f_{SS}^{M,N}(\lambda)]^{-1}$$

where

$$f_{XS}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^X(\lambda_{j(m),N})\boldsymbol{d}_N^{S'}(\lambda_{j(m),N});$$

$$f_{SS}^{M,N}(\lambda) = M^{-1} \sum_{m=1}^{M} \boldsymbol{d}_N^S(\lambda_{j(m),N})\boldsymbol{d}_N^{S'}(\lambda_{j(m),N});$$

$$\boldsymbol{d}_N^X(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=0}^{N-1} \boldsymbol{X}(n)W(n, \lambda_{j(m),N});$$

$$\boldsymbol{d}_N^S(\lambda_{j(m),N}) = N^{-1/2} \sum_{n=1}^{N-1} \{\boldsymbol{X}(n-1) + J_{r-1}\boldsymbol{X}(n-1)\}W(n, \lambda_{j(m),N});$$

where J_{r-1} is an $(r-1) \times (r-1)$ matrix of ones.

In this case we also obtain a consistent estimate for $\boldsymbol{\theta}$, namely,

$$\hat{\theta}_l(\lambda) = T\left(1 - \sum_{j=1}^{r-1} \hat{p}_{jl}(\lambda)\right) \qquad (l = 1, \ldots, r-1)$$

where $\hat{p}_{jl}(\lambda)$ is the jlth element of $\hat{P}(\lambda)$. The consistent estimate of the error spectrum is

$$\hat{f}_{\varepsilon\varepsilon}(\lambda) = f_{XX}^{M,N}(\lambda) - \hat{P}(\lambda) f_{SS}^{M,N}(\lambda) \hat{P}'(\lambda).$$

The actual implementation of this estimation procedure is then based on the methods described in Appendix A.

APPENDIX A

In this appendix we discuss the computation of the finite Walsh–Fourier transform for samples of length $N = 2^p$, $p > 0$ integer. The discrete Walsh functions are calculated via the Hadamard matrix $H(p)$, which is defined to be the symmetric orthogonal $N \times N$ matrix whose (u, v)th element, $u, v = 0, 1, \ldots, N-1$, is equal to $\phi(u, v) = \prod_{j=0}^{p-1} (-1)^{u_j v_j}$, where the binary representations of u and v are given by $(u_{p-1}, u_{p-2}, \ldots, u_0)$ and $(v_{p-1}, v_{p-2}, \ldots, v_0)$, respectively, u_j and v_j are either 0 or 1. For example, with $N = 8$, the (1, 5) element of $H(3)$ is $\phi(1, 5) = (-1)^{0+0+1} = -1$, whereas the (3, 7) element is $\phi(3, 7) = (-1)^{0+1+1} = 1$. The Hadamard matrix gives the discrete Walsh functions as rows (or columns) in what is called natural or Hadamard ordering. To obtain the Walsh functions in sequency order, we can reorder the rows of $H(p)$ according to the number of sign changes. We denote the sequency or Walsh-ordered Hadamard matrix by $H_W(p)$. An alternate method of obtaining $H_W(p)$ from $H(p)$ uses 'bit-reversal Gray code' to rearrange the rows; however this is essentially the same technique as counting the sign changes (see Ahmed and Rao, 1975, for details). Since these approaches involve counting sign changes, they are not very efficient procedures. We shall discuss a fast Walsh-ordered Hadamard transform in a moment. The Hadamard matrix can be generated recursively, $H(0) = 1$, and

$$H(k+1) = \begin{bmatrix} H(k) & H(k) \\ H(k) & -H(k) \end{bmatrix} \qquad (k = 0, 1, 2, \ldots)$$

so, for example,

$$H(1) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad H(2) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Denote $H(2) = [\boldsymbol{h}_0(2), \boldsymbol{h}_1(2), \boldsymbol{h}_2(2), \boldsymbol{h}_3(2)]$, where $\boldsymbol{h}_i(2)$, $i = 0, 1, 2, 3$, is the ith column of $H(2)$, then the corresponding Walsh-ordered Hadamard matrix is $H_W(2) = [\boldsymbol{h}_0(2), \boldsymbol{h}_2(2), \boldsymbol{h}_3(2), \boldsymbol{h}_1(2)]$. The procedure of obtaining the Walsh-ordered Hadamard transform from its definition requires either storage of the Hadamard matrix, or recomputation whenever the elements of $H_W(p)$ are needed. Hence, either the sample length is restricted to about $p = 10$ or 20, or the procedure is extremely slow. There are, however, fast methods which can reduce the number of computations (additions and subtractions) by about $2^{p-1}/p$ times less than by using the definition. The method we discuss here may be found in Ahmed and Rao (1975, chapter 6), who also gives a computer subroutine. The Walsh–Hadamard matrix can be computed as

$$H_W(p) = \prod_{i=1}^{p} H_i(p) \cdot B \tag{A1}$$

where

$$H_i(p) = \begin{bmatrix} F_s & & & & & & & 0 \\ & G_s & & & & & & \\ & & \cdots & & & & & \\ & & & F_s & & & & \\ & & & & G_s & & & \\ & & & & & \cdots & & \\ & & & & & & F_s & \\ 0 & & & & & & & G_s \end{bmatrix}, \qquad s = 2^{i-1}$$

with

$$F_s = \begin{bmatrix} I_s & I_s \\ I_s & -I_s \end{bmatrix} \quad G_s = \begin{bmatrix} I_s & -I_s \\ I_s & I_s \end{bmatrix}$$

and I_s being the $s \times s$ identity matrix. The matrix B in (A1) is a matrix which bit-reverses the order of the data. For example, with $N = 2^3$, the bit-reversal of $1 = (0, 0, 1)$ is $4 = (1, 0, 0)$, and the bit-reversal of $3 = (0, 1, 1)$ is $6 = (1, 1, 0)$, so that $X(1)$ is exchanged with $X(4)$ and $X(3)$ is exchanged with $X(6)$ in the data vector. If $\boldsymbol{X} = (X(0), \ldots, X(N-1))'$ is the data vector, $N = 2^p$, the fast finite Walsh–Fourier transform is computed as

$$\boldsymbol{d}_N(\boldsymbol{\lambda}_N) = N^{-1/2} H_W(p) \boldsymbol{X} = N^{-1/2} \prod_{i=1}^{p} H_i(p) \cdot B\boldsymbol{X} \tag{A2}$$

where $\boldsymbol{\lambda}_N = (0/N, 1/N, \ldots, (N-1)/N)'$. For example, if $N = 2^3$, the Walsh-ordered Hadamard matrix can be decomposed as

$$H_W(3) = \begin{bmatrix} F_1 & & & 0 \\ & G_1 & & \\ & & F_1 & \\ 0 & & & G_1 \end{bmatrix} \begin{bmatrix} F_2 & 0 \\ & \\ 0 & G_2 \end{bmatrix} F_4 B$$

where

$$B = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The vector of periodogram ordinates is obtained by squaring each element of $\boldsymbol{d}_N(\boldsymbol{\lambda}_N)$ given in (A2). Let $I_N(m) = d_N^2(m/N)$, denote the mth periodogram ordinate, $m = 0, 1, \ldots, N-1$. It is seen that

$$I_N(m) = \sum_{j=0}^{N-1} \hat{\tau}(j) W(j, m/N)$$

where

$$\hat{\tau}(j) = N^{-1} \sum_{k=0}^{N-1} X(k) X(j \oplus k).$$

Employing relationships (2.5) and (2.6) we may write

$$\hat{\tau}(j) = N^{-1} \sum_{m=0}^{N-1} I_N(m) W(m, j/N). \tag{A3}$$

Thus, for large N, the quickest way to compute $\hat{\tau}(j)$ is to use the fast Walsh–Fourier transform twice, once to compute $I_N(m)$ and once to compute the right-hand side of (A3).

APPENDIX B

In this appendix we prove the central limit theorem stated in Section 2, which is based on condition (2.4) and Assumption 2.3. The proof of the theorem is based on a martingale central limit theorem given in Brown (1971, theorem 2) and a limit theorem for approximating sequences given in Anderson (1971, theorem 7.7.1). We now state and prove the theorem.

THEOREM. *Let $\{X(n); n = 0, \pm 1, \pm 2, \ldots\}$ satisfy condition (2.4) and Assumption 2.3. Then the discrete Walsh–Fourier transform, $d_N(\lambda)$, converges in distribution to a normal variate with mean zero and variance $f(\lambda)$ given by (2.3).*

PROOF. Fix $\lambda \in [0, 1)$ and write $d_N(\lambda) = N^{-1/2} \sum_{n=0}^{N-1} X(n)W(n, \lambda) \equiv N^{-1/2} \sum_{n=1}^{N} y_n$. Choose an integer k such that $2\kappa < k < N$ and put $T = [N/k]$, where $[\cdot]$ is the greatest integer function. For $j = 1, 2, \ldots, T$, define $Y_{j,k} = y_{(j-1)k+1} + y_{(j-1)k+2} + \cdots + y_{jk-\kappa}$ and $Z_{j,k} = y_{jk-\kappa+1} + y_{jk-\kappa+2} + \cdots + y_{jk}$, and let $u_{k,N} = N^{-1/2} \sum_{j=1}^{T} Z_{j,k}$ and $v_{k,N} = N^{-1/2}(y_{kT+1} + y_{kT+2} + \cdots + y_N)$. Clearly $d_N(\lambda) = N^{-1/2} \sum_{j=1}^{T} Y_{j,k} + u_{k,N} + v_{k,N}$. We will show that $v_{k,N} \to 0$ in mean square as $N \to \infty$ uniformly in k, $u_{k,N} \to 0$ in mean square as $k \to \infty$ uniformly in N, and $N^{-1/2} \sum_{j=1}^{T} Y_{j,k}$ converges in law to a normal variate as $N \to \infty$ and $k \to \infty$. Thus, the theorem will follow from Anderson (1971, theorem 7.7.1).

Using Minkowski's inequality with $r = N - Tk < k$ and the fact that $|W(n, \lambda)| = 1$,

$$E\{v_{k,N}^2\}^{1/2} = N^{-1/2}\{E(y_{kT+1} + \cdots + y_N)^2\}^{1/2} \leqslant N^{-1/2} \sum_{j=1}^{r} E\{|X(kT+j)|^2\}^{1/2},$$

from which we obtain

$$E\{v_{k,N}^2\} \leqslant r^2 N^{-1}\gamma(0) \leqslant k^2 N^{-1}\gamma(0),$$

and hence the desired result follows. In a similar manner we can show that

$$E\{u_{k,N}^2\} \leqslant TN^{-1}\kappa^2\gamma(0) \leqslant k^{-1}\kappa^2\gamma(0)$$

to obtain the desired result for $u_{k,N}$.

To establish a central limit theorem for $N^{-1/2} \sum_{j=1}^{T} Y_{j,k}$ we use a martingale central limit theorem found in Brown (1971, theorem 2). We shall follow Brown's notation. Let k be fixed and let $S_{k,T} = \sum_{j=1}^{T} Y_{j,k}$; recall that $\mathscr{F}_n$ is the σ-field generated by $\{X(j), j \leqslant n\}$. Then, $\{S_{k,T}, \mathscr{F}_{kT-\kappa}; T = 1, 2, \ldots\}$ form a martingale. To see this note that

$$E\{S_{k,T+1} | \mathscr{F}_{kT-\kappa}\} = \sum_{j=1}^{T} Y_{j,k} + E\{Y_{T+1,k} | \mathscr{F}_{kT-\kappa}\}.$$

The terms in $Y_{T+1,k}$ are $\{X(n)W(n, \lambda), n = kT + j; j = 1, 2, \ldots, k - \kappa\}$. We are given that $E\{X(kT + j) | \mathscr{F}_{kT+j-\kappa}\} = 0$ a.s. and since $\mathscr{F}_{kT-\kappa} \subset \mathscr{F}_{kT+j-\kappa}$, $1 \leqslant j \leqslant k - \kappa$, we have

$$E\{X(kT + j) | \mathscr{F}_{kT-\kappa}\} = EE\{X(kT + j) | \mathscr{F}_{kT+j-\kappa}\} | \mathscr{F}_{kT-\kappa} = 0 \quad \text{a.s.}$$

and hence $E\{Y_{T+1,k} | \mathscr{F}_{kT-\kappa}\} = 0$ a.s. as was to be shown. Following Brown (1971), put $V_{k,T}^2 = \sum_{j=1}^{T} E(Y_{j,k}^2 | \mathscr{F}_{k(j-1)-\kappa})$ and $s_{k,T}^2 = EV_{k,T}^2$. A typical term in $V_{k,T}^2$ is $E\{y_{(j-1)k+l}\, y_{(j-1)k+m} | \mathscr{F}_{k(j-1)-\kappa}\}$, $1 \leqslant l, m \leqslant k - \kappa$; $1 \leqslant j \leqslant T$. Let $n_1 = (j-1)k + l$ and $n_2 = (j-1)k + m$, then by hypotheses of the theorem,

$$\begin{aligned} E\{y_{n_1} y_{n_2} | \mathscr{F}_{(j-1)k-\kappa}\} &= W(n_1 \oplus n_2, \lambda) E\{X(n_1)X(n_2) | \mathscr{F}_{(j-1)k-\kappa}\} \\ &= W(n_1 \oplus n_2, \lambda)\gamma(n_2 - n_1) \quad \text{a.s.} \\ &= E\{y_{n_1} y_{n_2}\}. \end{aligned}$$

Hence, $V_{k,T}^2 = \sum_{j=1}^{T} EY_{j,k}^2 = EV_{k,T}^2 = s_{k,T}^2$ a.s. The application of Theorem 2 of Brown (1971) entails showing that as $T \to \infty$,

(i) $s_{k,T}^{-2} V_{k,T}^2 \xrightarrow[p]{} 1$

(ii) $s_{k,T}^{-2} \sum_{j=1}^{T} E\{Y_{j,k}^2 I(|Y_{j,k}| \geqslant \varepsilon s_{k,T})\} \to 0$

for any $\varepsilon > 0$, where $I(\cdot)$ is the indicator function. We have already established (i) and we proceed with establishing (ii). First, we show that $\sup_n E\{|X(n)|^{2+\delta}\} < \infty$ for some $\delta > 0$ implies that $\sup_j E\{|Y_{j,k}|^{2+\delta}\} < \infty$. Let $n_l = (j-1)k + l$ and $a(n_l) = W(n_l, \lambda)$. Then by Minkowski's inequality

$$\begin{aligned} E\{|Y_{j,k}|^{2+\delta}\}^{1/(2+\delta)} &= E\{|a(n_1)X(n_1) + \cdots + a(n_{k-\kappa})X(n_{k-\kappa})|^{2+\delta}\}^{1/(2+\delta)} \\ &\leqslant \sum_{l=1}^{k-\kappa} E\{|X(n_l)|^{2+\delta}\}^{1/(2+\delta)} \\ &\leqslant (k - \kappa)[\sup_n E\{|X(n)|^{2+\delta}\}]^{1/(2+\delta)} \end{aligned}$$

so that for any j,

$$E\{|Y_{j,k}|^{2+\delta}\} \leqslant (k-\kappa)^{2+\delta} \sup_n E\{|X(n)|^{2+\delta}\} < \infty$$

as was to be shown. Following a similar argument to the one above, we can show that $EY_{j,k}^2$ is uniformly bounded, and since $EY_{j,k}^2$ does not vanish as $j \to \infty$ we see that $s_{k,T}^2 = \mathcal{O}(T)$. Also, one can show that since $\sup_j E\{|Y_{j,k}|^{2+\delta}\} < \infty$ for some $\delta > 0$, there exists a random variable Y_k with $EY_k^2 < \infty$ such that

$$E\{Y_{j,k}^2 I(|Y_{j,k}| \geqslant u)\} \leqslant E\{Y_k^2 I(|Y_k| \geqslant u)\}$$

for all j and all $u \geqslant 0$. See Stoffer (1985a, lemma 3.2). Thus,

$$s_{k,T}^{-2} \sum_{j=1}^{T} E\{Y_{j,k}^2 I(|Y_{j,k}| \geqslant \varepsilon s_{k,T})\} \leqslant s_{k,T}^{-2} TE\{Y_k^2 I(|Y_k| \geqslant \varepsilon s_{k,T})\} \to 0$$

as $T \to \infty$. Hence, by the Brown (1971) martingale central limit theorem we have that for fixed k,

$$T^{-1/2} \sum_{j=1}^{T} Y_{j,k}^2 = (N^{1/2}T^{-1/2})N^{-1/2} \sum_{j=1}^{T} Y_{j,k}^2$$

converges in law to a normal variate whose variance depends on k as $T \to \infty$ or equivalently $N \to \infty$. Clearly then, for fixed k, $(N^{1/2}T^{-1/2}k^{-1/2})N^{-1/2} \sum_{j=1}^{T} Y_{j,k}^2$ converges in law to a normal variate as $N \to \infty$. Letting $k \to \infty$ establishes the desired result.

ACKNOWLEDGEMENT

Part of this work was supported by the Air Force Office of Scientific Research under Contracts F49620-85-C-0008 and AFOSR-84-0113.

REFERENCES

AHMED, N. and RAO, K. R. (1975) *Orthogonal Transforms for Digital Signal Processing*. New York: Springer-Verlag.

ANDERSON, T. W. (1958) *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

—— (1971) *The Statistical Analysis of Time Series*. New York: Wiley.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980) *Statistical Inference for Stochastic Processes*. London: Academic Press.

BEAUCHAMP, K. G. (1975) *Walsh Functions and Their Applications*. London: Academic Press.

BRILLINGER, D. R. (1973) The analysis of time series collected in an experimental design. In *Multivariate Analysis III* (ed. P. R. Krishnaiah). New York: Academic Press, pp. 241–56.

—— (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

—— (1980) Analysis of variance problems under time series models. In *Handbook of Statistics*, Vol. 1 (ed. P. R. Krishnaiah). Amsterdam: North Holland, pp. 237–78.

BRILLINGER, D. R. and TUKEY, J. W. (1982) Spectrum estimation and system identification relying on a Fourier transform. *Technical Report No. 5*. Department of Statistics, Berkeley, California.

BROWN, B. M. (1971) Martingale central limit theorems. *Ann. Math. Statist.* 42, 59–66.

CHUNG, K. L. (1974) *A Course in Probability Theory*. New York: Academic Press.

HANNAN, E. J. (1973) Central limit theorems for time series regression. *Z. Warscheinlichkeits* 26, 157–70.

HARMUTH, H. (1972) *Transmission of Information by Orthogonal Functions*. Berlin: Springer-Verlag.

KOHN, R. (1980a) On the spectral decomposition of stationary time series using Walsh functions. I. *Adv. Appl. Prob.* 12, 183–99.

—— (1980b) On the spectral decomposition of stationary time series using Walsh functions. II. *Adv. Appl. Prob.* 12, 462–74.

JACOBS, P. A. and LEWIS, P. A. W. (1978a). Discrete time series generated by mixtures I: correlational and runs properties. *J. Roy. Statist. Soc. B* 40(1), 94–105.

—— (1978b) Discrete time series generated by mixtures II: asymptotic properties. *J. Roy. Statist. Soc. B* 40(2), 222–8.

—— (1983) Stationary discrete autoregressive moving average time series generated by mixtures. *J. Time Series Anal.* 4, 18–36.

LANGBERG, N. A. and STOFFER, D. S. (1987) Moving average models with bivariate exponential and geometric distributions. *J. Appl. Prob.*, 24, 48–61.

LEWIS, P. A. W. (1980) Simple models for positive-valued and discrete-valued time series with ARMA correlation structure. In *Multivariate Analysis V* (ed. P. R. Krishnaiah), Amsterdam: North-Holland, pp. 151–66.

MORETTIN, P. A. (1974a) Limit theorems for stationary and dyadic-stationary processes. *Biol. Soc. Brasil. Mat.* 5, 97–104.

—— (1974b) Walsh-function analysis of a certain class of time series. *Stoch. Process. Appl.* 2, 183–93.

—— (1981) Walsh spectral analysis. *SIAM Rev.* 23, 279–91.

—— (1983) A note on a central limit theorem for stationary processes. *J. Time Series Anal.* 4, 49–52.

OTT, J. and KRONMAL, R. A. (1976) Some classification procedures for multivariate binary data using orthogonal functions. *J. Amer. Statist. Ass.* 71, 391–9.

ROBINSON, G. S. (1972) Discrete Walsh and Fourier power spectra. *Proc. Symp. Appl. Walsh Functions*, Washington, DC, 298–309.

SCHEFFÉ, H. (1959) *The Analysis of Variance.* New York: Wiley.

SKOROKHOD, A. V. (1956) Limit theorems for stochastic processes. *Theor. Prob. Appl.* 1, 261–90.

STOFFER, D. S. (1985a) Central limit theorems for finite Walsh–Fourier transforms of weakly stationary time series. *J. Time Series Anal.* 6, 261–7.

—— (1985b) Walsh–Fourier analysis of discrete-valued time series. *Technical Report* #85-46. Center for Multivariate Analysis, University of Pittsburgh.

STOFFER, D. S. and PANCHALINGAM, T. (1987) A Walsh–Fourier approach to the analysis of binary time series. In *Advances in the Statistical Sciences—Time Series and Econometric Modelling*, Vol. III. (ed. J. B. MacNeill and G. J. Umphrey) Netherlands: Reidel, in the press.