Ann. Inst. Statist. Math.
Vol. 41, No. 2, 205-225 (1989)

STATISTICAL ANALYSIS OF DYADIC STATIONARY PROCESSES*

M. TANIGUCHI**, L. C. ZHAO, P. R. KRISHNAIAH AND Z. D. BAI

Center for Multivariate Analysis, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

(Received December 24, 1987; revised May 11, 1988)

Abstract. In this paper we consider a multiple dyadic stationary process with the Walsh spectral density matrix $f_\theta(\lambda)$, where θ is an unknown parameter vector. We define a quasi-maximum likelihood estimator $\hat{\theta}$ of θ, and give the asymptotic distribution of $\hat{\theta}$ under appropriate conditions. Then we propose an information criterion which determines the order of the model, and show that this criterion gives a consistent order estimate. As for a finite order dyadic autoregressive model, we propose a simpler order determination criterion, and discuss its asymptotic properties in detail. This criterion gives a strong consistent order estimate. In Section 5 we discuss testing whether an unknown parameter θ satisfies a linear restriction. Then we give the asymptotic distribution of the likelihood ratio criterion under the null hypothesis.

Key words and phrases: Dyadic stationary process, information criterion, likelihood ratio criterion, quasi-maximum likelihood estimator, Walsh spectral density.

1. Introduction

There has been much discussion of Walsh spectral analysis for dyadic stationary processes. Morettin (1974) investigated some asymptotic properties of the finite Walsh transforms of dyadic stationary processes. Nagai (1977) gave the spectral representations for dyadic stationary processes. If we consider finite dyadic linear models, then the greatest differences between dyadic stationary processes and ordinary stationary processes appear. Nagai (1980) and Nagai and Taniguchi (1987) established that a dyadic autoregressive and moving average (DARMA) process of finite

*This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notations hereon.

**The work of this author was done at the Center for Multivariate Analysis. His permanent address is Department of Mathematics, Hiroshima University, Hiroshima 730, Japan.

order can be expressed as a dyadic autoregressive (DAR) process of finite order, and also as a dyadic moving average (DMA) process of finite order. Nagai and Taniguchi (1987) discussed the principal component analysis of a multiple dyadic process, and also the canonical correlation analysis. Morettin (1981) gave a convenient survey for Walsh spectral analysis.

In this paper we consider a multiple dyadic stationary process with the Walsh spectral density matrix $f_{\boldsymbol{\theta}}(\lambda)$, where $\boldsymbol{\theta}$ is an unknown parameter vector. We define a quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and give the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ under appropriate conditions. In Section 3 we propose an information criterion which determines the order of the model, and show that this criterion gives a consistent order estimate. In Section 4, for a finite order dyadic autoregressive model, we propose a simpler order determination criterion, and show that the estimated order has strong consistency. Also, some interesting examples are given in the identification problem for Walsh spectra. In Section 5 we discuss a testing problem to check whether the unknown parameter $\boldsymbol{\theta}$ satisfies a linear restriction. Then we give the asymptotic distribution of the likelihood ratio criterion under the null hypothesis. Throughout this paper we are dealing with one mode of development of Walsh spectral analysis, via the concept of dyadic stationarity. We also remark that the applications of the dyadic approach seem limited in the existing circumstances.

2. Dyadic stationary processes and estimation theory

First we introduce some basic concepts and notations. Denote by T the set of all nonnegative integers. Let x and y be nonnegative real numbers and have the following binary expansions:

$$x = \sum_{l=-\infty}^{\infty} x_l 2^l \quad \text{with} \quad x_l = 0 \text{ or } 1 \,,$$

$$y = \sum_{l=-\infty}^{\infty} y_l 2^l \quad \text{with} \quad y_l = 0 \text{ or } 1 \,.$$

Then the dyadic addition $\oplus$ is defined by

$$x \oplus y = \sum_{l=-\infty}^{\infty} |x_l - y_l| 2^l \,.$$

A stochastic process (possibly vector process) $\{Y(t) : t \in T\}$ is said to be dyadic stationary if the joint distribution of $Y(t_1), Y(t_2), \ldots, Y(t_n)$ is the same as that of $Y(t_1 \oplus t), Y(t_2 \oplus t), \ldots, Y(t_n \oplus t)$ for every finite set of nonnegative integers $\{t_1, \ldots, t_n\}$ and for every nonnegative integer t. For $\lambda \in [0, 1)$, we write it as

$$\lambda = \sum_{j=1}^{\infty} \lambda_j 2^{-j} ,$$

where the λ_j is either 1 or 0. We define the j-th Rademacher function, $\phi_j(\lambda)$, as

$$\phi_j(\lambda) = (-1)^{\lambda_{j+1}}, \quad j \in T .$$

The Walsh functions $\{W(t,\lambda), t \in T, \lambda \in [0,1)\}$ are defined as follows:

(i) $W(0,\lambda) = 1, \lambda \in [0,1)$,

(ii) If $t = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_r}$, with $n_1 > n_2 > \cdots > n_r \geq 0$,

then $W(t,\lambda) = \phi_{n_1}(\lambda)\phi_{n_2}(\lambda)\cdots\phi_{n_r}(\lambda)$, $\lambda \in [0,1)$. $W(t,\lambda)$ is called the t-th Walsh function in Paley ordering. The properties of Walsh functions are well known:

(i) for each $t \in T$ and $\lambda \in [0,1)$, the value of $W(t,\lambda)$ is only $+1$ or -1,

(ii) for any s, $t \in T$,

$$W(t,\lambda)W(s,\lambda) = W(t \oplus s, \lambda), \quad \text{a.e. } \lambda ,$$

(iii) for each $t \in T$ and $\lambda \in [0,1)$,

$$W(t,\lambda)W(t,\mu) = W(t, \lambda \oplus \mu), \quad \text{a.e. } \mu .$$

(See Morettin (1974).)

Let $Y(t) = (Y_1(t),\ldots, Y_q(t))'$; $t \in T$ be a q-dimensional dyadic stationary process with zero mean and k-th order cumulants

$$c_{a_1\cdots a_k}(t_1,\ldots,t_{k-1}) = \text{cum}\{Y_{a_1}(t_1 \oplus t_k), Y_{a_2}(t_2 \oplus t_k),\ldots, Y_{a_k}(t_k)\} ,$$

$t_1,\ldots,t_{k-1} \in T$, $a_1,\ldots,a_k = 1,\ldots,q$. We denote the covariance matrices

$$\Gamma(t_1) = \{c_{a_1a_2}(t_1)\}, \quad q \times q \text{ matrices} .$$

The statistic

$$d^{(N)}(\lambda) = \sum_{t=0}^{N-1} Y(t)W(t,\lambda) \tag{2.1}$$

is called the finite Walsh transform of $\{Y(t): t = 0,1,\ldots,N-1\}$. Throughout this paper we assume that $N = 2^m$, with m a nonnegative integer and denote $d^{(N)}(\lambda) = (d_1^{(N)}(\lambda),\ldots, d_q^{(N)}(\lambda))'$. Here we assume the following.

ASSUMPTION 1. For every k and $j = 1,2,\ldots,k-1$,

$$\sum_{t_1=0}^{\infty} \cdots \sum_{t_{k-1}=0}^{\infty} |c_{a_1 \cdots a_k}(t_1,\ldots,t_{k-1})|\,|t_j| < \infty\,, \tag{2.2}$$

for all $a_1,\ldots,a_k$.

Then the Walsh spectral density matrix and the Walsh cumulant spectrum of order k of $\{Y(t)\}$ are defined by

$$f(\lambda) = \sum_{t=0}^{\infty} \Gamma(t) W(t,\lambda)\,,$$

and

$$\begin{aligned} f_{a_1\cdots a_k}(\lambda_1,\ldots,\lambda_{k-1}) \\ = \sum_{t_1} \cdots \sum_{t_{k-1}} c_{a_1\cdots a_k}(t_1,\ldots,t_{k-1}) \prod_{j=1}^{k-1} W(t_j,\lambda_j)\,, \end{aligned} \tag{2.3}$$

respectively. From Assumption 1, it is easy to see that $f(\lambda)$ and $f_{a_1\cdots a_k}(\lambda_1,\ldots,\lambda_{k-1})$ are integrable on $[0,1]$ and $[0,1]^{k-1}$, respectively. The following proposition is due to Morettin (1974).

PROPOSITION 2.1. *Under Assumption* 1,

$$\begin{aligned} &\mathrm{cum}\,\{d_{a_1}^{(N)}(\lambda_1),\ldots,d_{a_k}^{(N)}(\lambda_k)\} \\ &\quad = D_N(\lambda_1 \oplus \cdots \oplus \lambda_k)\{f_{a_1\cdots a_k}(\lambda_1,\ldots,\lambda_{k-1}) + O(N^{-1})\}\,, \end{aligned} \tag{2.4}$$

where $D_N(\lambda) = \sum_{t=0}^{N-1} W(t,\lambda)$, *and the term* $O(N^{-1})$ *is uniform with respect to* $\lambda_1,\ldots,\lambda_k$.

Although we do not assume the Gaussianity of $\{Y(t)\}$, we can compute the Gaussian likelihood function L of $\{Y(0),\ldots,Y(N-1)\}$, formally, and approximate L. That is, we get

$$\log L \cong -\frac{N}{2}\int_0^1 \{\log\det f_\theta(\lambda) + \mathrm{tr}\, I_N(\lambda) f_\theta(\lambda)^{-1}\}\,d\lambda + \text{constant}\,, \tag{2.5}$$

where the fitted Walsh spectral density matrix of $\{Y(t)\}$ is parameterized as $f_\theta(\lambda)$, $\theta = (\theta_1,\ldots,\theta_r)' \in \Theta \subset \mathbb{R}^r$, and

$$I_N(\lambda) = F_N(\lambda)F_N(\lambda)' = \{I_{ab}(\lambda)\}, \quad \text{say}\,,$$

$$F_N(\lambda) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} Y(t) W(t,\lambda)\,.$$

The fitted model $f_\theta(\lambda)$ may be different from the true one $f(\lambda)$, and we assume that $f_\theta(\lambda)$ and $f_\theta(\lambda)^{-1}$ are integrable on $[0, 1]$. Thus we estimate $\boldsymbol{\theta}$ by the value $\hat{\boldsymbol{\theta}}$ which minimizes

$$D(f_\theta, I_N) = \int_0^1 \{\log \det f_\theta(\lambda) + \operatorname{tr} I_N(\lambda) f_\theta(\lambda)^{-1}\} d\lambda , \tag{2.6}$$

with respect to $\boldsymbol{\theta}$. Henceforth we call $\hat{\boldsymbol{\theta}}$ the quasi-maximum likelihood estimator of $\boldsymbol{\theta}$. To discuss the asymptotic properties of $\hat{\boldsymbol{\theta}}$, the following lemma is a keystone.

LEMMA 2.1. *Let $\phi_j(\lambda) = \{\phi_{ab}^{(j)}(\lambda)\}$, $j = 1, \ldots, r$, be $q \times q$ matrix-valued continuous functions on* $[0, 1]$ *such that $\phi_j(\lambda) = \phi_j(\lambda)'$. Under Assumption* 1 *we can show that*

(2.7)

$$\lim_{N\to\infty} \int_0^1 \operatorname{tr} I_N(\lambda)\phi_j(\lambda) d\lambda = \int_0^1 \operatorname{tr} f(\lambda)\phi_j(\lambda) d\lambda, \tag{1}$$

in probability and the quantities

$$A_j = \sqrt{N} \int_0^1 \operatorname{tr} \{(I_N(\lambda) - f(\lambda))\phi_j(\lambda)\} d\lambda, \quad j = 1, \ldots, r , \tag{2}$$

have, asymptotically, a normal distribution with zero mean vector and covariance matrix V whose (j, m)-th element is

$$2\int_0^1 \operatorname{tr} \{f(\lambda)\phi_m(\lambda) f(\lambda)\phi_j(\lambda)\} d\lambda + \sum_{a,b,c,d=1}^{q} \iint_0^1 \phi_{ba}^{(j)}(\lambda)\phi_{dc}^{(m)}(\mu) f_{abcd}(\lambda, \lambda, \mu) d\lambda d\mu . \tag{2.8}$$

PROOF. Notice that

$$A_j = \sqrt{N} \sum_{a,b=1}^{q} \int_0^1 \{I_{ab}(\lambda) - f_{ab}(\lambda)\} \phi_{ba}^{(j)}(\lambda) d\lambda .$$

By Proposition 2.1, we have

$$E(I_{ab}(\lambda) - f_{ab}(\lambda)) = O\left(\frac{1}{N}\right),$$

where $O(1/N)$ is uniform with respect to λ. Hence we obtain

$$E(A_j) = O(N^{-1/2}) . \tag{2.9}$$

Since

$$\begin{aligned}\operatorname{cum}\{I_{ab}(\lambda), I_{cd}(\mu)\} &= \frac{1}{N^2}\{\operatorname{cum}(d_a(\lambda), d_d(\mu))\operatorname{cum}(d_b(\lambda), d_c(\mu)) \\ &\qquad + \operatorname{cum}(d_a(\lambda), d_c(\mu))\operatorname{cum}(d_b(\lambda), d_d(\mu)) \\ &\qquad + \operatorname{cum}(d_a(\lambda), d_b(\lambda), d_c(\mu), d_d(\mu))\} \\ &= \frac{1}{N^2}\Bigg\{ D_N^2(\lambda \oplus \mu)[f_{ad}(\lambda)f_{bc}(\lambda) + f_{ac}(\lambda)f_{bd}(\lambda)] \\ &\qquad + D_N(\lambda \oplus \lambda \oplus \mu \oplus \mu) f_{abcd}(\lambda, \lambda, \mu) + O\left(\frac{1}{N}\right)\Bigg\}\end{aligned}$$

and

$$D_N(\lambda \oplus \lambda \oplus \mu \oplus \mu) = D_N(0) = N,$$

we have

$$\begin{aligned}(2.10)\quad \operatorname{cum}(A_j, A_m) &= N\iint_0^1 \sum_{a,b,c,d=1}^{q} \phi_{ba}^{(j)}(\lambda)\phi_{dc}^{(m)}(\mu) \\ &\qquad \cdot \operatorname{cum}(I_{ab}(\lambda), I_{cd}(\mu))\, d\lambda_1 d\lambda_2 \\ &= \sum_{a,b,c,d=1}^{q}\Bigg[\iint_0^1 \phi_{ba}^{(j)}(\lambda)\phi_{dc}^{(m)}(\mu) f_{abcd}(\lambda, \lambda, \mu)\, d\lambda d\mu \\ &\qquad + 2\int_0^1 \phi_{ba}^{(j)}(\lambda) f_{ac}(\lambda) f_{bd}(\lambda)\, d\lambda \\ &\qquad \cdot \left[\frac{1}{N}\int_0^1 \phi_{dc}^{(m)}(\mu) D_N^2(\lambda \oplus \mu)\, d\mu\right]\Bigg] + O\left(\frac{1}{N}\right).\end{aligned}$$

Noting that

$$D_N(\lambda \oplus \lambda) = \begin{cases} N & \text{if } \ |\lambda \oplus \mu| < \dfrac{1}{N}, \\ 0 & \text{otherwise}, \end{cases}$$

we get that by the continuity of $\phi_{dc}^{(m)}(\mu)$,

$$\frac{1}{N}\int_0^1 \phi_{dc}^{(m)}(\mu) D_N^2(\lambda \oplus \mu)\, d\mu = \begin{cases} \phi_{dc}^{(m)}(\lambda) + o(1), & \text{if } \ \dfrac{1}{2N} < \lambda < 1 - \dfrac{1}{2N}, \\ O(1), & \text{otherwise}. \end{cases}$$

Substituting the above into (2.10), we get

$$\text{(2.11)}\quad \operatorname{cum}(A_j, A_m) = 2\int_0^1 \operatorname{tr}\{f(\lambda)\phi_m(\lambda)f(\lambda)\phi_j(\lambda)\}d\lambda$$
$$+\sum_{a,b,c,d=1}^{q} \iint_0^1 \phi_{ba}^{(j)}(\lambda)\phi_{dc}^{(m)}(\mu)f_{abcd}(\lambda,\lambda,\mu)d\lambda d\mu + o(1)\,.$$

Thus (2.9) and (2.11) imply our result (1) of (2.7). Also (2.11) gives the asymptotic variance (2.8). As for the asymptotic normality of A_j, we have only to evaluate $J(J \geq 3)$-th order cumulant, cum $\{A_{i_1}, A_{i_2}, \ldots, A_{i_J}\}$ and show that they are zero, $J \geq 3$. Here, without loss of generality, we evaluate it for scalar process.

By Theorem 2.3.2, p. 21 of Brillinger (1975), we have

$$\text{(2.12)}\quad \operatorname{cum}(d_{11}(\lambda_1)d_{12}(\lambda_1), \ldots, d_{J1}(\lambda_J)d_{J2}(\lambda_J))$$
$$= \sum_{\nu} \operatorname{cum}(d_{ji}(\lambda_j), (j,i) \in \nu_1) \cdots \operatorname{cum}(d_{ji}(\lambda_j), (j,i) \in \nu_S)\,,$$

where the summation runs over all indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_S$ of the set $\{(j,i), j = 1,2,\ldots,J, i = 1,2\}$ (the definition of indecomposability can be found on p. 20 of Brillinger (1975)). By indecomposability of the partitions, each ν_n contains at least two elements, so we have

$$S \leq J/2\,.$$

By Proposition 2.1, we have

$$\operatorname{cum}(d_{ji}(\lambda_j), (j,i) \in \nu_1) \cdots \operatorname{cum}(d_{ji}(\lambda_j), (j,i) \in \nu_S) = O\left(\prod_{n=1}^{S} D_N\left(\bigoplus_{(j,i)\in\nu_n} \lambda_j\right)\right).$$

Since

$$D_N(\lambda) = \begin{cases} N, & \text{if} \quad 0 \leq \lambda < \dfrac{1}{N}\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

we have, for $J \geq 2$

$$\int_0^1 \cdots \int_0^1 \prod_{n=1}^{S} D_N\left(\bigoplus_{(j,i)\in\nu_n} \lambda_j\right) d\lambda_1 \cdots d\lambda_J$$
$$= \int_0^1 \cdots \int_0^1 D_N(\mu_1 \oplus \mu_2)D_N(\mu_2 \oplus \mu_3) \cdots D_N(\mu_s \oplus \mu_1)d\mu_1 \cdots d\mu_s = O(N)\,,$$

and for $J=1$

$$\int_0^1 \cdots \int_0^1 D_N(\lambda_1 \oplus \cdots \oplus \lambda_J) d\lambda_1 \cdots d\lambda_J = O(1) .$$

Thus,

$$\int_0^1 \cdots \int_0^1 \operatorname{cum}\,(d_{11}(\lambda_1)d_{12}(\lambda_1),\ldots, d_{J1}(\lambda_J)d_{J2}(\lambda_J))d\lambda_1 \cdots d\lambda_J = O(N) ,$$

and consequently

$$\begin{aligned}\operatorname{cum}\,(A_{i_1},\ldots, A_{i_J}) &= N^{-J/2}\int_0^1 \cdots \int_0^1 \phi_{i_1}(\lambda_1)\cdots \phi_{i_J}(\lambda_J)\\ &\quad \cdot \operatorname{cum}\,(d_{11}(\lambda_1)d_{12}(\lambda_1),\ldots, d_{J1}(\lambda_J)d_{J2}(\lambda_J))d\lambda_1 \cdots d\lambda_J\\ &= O(N^{-J/2+1}) ,\end{aligned}$$

which implies the asymptotic normality. □

Suppose $f(\lambda)$ is the spectral density of a stationary process and $\{f_{\theta}(\lambda)\}$ is a family of fitted spectral densities which are parameterized by $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^r$, where Θ is a compact set in $\mathbb{R}^r$. We define a pseudo-true value $\bar{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^r$, by a value which minimizes

$$D(f_{\theta}, f) = \int_0^1 \{\log \det f_{\theta}(\lambda) + \operatorname{tr} f(\lambda) f_{\theta}(\lambda)^{-1}\} d\lambda ,$$

with respect to $\boldsymbol{\theta} \in \Theta$.

ASSUMPTION 2. The fitted model $f_{\theta}(\lambda)$ is twice continuously differentiable with respect to $\boldsymbol{\theta} \in \Theta$.

ASSUMPTION 3. If $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$, then $f_{\theta}(\lambda) \neq f_{\theta^*}(\lambda)$ on a set of positive Lebesgue measure. The matrix

$$M_f(\boldsymbol{\theta}) = \int_0^1 \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} [\log \det f_{\theta}(\lambda) + \operatorname{tr} f_{\theta}(\lambda)^{-1} f(\lambda)] d\lambda \tag{2.13}$$

is nonsingular for all $\boldsymbol{\theta} \in \Theta$, and $M_f = M_f(\bar{\boldsymbol{\theta}})$.

The first statement of Assumption 3 is an identifiability condition. In Section 4 some nonidentifiable examples will be given. Then we have the following theorem.

THEOREM 2.1. *Let $\{Y(t)\}$ be a q-dimensional dyadic stationary process with mean zero and the spectral density $f(\lambda)$. Suppose that Assumptions* 1–3 *are satisfied, and that $\bar{\boldsymbol{\theta}}$ exists uniquely and lies in* Int Θ. *Then*

(i) $\lim_{N\to\infty} \hat{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}$ *in probability,*

(ii) *the distribution of the vector $\sqrt{N}\{\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\}$, as $N \to \infty$, tends to the normal distribution with mean zero and covariance matrix $M_f^{-1} V M_f^{-1}$, where $V = \{V_{jm}\}$ is an $r \times r$ matrix such that*

$$V_{jm} = 2\int_0^1 \operatorname{tr}\left[f(\lambda) \frac{\partial}{\partial \theta_j} \{f_\theta(\lambda)\}^{-1} f(\lambda) \frac{\partial}{\partial \theta_m} \{f_\theta(\lambda)\}^{-1} \right]_{\theta=\bar{\theta}} d\lambda$$

$$+ \sum_{a,b,c,d=1}^{q} \iint_0^1 \left\{ \frac{\partial f_\theta^{(b,a)}(\lambda)}{\partial \theta_j} \cdot \frac{\partial f_\theta^{(d,c)}(\lambda)}{\partial \theta_m} \right\}_{\theta=\bar{\theta}} f_{abcd}(\lambda, \lambda, \mu)\, d\lambda d\mu\ ,$$

where $f_\theta^{(b,a)}(\lambda)$ is the (b,a)-th element of $f_\theta(\lambda)^{-1}$.

PROOF. From the definitions of $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} D(f_\theta, I_N)_{\theta=\hat{\theta}} = 0\ , \tag{2.14}$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} D(f_\theta, f)_{\theta=\bar{\theta}} = 0\ . \tag{2.15}$$

Expanding (2.14) around $\bar{\boldsymbol{\theta}}$, we have

$$0 = \frac{\partial}{\partial \boldsymbol{\theta}} D(f_{\bar{\theta}}, I_N) + \tilde{M}_f(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})\ , \tag{2.16}$$

where $\boldsymbol{\theta}^*$ lies on the straight section with end points $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$, and

$$\tilde{M}_f(\boldsymbol{\theta}^*) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} D(f_{\theta^*}, I_N)\ .$$

By Lemma 2.1, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} D(f_{\bar{\theta}}, I_N) \to 0\ ,$$

in probability and

$$\tilde{M}_f(\boldsymbol{\theta}) \to M_f(\boldsymbol{\theta})\ ,$$

in probability for each $\boldsymbol{\theta} \in \Theta$. By Assumptions 2 and 3, absolute values of eigenvalues of $M_f(\theta)$ have a positive lower bound for all $\boldsymbol{\theta} \in \Theta$. Hence when n is large enough, with a probability arbitrarily nearing one, so do the absolute values of eigenvalues of $\tilde{M}_f(\boldsymbol{\theta})$. By (2.16) we have

$$\hat{\boldsymbol{\theta}} \to \bar{\boldsymbol{\theta}},$$

in probability and consequently

$$\tilde{M}_f(\boldsymbol{\theta}^*) \to M_f(\bar{\boldsymbol{\theta}}),$$

in probability. Then the limiting distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})$ is equivalent to that of

$$(2.17) \qquad -M_f^{-1}\sqrt{N}\frac{\partial}{\partial\boldsymbol{\theta}} D(f_{\bar{\theta}}, I_N)$$

$$= -M_f^{-1}\sqrt{N}\int_0^1 \frac{\partial}{\partial\boldsymbol{\theta}}\{\log\det f_{\bar{\theta}}(\lambda) + \operatorname{tr} f_{\bar{\theta}}(\lambda)^{-1} I_N(\lambda)\}d\lambda$$

$$= -M_f^{-1}\sqrt{N}\int_0^1 \frac{\partial}{\partial\boldsymbol{\theta}}[\operatorname{tr} f_{\bar{\theta}}(\lambda)^{-1}\{I_N(\lambda) - f(\lambda)\}]d\lambda,$$

by (2.15). Again applying Lemma 2.1 to (2.17), we have completed the proof. □

Remark. If the true Walsh spectral density matrix $f(\lambda) = f_{\theta}(\lambda)$, the pseudo-true value is equal to the true value, i.e., $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}$ (see Hosoya and Taniguchi (1982)).

3. Model selection of Walsh spectral models

In this section we assume that the process $\{Y(t)\}$ has the true Walsh spectral density matrix $f(\lambda) = f_{\theta_r}(\lambda)$, $\boldsymbol{\theta}_r = (\theta_1, \ldots, \theta_r)'$, where $\boldsymbol{\theta}_r$ is an unknown parameter vector. (We use the suffix r to stress the dimension.) Since the order $\dim \boldsymbol{\theta} = r$ is unknown in many situations, we must estimate r from the data. Here we fit the Walsh spectral model $f_{\theta_k}(\lambda)$, $0 \le k \le L$, where L is a preassigned upper limit to the order. We determine the true order r by the value $\hat{k}$ which minimizes the following criterion:

$$(3.1) \qquad A(k) = D(f_{\hat{\theta}_k}, I_N) + \frac{kC_N}{N} \quad \text{for} \quad k = 0, 1, \ldots, L,$$

where $C_N \to \infty$ and $C_N/N \to 0$ as $N \to \infty$. For this estimated order $\hat{k}$ we have:

THEOREM 3.1. *Suppose that all the assumptions in Section 2 for* $f(\lambda) = f_{\theta_r}(\lambda)$ *and* $f_\theta(\lambda) = f_{\theta_k}(\lambda)$ *are satisfied. Then* $\lim_{N\to\infty} \hat{k} = r$, *in probability.*

PROOF. From (2.16) we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k) = -\tilde{M}_f^{-1}(\boldsymbol{\theta}_k^*)\sqrt{N}\int_0^1 \frac{\partial}{\partial \boldsymbol{\theta}_k}[\operatorname{tr} f_{\bar{\theta}_k}(\lambda)^{-1}\{I_N(\lambda) - f(\lambda)\}]d\lambda\,, \tag{3.2}$$

which tends to normal by Theorem 2.1. Thus we have, for any sequence of positive numbers $\tilde{C}_N \to \infty$,

$$P[\|\sqrt{N}(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)\| > \tilde{C}_N] = o(1)\,, \tag{3.3}$$

where $\|\cdot\|$ is the Euclidian norm. Taking $\tilde{C}_N = \sqrt[4]{C_N}$, we obtain

$$\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k = O_p(\sqrt[4]{C_N}/\sqrt{N})\,. \tag{3.4}$$

Expanding around $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_k$, and noting (3.4) we can see that

$$\begin{aligned} D(f_{\bar{\theta}_k}, I_N) = {} & D(f_{\hat{\theta}_k}, I_N) + (\bar{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_k)' \left.\frac{\partial D(f_{\theta_k}, I_N)}{\partial \boldsymbol{\theta}_k}\right|_{\theta_k = \hat{\theta}_k} \\ & + \frac{1}{2}(\bar{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_k)'\tilde{M}_f(\boldsymbol{\theta}_k^*)(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k)\,. \end{aligned} \tag{3.5}$$

Since $\partial D(f_{\theta_k}, I_N)/\partial \boldsymbol{\theta}_k|_{\theta_k = \hat{\theta}_k} = \mathbf{0}$, we have

$$D(f_{\hat{\theta}_k}, I_N) = D(f_{\bar{\theta}_k}, I_N) - \frac{1}{2}(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)'\tilde{M}_f(\boldsymbol{\theta}_k^*)(\hat{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_k)\,. \tag{3.6}$$

As a first step we show that

$$P(\hat{k} < r) \to 0 \quad \text{as} \quad N \to \infty\,. \tag{3.7}$$

For $k < r$, we evaluate

$$P_1 = P(A(k) < A(r)) = P\left\{ D(f_{\hat{\theta}_k}, I_N) - D(f_{\hat{\theta}_r}, I_N) < \frac{(r-k)C_N}{N} \right\}.$$

Using the relation (3.6), the above probability is approximated as

$$P\left\{ D(f_{\bar{\theta}_k}, I_N) - D(f_{\bar{\theta}_r}, I_N) < \frac{(r-k)C_N}{N} + \frac{1}{2}(\hat{\theta}_k - \bar{\theta}_k)'\tilde{M}_f(\theta_k^*)(\hat{\theta}_k - \bar{\theta}_k) - \frac{1}{2}(\hat{\theta}_r - \bar{\theta}_r)'M_f(\theta_r^*)(\hat{\theta}_r - \bar{\theta}_r) \right\}. \tag{3.8}$$

Using Lemma 2.1 the left-hand side of the above $\{\cdot\}$ converges to $D(f_{\bar{\theta}_k}, f) - D(f_{\bar{\theta}_r}, f)$, which is strictly positive for $k < r$. On the other hand, by (3.4), the right-hand side of $\{\cdot\}$ converges to zero in probability, which implies the probability $P_1 \to 0$ as $N \to \infty$. As a second step we show

$$P(\hat{k} > r) \to 0 \quad \text{as} \quad N \to \infty\,. \tag{3.9}$$

We have for $k > r$,

$$P_2 = P\{A(k) < A(r)\} = P\left\{ D(f_{\bar{\theta}_k}, I_N) - D(f_{\bar{\theta}_r}, I_N) < \frac{(r-k)C_N}{N} \right\}.$$

Using the relation (3.6), the above probability is approximated as

$$P\left\{ D(f_{\bar{\theta}_k}, I_N) - D(f_{\bar{\theta}_r}, I_N) - \frac{1}{2}(\hat{\theta}_k - \bar{\theta}_k)'\tilde{M}_f(\theta_k^*)(\hat{\theta}_k - \bar{\theta}_k) + \frac{1}{2}(\hat{\theta}_r - \bar{\theta}_r)'\tilde{M}_f(\theta_r^*)(\hat{\theta}_r - \bar{\theta}_r) < \frac{(r-k)C_N}{N} \right\}. \tag{3.10}$$

Because $f_{\bar{\theta}_k}(\lambda) = f_{\bar{\theta}_r}(\lambda)$, for $k \geq r$, we can see that

$$D(f_{\bar{\theta}_k}, I_N) - D(f_{\bar{\theta}_r}, I_N) = 0\,. \tag{3.11}$$

While, by (3.4), we can see that

$$-\frac{1}{2}(\hat{\theta}_k - \bar{\theta}_k)'\tilde{M}_f(\theta_k^*)(\hat{\theta}_k - \bar{\theta}_k) + \frac{1}{2}(\hat{\theta}_r - \bar{\theta}_r)'\tilde{M}_f(\theta_r^*)(\hat{\theta}_r - \bar{\theta}_r)$$

is at most of order $O_p(\sqrt{C_N}/N)$. However, the right-hand side of $\{\cdot\}$ in (3.10) is $(r-k)C_N/N$, $(r < k)$, which implies $P_2 \to 0$, as $N \to \infty$. Thus we have completed the proof. □

4. Determination of the order of dyadic autoregressive models

In the previous sections we could proceed in ways fairly analogous to those used in the ordinary stationary processes. However, if we consider finite parametric models, for example, dyadic autoregressive processes of finite order (DAR-processes), dyadic moving average processes of finite order (DMA-processes) and dyadic autoregressive moving average processes of finite order (DARMA-processes), then the greatest differences exist between dyadic stationary processes and ordinary stationary ones. That is, it is known that DAR, DMA and DARMA are equivalent, in the sense that a DAR or DARMA-process of finite order can be expressed as a DMA-process of finite order (see Nagai (1980) or Nagai and Taniguchi (1987)).

In this section, for a finite order dyadic autoregressive model, we can propose a simpler order determination criterion. Then we show that this criterion gives a strong consistent order estimate.

A q-dimensional dyadic stationary process $\{Y(t): t \in T\}$ is called a dyadic autoregressive process, if it can be expressed by

$$\sum_{j=0}^{p} A_j Y(t \oplus j) = \varepsilon(t), \qquad t \in T, \tag{4.1}$$

where

(i) A_j's are $q \times q$ matrices, $A_0 = I_q$ and $p = 2^r - 1$, where r is a non-negative integer,

(ii) $\varepsilon(t)$, $t \in T$ are i.i.d. random vectors such that

$$E\varepsilon(t) = 0, \qquad E\varepsilon(t)\varepsilon(t)' = G > 0, \tag{4.2}$$

(iii)

$$\det \Phi(\lambda) \neq 0, \qquad \text{a.e. } \lambda, \tag{4.3}$$

where $\Phi(\lambda) = \sum_{j=0}^{p} A_j W(j, \lambda)$.

If (4.2) and (4.3) hold true, then the Walsh spectral density of $\{Y(t)\}$ is

$$f(\lambda) = \Phi(\lambda)^{-1} G\{\Phi(\lambda)^{-1}\}'. \tag{4.4}$$

We call a DAR-process (4.1) irreducible if there does not exist a matrix

$$\Phi_1(\lambda) = \sum_{j=0}^{2^{r-1}-1} K_j W(j, \lambda),$$

which satisfies

$$f(\lambda) = \Phi_1(\lambda)^{-1} G\{\Phi_1(\lambda)^{-1}\}', \quad \text{a.e. } \lambda . \tag{4.5}$$

Moreover, for an irreducible DAR-process (4.1), there exists a t_0, $2^{r-1} \le t_0 \le 2^r - 1$, such that $A_{t_0} \ne 0$.

For an irreducible DAR model (4.1), p is called the order of the model. For simplicity, such a model is written as DAR(p). Note that in the above definition, the order of the model (4.1) is defined as $p = 2^r - 1$, not as $\max\{t: A_t \ne 0\}$. The advantage of such a definition is that it suits the Walsh spectral analysis, and is convenient for estimating the parameters of the model. To see this, consider the following two scalar irreducible DAR models:

$$X(t) + X(t \oplus 1) + \alpha X(t \oplus 2) = \varepsilon(t),$$

and

$$Y(t) + Y(t \oplus 1) + \alpha Y(t \oplus 3) = \varepsilon(t), \qquad t \in T,$$

where $\alpha \ne 0$, $\alpha \ne \pm 2$, $\varepsilon(t)$'s are i.i.d. with $E\varepsilon(t) = 0$, $E\varepsilon(t)^2 = \sigma^2$. It is easily seen that they have the same Walsh spectral density

$$\sigma^2[1 + W(1,\lambda) + \alpha W(2,\lambda)]^{-2} .$$

But if we define the order of the model as $\max\{t: A_t \ne 0\}$, then their order may be 2 and 3, respectively. Obviously such a definition is not convenient for Walsh spectral analysis. It is easy to see that these two models are not essentially different. For a $q \times q$ matrix $A = (a_{ij}, 1 \le i, j \le q)$, denote $\|A\| = \sum_{i,j=1}^{q} |a_{ij}|$. To determine the order $p = 2^r - 1$ of the irreducible model (4.1), we suggest the following criterion:

$$L_N(k) = \frac{1}{2^k} \sum_{n=0}^{2^k-1} \left\| \frac{1}{N} \sum_{t=0}^{N-1} Y(t) Y(t \oplus (2^k + n))' \right\|^2 - \frac{C_N}{N}, \tag{4.6}$$

where $Y(0), \dots, Y(N-1)$ are the observations of the model (4.1), $N = 2^m$ with m positive integer, and C_N satisfies the following conditions:

$$\lim_{N\to\infty} \frac{C_N}{N} = 0 \quad \text{and} \quad \lim_{N\to\infty} \frac{C_N}{\log\log N} = \infty . \tag{4.7}$$

Define

(4.8) $$\hat{r}_N = \max\{k \geq 0 : L_N(k-1) > 0, L_N(k) < 0\},$$

where $L_N(-1) = 1$ for convenience. We can use $\hat{r}_N$ as an estimate of the true value r of the model (4.1). We have the following:

THEOREM 4.1. *If the model* (4.1) *is irreducible and* (i), (ii) *and* (iii) *are satisfied, then*

(4.9) $$\lim_{N\to\infty} \hat{r}_N = r, \qquad a.s.$$

PROOF. Suppose that p is the true order of the model (4.1) and $p = 2^r - 1$. According to Nagai and Taniguchi (1987), if $\det\{\Phi(\lambda)\} \neq 0$, then $\{Y(t) : t \in T\}$ is a DMA-process written as

(4.10) $$Y(t) = \sum_{j=0}^{2^r-1} K_{j\oplus t}\varepsilon(j), \qquad t \in T.$$

Put $\Gamma(n) = EY(0)Y'(n)$. By (4.10) and condition (ii), it is easily seen that for any n,

(4.11) $$\left\| \frac{1}{N}\sum_{t=0}^{N-1} Y(t)Y'(t\oplus n) - \Gamma(n) \right\| = O\left(\sqrt{\frac{\log\log N}{N}}\right), \qquad \text{a.s.},$$

as $N \to \infty$ (e.g., Petrov (1975)). By (4.10), for $n \geq 2^r$, $\Gamma(n) = 0$. Thus, if $k \geq r$, then

(4.12) $$L_N(k) = O\left(\frac{\log\log N}{N}\right) - \frac{C_N}{N}, \qquad \text{a.s.},$$

as $N \to \infty$. From this and $\lim_{N\to\infty}(C_N/\log\log N) = \infty$, it follows that with probability one for large N,

(4.13) $$L_N(k) < 0, \qquad k \geq r.$$

If $r = 0$, the theorem is proved.

Now assume that $r > 0$. We have

(4.14) $$\lim_{N\to\infty} \sum_{n=0}^{2^{r-1}-1} \left\| \frac{1}{N}\sum_{t=0}^{N-1} Y(t)Y(t\oplus(2^{r-1}+n))' \right\|^2 = \sum_{n=0}^{2^{r-1}-1} \|\Gamma(2^{r-1}+n)\|^2, \qquad \text{a.s.}$$

We proceed to prove that

(4.15) $$\sum_{n=0}^{2^{r-1}-1} \|\Gamma(2^{r-1}+n)\|^2 > 0\,.$$

If we assume that (4.15) does not hold, we have

(4.16) $$f(\lambda) = \sum_{l=0}^{2^{r-1}-1} \Gamma(l)W(l,\lambda), \qquad \lambda \in [0,1]\,.$$

Put $h = 2^{r-1} - 1$, $\lambda_j = j/(h+1)$, $j = 0, 1, \dots, h$. We know that for all $l \le h$, $W(l,\lambda) = W(l,\lambda_j)$ for $\lambda \in [\lambda_j, \lambda_{j+1})$. From this it is easily seen that $f(\lambda)$ takes only at most $h+1$ different values, say, $f(\lambda_0), \dots, f(\lambda_h)$. By (4.4), $G > 0$ and $\Phi(\lambda) \neq 0$, it is easily seen that $f(\lambda_j) > 0$, $j = 0, 1, \dots, h$. Hence we can write

$$G = G^{1/2}(G^{1/2})', \qquad f(\lambda_j) = f^{1/2}(\lambda_j)f^{1/2}(\lambda_j)'\,,$$

$j = 0, 1, \dots, h$. Put

$$H_{h+1} = \begin{bmatrix} W(0,\lambda_0) & W(1,\lambda_0) & \dots & W(h,\lambda_0) \\ W(0,\lambda_1) & W(1,\lambda_1) & \dots & W(h,\lambda_1) \\ \vdots & \vdots & \vdots & \vdots \\ W(0,\lambda_h) & W(1,\lambda_h) & \dots & W(h,\lambda_h) \end{bmatrix}.$$

Then $H'_{h+1}H_{h+1} = (h+1)I_{h+1}$. Thus the matrix equation

(4.17) $$(H_{h+1} \otimes I_q)\begin{pmatrix} B_0 \\ \vdots \\ B_h \end{pmatrix} = \begin{pmatrix} G^{-1/2}f^{1/2}(\lambda_0) \\ \vdots \\ G^{-1/2}f^{1/2}(\lambda_h) \end{pmatrix}$$

has a unique solution $(B'_0, \dots, B'_h)$, where B_j's are all $q \times q$ matrices. From (4.17) we can see that

(4.18) $$\left(\sum_{l=0}^{h} B_l W(l,\lambda_j)\right) G \left(\sum_{l=0}^{h} B_l W(l,\lambda_j)\right)' = f(\lambda_j), \qquad j = 0, 1, \dots, h\,,$$

which implies

(4.19) $$\eta(\lambda)G\eta(\lambda)' = f(\lambda), \qquad \lambda \in [0,1]\,,$$

where

$$\eta(\lambda) = \sum_{l=0}^{h} B_l W(l, \lambda) . \tag{4.20}$$

By Nagai and Taniguchi (1987) there exists

$$\Phi_1(\lambda) = \sum_{l=0}^{2^{r-1}-1} K_l W(l, \lambda) ,$$

such that

$$\Phi_1(\lambda)\eta(\lambda) = I_q, \quad \text{a.e. } \lambda . \tag{4.21}$$

Thus we have

$$f(\lambda) = \Phi_1(\lambda)^{-1} G\{\Phi_1(\lambda)^{-1}\}', \quad \text{a.e. } \lambda , \tag{4.22}$$

which contradicts our irreducibility assumption. Now (4.15) has been proved. By (4.6), (4.7), (4.14) and (4.15), with probability one for large N,

$$L_N(r-1) > 0, \quad r > 0 . \tag{4.23}$$

Noting (4.13) and (4.23), with probability one for large N, we have

$$\hat{r}_N = r . \tag{4.24}$$

□

Remark. The following scalar process $\{Y(t): t \in T\}$ is a reducible DAR-process:

$$X(t) + X(t \oplus 1) + X(t \oplus 2) - X(t \oplus 3) = \varepsilon(t), \quad t \in T ,$$

where $\varepsilon(t)$'s are i.i.d. with $E\varepsilon(t) = 0$ and $E\varepsilon(t)^2 = \sigma^2$. Then

$$\Phi(\lambda) = 1 + W(1, \lambda) + W(2, \lambda) - W(3, \lambda) ,$$

but

$$\frac{\sigma^2}{\{\Phi(\lambda)\}^2} = \frac{\sigma^2}{4} .$$

5. Test of hypothesis for linear restriction of parameters

Let $\{Y(t)\}$ be a q-dimensional dyadic stationary process with Walsh spectral density $f_\theta(\lambda)$ depending on an unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$. We assume that $\{Y(t)\}$ satisfies all the assumptions in Theorem 2.1. The first problem is to test a composite hypothesis H_0: $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$, against H: $\boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{20}$, where $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$, $\boldsymbol{\theta}_1' = (\theta_1, \ldots, \theta_l)$, $\boldsymbol{\theta}_2' = (\theta_{l+1}, \ldots, \theta_p)$ and $\boldsymbol{\theta}_{20}' = (\theta_{l+1,0}, \ldots, \theta_{p,0})$, a specified vector and $(\boldsymbol{\theta}_1', \boldsymbol{\theta}_{20}') \in \operatorname{Int} \Theta$. Although we do not assume the Gaussianity of $\{Y(t)\}$, we can formally form the following log-likelihood ratio criterion

$$G = 2 \log L = N\{D(f_{(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)}, I_N) - D(f_{(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})}, I_N)\}\,, \tag{5.1}$$

where $\hat{\boldsymbol{\theta}}' = (\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2')$ is the quasi-maximum likelihood estimator for $\boldsymbol{\theta}$ under H, and $\tilde{\boldsymbol{\theta}}_1$ is that for $\boldsymbol{\theta}_1$ under H_0. Put $v = \sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, $w = \sqrt{N}(\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$ and $u' = (w', \mathbf{0}')$. Expanding in a Taylor expansion around $\hat{\boldsymbol{\theta}}$, we have

$$\begin{aligned} -G &= N\{D(f_{(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})}, I_N) - D(f_{(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)}, I_N)\} \\ &= \frac{1}{2}(u - v)' \frac{\partial^2 D(f_{\boldsymbol{\theta}}, I_N)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(u - v)(1 + o_p(1)) \\ &= \frac{1}{2}(u - v)' \frac{\partial^2 D(f_{\boldsymbol{\theta}}, I_N)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(u - v)(1 + o_p(1)) \\ &= \frac{1}{2}(u - v)' M_f (u - v)'(1 + o_p(1))\,. \end{aligned} \tag{5.2}$$

From Theorem 2.1 we have

$$v = -M_f^{-1}\sqrt{N}\frac{\partial}{\partial \boldsymbol{\theta}} D(f_{\boldsymbol{\theta}}, I_N)(1 + o_p(1))\,. \tag{5.3}$$

Similarly we have

$$u = -L_f\sqrt{N}\frac{\partial}{\partial \boldsymbol{\theta}} D(f_{\boldsymbol{\theta}}, I_N)(1 + o_p(1))\,, \tag{5.4}$$

where

$$L_f = \begin{bmatrix} I_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

and

$$E\frac{\partial^2}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} D(f_{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{20})}, I_N) = I_{11} + O(N^{-1})\,.$$

From (5.2), (5.3) and (5.4) we have

$$(5.5) \quad -G = \frac{1}{2}\sqrt{N}\,\frac{\partial D(f_\theta, I_N)}{\partial \theta'}[M_f^{-1} - L_f]M_f[M_f^{-1} - L_f] \cdot \sqrt{N}\,\frac{\partial D(f_\theta, I_N)}{\partial \theta}(1 + o_p(1))$$

$$= \frac{1}{2}\sqrt{N}\,\frac{\partial D(f_\theta, I_N)}{\partial \theta'}[M_f^{-1} - L_f]\sqrt{N}\,\frac{\partial D(f_\theta, I_N)}{\partial \theta}(1 + o_p(1)) .$$

Here we set down the following assumptions:

ASSUMPTION 4. The process $\{Y(t)\}$ is a linear dyadic stationary process represented as

$$(5.6) \qquad Y(t) = \sum_{j=0}^{\infty} A_j \boldsymbol{e}(t \oplus j) ,$$

where $\sum_{j=0}^{\infty} \|A_j\| < \infty$ and A_j are $q \times q$-matrices, and the $\boldsymbol{e}(t)$'s are independent random variables.

ASSUMPTION 5. The unknown parameter $\boldsymbol{\theta}$ of $f_\theta(\lambda)$ is innovation-free, i.e.,

$$(5.7) \qquad \frac{\partial}{\partial \boldsymbol{\theta}} \int_0^1 \operatorname{tr}\{f_\theta(\lambda)^{-1} f_{\theta_0}(\lambda)\} d\lambda = \mathbf{0}_p .$$

(See Hosoya and Taniguchi (1982).)

LEMMA 5.1. *Suppose that Assumptions* 1–5 *are satisfied, and that* $\sum_{j=0}^{\infty} A_j W(j, \lambda) \neq 0$ *for all* $\lambda \in [0, 1]$. *For an innovation-free parameter* $\boldsymbol{\theta}$ *we have*

$$\sqrt{\frac{N}{2}}\,\frac{\partial D(f_\theta, I_N)}{\partial \boldsymbol{\theta}} \xrightarrow{\mathscr{L}} N(\mathbf{0}_p, M_f) .$$

PROOF. Using an argument similar to Hosoya and Taniguchi (1982), we can see that

$$\iint_0^1 \frac{\partial}{\partial \theta_j} f_\theta^{(b,a)}(\lambda) \cdot \frac{\partial}{\partial \theta_m} f_\theta^{(d,c)}(\lambda) \cdot f_{abcd}(\lambda, \lambda, \mu) d\lambda d\mu = 0 ,$$

for $j, m = 1, \ldots, p$; $a, b, c, d = 1, \ldots, q$. Putting $\phi_j(\lambda) = (\partial/\partial\theta_j) f_\theta(\lambda)$ in Lemma 2.1, we have the desired result. □

Applying Lemma 5.1 to (5.5) we have:

THEOREM 5.1. *Suppose that Assumptions* 1–5 *are satisfied. Then the distribution of* $-G$ *under* H_0 *tends to* $\chi^2(p-l)$ *as* $N \to \infty$.

Now we consider a more general test of the hypothesis.

$$H_0\colon B\boldsymbol{\theta} = u_{20} \qquad \text{against} \qquad H\colon B\boldsymbol{\theta} \neq u_{20}\,,$$

where B is a $(p-l)\times p$ matrix with rank $B = p - l$, and $u_{20}' = (u_{l+1,0},\dots,u_{p,0})$. Then there exists an $l\times p$ matrix A such that

$$\begin{pmatrix} A \\ B \end{pmatrix}\boldsymbol{\theta} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u(\boldsymbol{\theta})\,,$$

where $\det\begin{pmatrix} A \\ B \end{pmatrix} \neq 0$. Let $\hat{\boldsymbol{\theta}}$ be the quasi-maximum likelihood estimator of $\boldsymbol{\theta} \in \Theta$, then $u(\hat{\boldsymbol{\theta}}) = \hat{u}$. Then the likelihood ratio criterion of testing

$$H_0\colon u_2 = u_{20} \qquad \text{against} \qquad H\colon u_2 \neq u_{20}$$

is given by

$$\tilde{G} = N\{D(f_{(\hat{u}_1,\hat{u}_2)}, I_N) - D(f_{(\tilde{u}_1,u_{20})}, I_N)\}\,, \tag{5.9}$$

where $\tilde{u}_1$ is the quasi-maximum likelihood estimator of u_1 under H_0. Then we have:

THEOREM 5.2. *Suppose that Assumptions* 1–5 *are satisfied. Then the distributions of* $-\tilde{G}$ *under* H_0 *tends to* $\chi^2(p-l)$ *as* $N \to \infty$.

Acknowledgements

The authors sincerely thank two referees for providing many insightful comments.

REFERENCES

Beauchamp, K. G. (1975). *Walsh Functions and Their Applications*, Academic Press, New York.

Brillinger, D. R. (1975). *Time Series; Data Analysis and Theory*, Holt, Rinehart & Winston.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Statist.*, **10**, 132–153.

Kohn, R. (1980). On the spectral decomposition of stationary time series using Walsh functions, I, *Adv. in Appl. Probab.*, **12**, 183–199.

Morettin, P. A. (1974). Walsh function analysis of a certain class of time series, *Stochastic Process. Appl.*, **2**, 183–193.

Morettin, P. A. (1981). Walsh spectral analysis, *SIAM Rev.*, **23**, 279–291.

Nagai, T. (1977). Dyadic stationary processes and their spectral representations, *Bull. Math. Statist.*, **17**, 65–73.

Nagai, T. (1980). On finite parametric linear models of dyadic stationary processes, *Bull. Math. Statist.*, **19**, 45–53.

Nagai, T. and Taniguchi, M. (1987). Walsh spectral analysis of multiple dyadic stationary processes and its applications, *Stochastic Process. Appl.*, **24**, 19–30.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer, New York-Berlin.