Nonparametric Log–Spectrum Estimation using Disconnected Regression Splines and Genetic Algorithms

Thomas C. M. Lee* and Tan F. Wong†

April 4, 2001; Revised: February 5, 2002

Abstract

This article proposes a new nonparametric procedure for estimating log spectra. This procedure consists of three major components: 1) a novel statistical model for modeling the unknown target log spectrum, 2) an AIC-based model selection criterion for choosing a "best" fitting model, and 3) a genetic algorithm for effectively searching the "best" fitting model. Numerical experiments are conducted to evaluate and compare the practical performance of the proposed procedure with some other common log spectral estimation procedures appearing in the literature. These other procedures include wavelet techniques, kernel smoothing and regression spline fitting. Empirical results suggest that the proposed procedure compares favourably against all these procedures, especially when the unknown log spectrum contains inhomogeneous structures.

*Corresponding Author. Postal: Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA. Phone: (970) 491-2185. Fax: (970) 491-7895. Email: `tlee@stat.colostate.edu`.

†Postal: Department of Electrical and Computer Engineering, University of Florida, 461 Engineering Building, PO Box 116130, Gainesville, FL 32611, USA. Phone: (352) 392-2665. Fax: (352) 392-0044. Email: `twong@ece.ufl.edu`.

List of Tables

List of Figures

Keywords: genetic algorithms, log–periodogram smoothing; disconnected regression splines; spectral density estimation

1 Introduction

This article studies the problem of nonparametric log spectral density estimation. Various log spectrum estimation procedures that adopt the idea of smoothing the log periodogram have been proposed. These include kernel smoothing (e.g., [6], [14], [17] and [20]), smoothing spline methods (e.g., [19] and [23]) and wavelet techniques (e.g., [9], [18] and [24]). The new procedure that this article proposes uses a different statistical model to model the target log spectrum: the target log spectrum is modeled by a series of disconnected regression splines that partition the domain of the spectrum (briefly, regression splines are a special kind of polynomial functions and a brief introduction is given in Section 2.2). As will be demonstrated below, such a model is extremely suitable for spectra with inhomogeneous structures.

It will be shown below that the problem of estimating a log spectrum using this disconnected regression spline approach can be posed as a statistical model selection problem, in which different candidate models may have different dimensions. In order to tackle this model selection problem, we employ a modified form of the Akaike's information criterion (AIC) [1] to construct an objective function for which the best fitting model for our problem is defined as its optimizer. However, optimizing this objective function would involve solving a hard and large scale optimization problem. In this work we propose using genetic algorithms for solving such problems (e.g., see [5], [7] and references given therein). Simulation results suggest that the use of genetic algorithms is very effective.

The rest of this article is organized as follows. Background material on the properties of log periodograms and regression splines is given in Section 2. In Section 3 we present our new disconnected regression spline model for log spectra. Section 4 describes the above mentioned AIC model selection method while Section 5 proposes a genetic algorithm for solving the related optimization problem. Section 6 reports simulation results. Finally, conclusions are given in Section 7.

2 Background

2.1 Log Periodogram

Suppose that $\{x_t\}$ is a real-valued, zero-mean strictly stationary process with unknown spectral density S, and that a finite-sized realization $x_0, \ldots, x_{2n-1}$ of $\{x_t\}$ is observed. The periodogram is defined as

$$I(\omega) = \frac{1}{2\pi \times 2n} \left| \sum_{t=0}^{2n-1} x_t \exp(-i\omega t) \right|^2, \qquad \omega \in [0, 2\pi).$$

To simplify notation, write $\omega_l = 2\pi l/(2n)$. Since the spectral density S is symmetric about $\omega = \pi$, we shall focus our discussion on $S(\omega_l)$ for $l = 0, \ldots, n-1$.

Let $\gamma_r = E(x_{t-r}x_t)$, $r = 0, 1, \ldots$, be the autocovariance function. If all moments of x_t exist, the sum of all $|\gamma_r|$'s is bounded and n is large (e.g., see [2, Theorem 5.2.6]), then

$$I(\omega_l) \approx S(\omega_l)\epsilon_l, \qquad j = 0, \ldots, n-1, \tag{1}$$

where ϵ_0, and $\epsilon_{n/2}$ if n is even, are independent χ_1^2 random variables, and all other ϵ_j's are independent random variables distributed as the standard exponential distribution. From now on, as in [6], [9], [14], [18], [19], [20], [22], [23] and [24], ϵ_0 and $\epsilon_{n/2}$ will be treated as if they were standard exponential random variables and the approximation in (1) is assumed to be exact. The effect of these changes is asymptotically negligible. Hence we have the following model:

$$I(\omega_l) = S(\omega_l)\epsilon_l, \qquad j = 0, \ldots, n-1,$$

where the ϵ_j's are independent standard exponential random variables. Thus $E\{I(\omega_l)\} = S(\omega_l)$ and $\mathrm{var}\{I(\omega_l)\} = S(\omega_l)^2$.

Due to its unacceptably large variance, $I(\omega_l)$ is seldom used as an estimate of $S(\omega_l)$. To stabilize these high variance $I(\omega_l)$'s, it is common to apply a logarithmic transform to convert the above multiplicative noise model into an additive noise model (e.g., see [6], [9], [14], [18], [19], [23] and [24]):

$$y(\omega_l) = \log I(\omega_l) + \gamma = \log S(\omega_l) + \xi_l = f(\omega_l) + \xi_l, \qquad l = 0, \ldots, n-1. \tag{2}$$

In the above $\gamma = 0.57721$ is the Euler's constant and the ξ_l's are independent random variables with probability density function $p_\xi(x) = \exp\{x - \gamma - \exp(x - \gamma)\}$. It can be shown that $E(\xi_l) = 0$ and $\mathrm{var}(\xi_l) = \pi^2/6$. Also, in the above $f = \log S$ is the log spectrum.

With the above set up, the problem that this article considers can be stated as: given the "data" $\boldsymbol{y} = (y(\omega_0), \ldots, y(\omega_{n-1}))^T$ that are obtained from the periodogram and the additive noise model (2), we would like to estimate f in a nonparametric fashion. It is anticipated that f may possess some abrupt changing structures but is otherwise smooth.

2.2 Regression Splines

In the statistics literature an increasingly popular nonparametric curve fitting method is regression spline smoothing; e.g., see [8], [10], [13], [11] and [16]. Regression splines are a special kind of piecewise polynomials. Those breakpoints that divide a regression spline into pieces is known as *knots*. Usually these divided pieces are low–order polynomials having the same order, and they are forced to join smoothly at the knots. Common choices for the order of the polynomials are linear, quadratic and cubic. In this article we shall use piecewise quadratic polynomials constrained to have continuous first derivative at the knots. This idea is illustrated in Fig. 1. If $k_1 < \ldots < k_m$ is a sequence of m knots, then mathematically a quadratic regression spline f_{Q} can be expressed as

$$f_{\mathrm{Q}}(\omega) = \alpha_0 + \alpha_1\omega + \alpha_2\omega^2 + \sum_{r=1}^{m} \beta_r(\omega - k_r)_+^2,$$

where α_i's and β_r's are parameters and $(\omega)_+$ denotes $\max(0, \omega)$. It is straightforward to verify that f_{Q} is a quadratic polynomial in any interval $[k_r, k_{r+1})$ and that f_{Q} possesses continuous first derivative.

It is widely known that the crucial aspect of regression spline fitting is the choice of the number and the placement of the knots: inadequate number of knots or badly placed knots would lead to oversmoothing in some regions of the underlying true function, while too many knots would inflate local variance. Once the number and locations of the knots are determined, *unique* maximum likelihood estimates of α_i's and β_r's can be obtained (or approximated) say by the Newton–Raphson algorithm (or least–squares regression); e.g., see [13] and [16].

As a remark we mention that regression splines are *not* the same as smoothing splines nor interpolating splines: a smoothing spline is defined as the $\hat{f}$ that minimizes $\sum\{y(\omega_l) - \hat{f}(\omega_l)\}^2 + \lambda \int\{\hat{f}''(\omega)\}^2 d\omega$ (λ is a pre–determined smoothing parameter) while an interpolating spline passes through all the data points (i.e., no smoothing is done).

3 log spectrum Model: Disconnected Regression Splines

This section presents our model for the log spectrum f. One characteristic of our model is that it is capable of handling f with inhomogeneous structures. The idea is to approximate f by a series of disconnected quadratic regression splines. In this way boundary points between any two adjacent quadratic regression splines can serve as locations of sudden changes in f; see Fig. 2 for an illustration. In the sequel we shall call these boundary points *break points*. Despite that regression splines have been used in the spectral density estimation context (e.g., [13]), to the best of our knowledge, the use of a series of regression splines to handle inhomogeneous structures in a log spectrum is novel.

Suppose it is known that there are $B-1$ break points in f and that these break points are located at $b_1, b_2, \ldots, b_{B-1}$. For convenience let $b_0 = \omega_0$ and $b_B = \omega_{n-1}$, and assume $b_0 < b_1 < \ldots < b_B$. Then one way to estimate f is to fit a separate quadratic regression spline to each of the B disjoint segments $[b_{j-1}, b_j)$, $j = 1, \ldots, B$. If I_E is the indicator function for the event E, then such a disconnected regression spline model for f can be expressed as

$$f(\omega) = f_1(\omega) I_{\{b_0 \leq \omega < b_1\}} + f_2(\omega) I_{\{b_1 \leq \omega < b_2\}} + \ldots + f_B(\omega) I_{\{b_{B-1} \leq \omega \leq b_B\}}, \qquad (3)$$

where each of the f_j's is a quadratic regression spline having m_j knots located at $k_{j1}, k_{j2}, \ldots, k_{jm_j}$. Furthermore,

$$f_j(\omega) = \alpha_{j0} + \alpha_{j1}\omega + \alpha_{j2}\omega^2 + \sum_{r=1}^{m_j} \beta_{jr}(\omega - k_{jr})_+^2, \qquad b_{j-1} \leq \omega < b_j, \qquad j = 1, \ldots, B, \qquad (4)$$

where $\boldsymbol{\alpha}_j = \{\alpha_{j0}, \alpha_{j1}, \alpha_{j2}\}$ and $\boldsymbol{\beta}_j = \{\beta_{j1}, \ldots, \beta_{jm_j}\}$ are model parameters. To simplify notation, let $\boldsymbol{b} = \{b_1, \ldots, b_{B-1}\}$, $\boldsymbol{m} = \{m_1, \ldots, m_B\}$ and $\boldsymbol{k}_j = \{k_{j1}, \ldots, k_{jm_j}\}$ for $j =$

$1, \ldots, B$. Although not necessary but for computational convenience, it is assumed that $\{b_j, k_{jr}; \text{for all } j, r\}$ is a subset of $\{\omega_0, \ldots, \omega_{n-1}\}$ and that

$$b_0 < k_{11} < \ldots < k_{1m_1} < b_1 < \ldots < b_{j-1} < k_{j1} < \ldots < k_{jm_j} < b_j < \ldots < b_B.$$

Of course, in most situations the number and locations of the break points and knots are not known and need to be estimated.

4 Model Selection and Parameter Estimation

If f is modeled by the above disconnected regression splines (3) and (4), then an estimate $\hat{f}$ of f can be obtained by first estimating $\boldsymbol{\theta} = \left\{B, \boldsymbol{b}, \boldsymbol{m}, \{\boldsymbol{k}_j, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j\}_{j=1}^{B}\right\}$ and then plugging the resulting estimate $\hat{\boldsymbol{\theta}} = \left\{\hat{B}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{m}}, \{\hat{\boldsymbol{k}}_j, \hat{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\beta}}_j\}_{j=1}^{\hat{B}}\right\}$ into (3) and (4). Hence using the disconnected regression spline approach, our original log spectrum estimation problem can be posed as a model selection problem, with each candidate model specified by a $\hat{\boldsymbol{\theta}}$. The goal, then, is to select a "best" $\hat{\boldsymbol{\theta}}$. Notice that different $\hat{\boldsymbol{\theta}}$'s may have different dimensions. Also notice that once $\hat{B}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{m}}$ and $\{\hat{\boldsymbol{k}}_j\}_{j=1}^{\hat{B}}$ are specified, natural maximum likelihood estimates of $\{\hat{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\beta}}_j\}_{j=1}^{\hat{B}}$ will be uniquely determined, and can be iteratively computed say by the Newton-Raphson algorithm.

Now the question is: how to select a "best" fitting model $\hat{\boldsymbol{\theta}}$? We shall consider the use of a generalization of the Akaike's information criterion (AIC) [1] (see also [3] for a comprehensive reference). With AIC the best fitting model is chosen as the one that minimizes an estimator of the Kullback–Leibler (KL) distance measure (e.g., see [3]) between a fitted model and the "true" model. If p is the number of parameters that need to be estimated in a fitted model, then under some mild regularity conditions one can show that such a KL distance estimator is $-2 \times$ "maximized log likelihood" $+ 2p$. For our log spectral density estimation problem this distance estimator amounts to

$$\mathrm{AIC}(\hat{f}) = \left[2 \sum_{l=0}^{n-1} \left[\exp\{y(\omega_l) - \hat{f}(\omega_l) - \gamma\} - \{y(\omega_l) - \hat{f}(\omega_l)\} - \gamma \right] + \lambda p \right]_{\lambda=2} . \qquad (5)$$

However, empirical results for similar but different model selection problems often suggest that $\lambda = 2$ is *not* a good choice. Instead, it is often suggested to replace $\lambda = 2$ by $\lambda = \log n$; e.g., see [11], [12] and [13]. Such a change of the value of λ is in fact making AIC the same as the Bayesian Information Criterion (BIC) [21].

We shall follow this suggestion and select the "best" model $\hat{f}$ as the one that minimizes $\mathrm{AIC}(\hat{f})$ with $\lambda = \log n$. Our empirical experience also supports this choice. Note that for our regression model $p = 3\hat{B} + \sum_{j=1}^{\hat{B}} \hat{m}_j$. Also, we use a Newton–Raphson algorithm to obtain maximum likelihood estimates of the parameters $\{\hat{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\beta}}_j\}_{j=1}^{\hat{B}}$.

5 Optimization by Genetic Algorithms

When the number of data points is large, finding the best estimate defined by the above AIC/BIC criterion would involve solving a hard, large scale minimization problem. Common techniques for dealing with these types of problems include knot addition, knot deletion, knot movement or combinations of them; e.g., [8], [13] and [16]. However, these techniques do not provision the inclusion of break points in our model. In this article we suggest using genetic algorithms, which are also known as evolutionary algorithms [7], for solving our needs. It has been demonstrated that, when correctly applied, genetic algorithms can tackle very hard and large scale optimization problems; e.g, see [5].

5.1 General Description

The use of genetic algorithms for solving optimization problems can be briefly described as follows. An initial set, or population, of possible solutions to an optimization problem is obtained and represented in vector form. These vectors are often called *chromosomes* and are free to "evolve" in the following way. Firstly parent chromosomes are randomly chosen from the initial population: chromosomes having lower (or higher) values of the objective criterion to be minimized (or maximized), respectively, would have a higher chance of being chosen. Offspring are then reproduced from either applying a *crossover* or a *mutation* operation to these chosen parents. Once a sufficient number of such second generation offspring are produced, third generation offspring are further produced from these second generation offspring in a similar manner. For the reason of preserving the best chromosome of a current generation, an additional step that one may perform is the *elitist* step: replace the worst chromosome of the next generation with the best chromosome of the current generation. This reproduction process continues for a number of generations. The expectation is that the objective criterion values of the offspring should gradually improve over generations; i.e., approaching the optimal value.

5.2 Chromosome Representation

The performance of a genetic algorithm would certainly depend on how a possible solution is represented as a chromosome. In traditional applications of genetic algorithms, solutions are often represented as binary vectors; that is, chromosomes with two types of genes. However it is more convenient to employ a "three–gene–type" representation for our problem.

First we point out that for our problem a possible solution $\hat{\boldsymbol{\theta}}$ can be uniquely specified by $\hat{B}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{m}}$ and $\{\hat{\boldsymbol{k}}_j\}_{j=1}^{\hat{B}}$. It is because once they are specified, the corresponding maximum likelihood estimates for $\{\hat{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\beta}}_j\}_{j=1}^{\hat{B}}$ can be uniquely calculated. Thus for the current problem a chromosome only needs to carry information about $\hat{B}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{m}}$ and $\{\hat{\boldsymbol{k}}_j\}_{j=1}^{\hat{B}}$. A simple example will be used to illustrate our representation scheme. Suppose $n = 20$, $\hat{B} = 2$, $\hat{\boldsymbol{b}} = \{12\}$, $\hat{\boldsymbol{m}} = \{2, 1\}$, $\hat{\boldsymbol{k}} = \{\hat{\boldsymbol{k}}_1, \hat{\boldsymbol{k}}_2\}$, $\hat{\boldsymbol{k}}_1 = \{6, 10\}$ and $\hat{\boldsymbol{k}}_2 = \{17\}$. That is, the log spectrum estimate is composed of two disconnected splines separated at ω_{12}, and there are two and one knots in the first and the second spline respectively. These knots are located at ω_6, ω_{10} and ω_{17}. If we use "$\wr$" to denote a break point gene, "$\diamond$"to denote a knot gene and "$\cdot$" to denote a normal gene, then the chromosome for this example is composed of $n = 20$ genes arranged as:

$$\cdot\cdot\cdot\cdot\cdot\diamond\cdot\cdot\cdot\diamond\cdot\wr\cdot\cdot\cdot\cdot\diamond\cdot\cdot\cdot \quad .$$

Empirical evidence suggests that the above representation scheme is extremely effective for the purpose of using genetic algorithms to minimize our AIC/BIC criterion. It is most likely due to the fact that the location information of the break points and the knots of a $\hat{\boldsymbol{\theta}}$ are explicitly represented.

5.3 Implementation

This section lists the major steps behind the implementation of the genetic algorithm that we use to minimize the modified AIC$(\hat{f})$. The steps are:

1. Randomly generate an initial population of S chromosomes.

2. Compute the AIC$(\hat{f})$ value for each of the S chromosomes.

3. Sort the AIC$(\hat{f})$ values in descending order. Assign rank 1 to the chromosome with

the largest AIC($\hat{f}$) value, rank 2 to the one with the second largest AIC($\hat{f}$) value, and so on. Denote the rank of chromosome i as r_i.

4. With probability P_{CROSS}, perform a crossover operation. Otherwise, perform a mutation operation.

   - *Crossover* — one child is reproduced from two parents: select two parent chromosomes from the initial population. The probability that chromosome i will be selected is $r_i/(r_1+\ldots+r_S)$. Then for each gene location of the child chromosome, with equal probability, assign it with either the corresponding father gene or the corresponding mother gene.

   - *Mutation* — one child is reproduced from one parent: select one parent using the same probability law as in crossover. Then for each child gene location, assign it with either a knot gene with probability P_{K}, a break point gene with probability P_{S}, a normal gene with probability P_{N}, or the corresponding father gene with probability $1 - P_{\text{K}} - P_{\text{S}} - P_{\text{N}}$.

5. Repeat Step 4 until S child chromosomes are reproduced. That is, until a whole new generation is obtained.

6. Apply the elitist step.

7. Repeat Steps 2 to 6 using the new generation as the initial population. Then repeat the whole process until the AIC($\hat{f}$) value of the best chromosome does not change for N_{SAME} generations.

8. The best chromosome of the youngest generation is taken as the minimizer of AIC($\hat{f}$).

In our implementation $S = 300$, $P_{\text{CROSS}} = 0.9$, $P_{\text{K}} = 0.1$, $P_{\text{S}} = 0.02$, $P_{\text{N}} = 0.1$ and $N_{\text{SAME}} = 20$.

6 Numerical Experiments

This section reports results of those numerical experiments that were conducted for evaluating and comparing the practical performance of the proposed log spectrum estimation procedure with some other procedures appearing in the literature.

6.1 Settings

Four testing spectra and three different sample sizes were used. The three sample sizes were $2n = 256$, 512 and 1024, and the log of the four spectra are displayed in Fig. 3. These four testing spectra have different characteristics and have been used by other authors (e.g., [9], [18] and [24]). Spectrum 1 contains sharp changes (according to [18], it is an approximation to the spectrum of a typical communication signal). Spectrum 2 corresponds to an AR(24) process and it possesses some broad as well as some fine structures that may be hard to estimate simultaneously. Spectrum 3 is from an AR(2) process. It has a peak at $\omega \approx 0.75$ which can be easily oversmoothed. Finally, Spectrum 4 is a MA(15001) process with a sharp peak at $\omega \approx \pi/4$. Detecting this peak is a challenging problem. For complete specifications of these four testing spectra, see for example [24].

For each combination of testing spectrum and sample size (totally there are $4 \times 3 = 12$ such combinations), 100 independent series were simulated. For each of these 100 series, four different log spectrum estimates were obtained from the following estimation procedures:

1. drs: the proposed disconnected regression spline procedure.

2. mwt: the multitaper scale–independent wavelet soft thresholding procedure proposed by [24]. It is demonstrated in [24] that this multitaper wavelet thresholding procedure is superior to the wavelet–based procedures proposed by [9] and [18]. These wavelet–based procedures are shown to be very effective in terms of recovering highly inhomogeneous log spectra.

3. lsp: the "logspline" procedure proposed by [13]. This procedure models the log spectrum by *one single* regression spline, and hence it was not designed to handle sharp changes. This procedure uses a knot deletion–addition algorithm (i.e., not a genetic algorithm) for searching the number and locations of the knots of the best fitting regression spline.

4. wla: the constant–span weighted local averaging procedure of [14]. This procedure is a member of the class of smoothed periodogram estimators described for example by [2, Sections 5.4 to 5.6]. It aims to choose the smoothing span that minimizes the mean squared error (MSE, defined below) between the true and the estimated log spectra. This simple procedure should perform well when the log spectrum does not possess any abrupt changes.

As one can see that, the above four log spectrum estimation procedures are very different in nature and a direct comparison of them should be interesting.

6.2 Ranking the Procedures

In this subsection we describe the approach we adopted to rank the performances of the procedures.

The numerical quantity that was used to measure the quality of a log spectrum estimate $\hat{f}$ is the mean squared error: $\mathrm{MSE}(\hat{f}) = \frac{1}{n}\sum_{l=0}^{n-1}\{f(\omega_l) - \hat{f}(\omega_l)\}^2$. For each combination of test spectrum and sample size, we computed the MSE values for all $\hat{f}$'s that we obtained, and the results are summarized in Table 7 in terms of means and standard deviations. We have also performed Fisher's Least Significant Difference Test (e.g., see [4]) to test if the difference between the mean MSE values of two procedures is significant or not. The significance level used was 5%. Based on the results of this test, we ranked the four procedures in the following manner. If the mean MSE value of a procedure is significantly less than the mean MSE values of the remaining three procedures, then this procedure will be assigned a rank of 1. If the mean MSE value of a procedure is significantly less than two but greater than one of the remaining three procedures, then this procedure will be assigned a rank of 2. Similarly for ranks 3 and 4 (i.e., 1 is the best while 4 is the worst). On the other hand, if the mean MSE values of two procedures are not significantly different, then these two procedures will share the same averaged rank. For example, if the mean MSE values of two procedures are not significantly different but are significantly less than the MSE values of the remaining two procedures, then these two procedures will be assigned the same averaged rank of $\frac{1+2}{2} = 1.5$. All resulting rankings are also provided in Table 7. This ranking scheme has been adopted by previous articles (e.g., see [15] and [25]).

We chose MSE as the measure of performance because it is commonly accepted as a reasonable measure of the difference between the true and an estimated log spectra. The MSE does have, to a reasonable extent, the capability of revealing differences in fine structures and peaks. For instance, consider Spectrum 2. The MSE–based ranking presented in Table 7 does in fact agree with the conclusion obtained from visually inspecting Figure 4 (see below for the description of Figure 4) that wla performs better than the other 3 techniques. The above ranking system, which is based on the Fisher's least significant difference test and the MSE measure, provides a reasonable quantitative means for ranking the four procedures.

6.3 Results

The relative performances of the four procedures, of course, depend on the structure of the true spectrum. For Spectrum 1, drs gave the best results, due to the fact that the proposed disconnected regression spline model is designed to handle sharp changes. For mwt, it suffered from Gibb's effects; for lsp, it oversmoothed as it approximated the log spectrum with one spline; and for wla, due to the right–most sharp rise, undersmoothed the target. For Spectrum 2, all procedures, with the exception of wla, suffered from oversmoothing. For Spectrum 3, the two spline–based procedures, drs and lsp, gave similar and better results. They could recover the peak reasonably well, while both mwt and wla oversmoothed it, with mwt to a lesser extent. Lastly, for Spectrum 4, both mwt and wla often missed the sharp change and hence gave poor MSE readings. While not shown in Figure 5, quite often drs did capture this sharp change and hence gave overall the best results.

From the above observations, no procedure can be claimed as universally the best nor the worst. However, the Fisher's Least Significant Difference Test rankings do suggest that drs is, on average, superior to others: the overall average test rankings (see below) for the procedures drs, mwt, lsp and wla are 1.54, 3.13, 2.58 and 2.75 respectively. The overall average ranking of a procedure is the average of all the rankings that the procedure scored for all combinations of test spectrum and sample size.

We also did the following in order to provide visual evaluation on the performances of the four procedures. For the case of $2n = 1024$ and for each estimation procedure, we first ranked the previously computed MSE values, and then plotted those log spectrum estimates that correspond to the 50th smallest MSE value in Figure 4 and Figure 5. These estimates can be taken as "typical" estimates of the four procedures. From a visual sense, these plots agree quite well with the average test rankings given in the previous paragraph.

We report the average computation times for the four procedures in Table 7. From the table one can see that drs is more computationally intensive than the other three procedures.

Although we do not have a universally best procedure, it seems that the proposed procedure drs is a good and conservative procedure to use. It is because, in this experiment, not only it never gave the worst performance, but it also came first for many cases and scored the best average test rankings. Also, it performed better than the other methods when sharp changes are present in the spectrum. However, because of its higher complexity, drs is more suitable for applications that allow off–line processing.

6.4 A Second Experiment

To further demonstrate the good performance of the proposed procedure drs for recovering log spectra with inhomogeneous structures, we conducted a second experiment.

We first generated 100 series with $2n = 256$ using Spectrum 1. Then for each of these

100 series, the procedures mwt, lsp and wla are applied to estimate the log spectrum. We also applied drs to estimate the spectrum, but with a smaller sample size: 75% of $2n$, which is 192. The whole experimental procedure was repeated two more times with $2n = 512$ and $2n = 1024$. Results are summarized in Table 7 in a similar manner as before. From this table it is evident that the proposed procedure drs, although computationally more expensive, is more reliable for spectra with inhomogeneous structures even when the sample size employed is smaller than the one used in other procedures.

7 Conclusion

In this article an automatic log spectrum estimation procedure based on the disconnected regression spline model is proposed. Numerical experiments have been conducted to compare the practical performances of the proposed procedure with some other log spectral density estimation procedures commonly found in the literature. Empirical results from the experiments suggest that the proposed procedure, despite of its high complexity, is a very promising and reliable procedure, especially when the unknown log spectrum possesses inhomogeneous structures.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[2] D. R. Brillinger. *Time Series: Data Analysis and Theory.* Holden–Day, Inc, expanded edition, 1981.

[3] K. P. Burnham and D. R. Anderson. *Model Selection and Inference: A Practical Information-Theoretic Approach.* Springer-Verlag New York Inc., 1998.

[4] G. W. Cobb. *Introduction to Design and Analysis of Experiments.* Springer–Verlag, New York, 1997.

[5] L. Davis. *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, New York, 1991.

[6] J. Fan and E. Kreutzberger. Automatic local smoothing for spectral density estimation. *Scandinavian Journal of Statistics*, 25:359–369, 1998.

[7] D. B. Fogel. Evolutionary computing. *IEEE Spectrum*, pages 26–32, 2000.

[8] J. H. Friedman and B. W. Silverman. Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics*, 31:3–21, 1989.

[9] H.-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis*, 18:231–251, 1997.

[10] T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models.* Chapman and Hall, London, 1990.

[11] C. Kooperberg, S. Bose, and C. J. Stone. Polychotomous regression. *Journal of the American Statistical Association*, 92:117–127, 1997.

[12] C. Kooperberg and C. J. Stone. Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1:301–328, 1992.

[13] C. Kooperberg, C. J. Stone, and Y. K. Troung. Logspline estimation of a possibly mixed spectral distribution. *Journal of Time Series Analysis*, 16:359–388, 1995.

[14] T. C. M. Lee. A simple span selector for periodogram smoothing. *Biometrika*, 84:965–969, 1997.

[15] T. C. M. Lee. Tree–based wavelet regression for correlated data using the minimum description length principle. *Australian & New Zealand Journal of Statistics*, 1999. accepted subject to revision.

[16] T. C. M. Lee. Regression spline smoothing using the minimum description length principle. *Statistics and Probability Letters*, 48:71–82, 2000.

[17] T. C. M. Lee. A stabilized bandwidth selection method for kernel smoothing of the periodogram. *Signal Processing*, 81:419–430, 2001.

[18] P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*, 42:3126–3136, 1994.

[19] Y. Pawitan and F. O'Sullivan. Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association*, 89:600–610, 1994.

[20] K. S. Riedel and A. Sidorenko. Adaptive smoothing of the log–spectrum with multiple tapering. *IEEE Transactions on Signal Processing*, 44:1794–1800, 1996.

[21] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:4611–464, 1978.

[22] P. Stoica and T. Sundin. Optimally smoothed periodogram. *Signal Processing*, 78:253–264, 1999.

[23] G. Wahba. Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75:122–132, 1980.

[24] A. T. Walden, D. B. Percival, and E. J. McCoy. Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, 46:3153–3165, 1998.

[25] M. P. Wand. A comparison of regression spline smoothing procedures. *Computational Statistics*, 15:443–462, 2000.

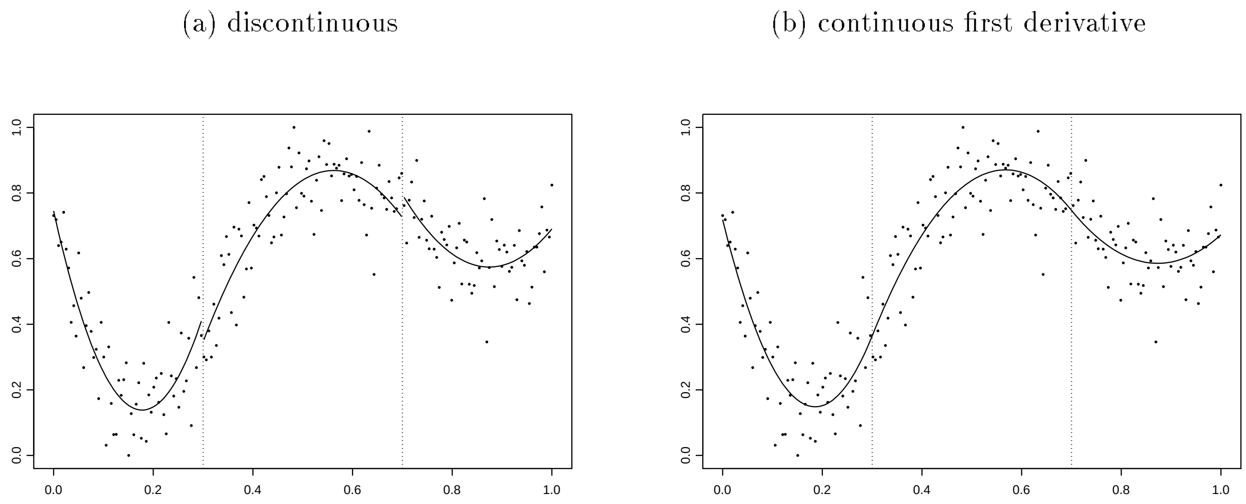

Figure 1: Piecewise quadratic regression splines: there are three pieces in each of the two quadratic splines. The one on the right is constrained to have continuous first derivative. Vertical lines indicate the locations of the knots.

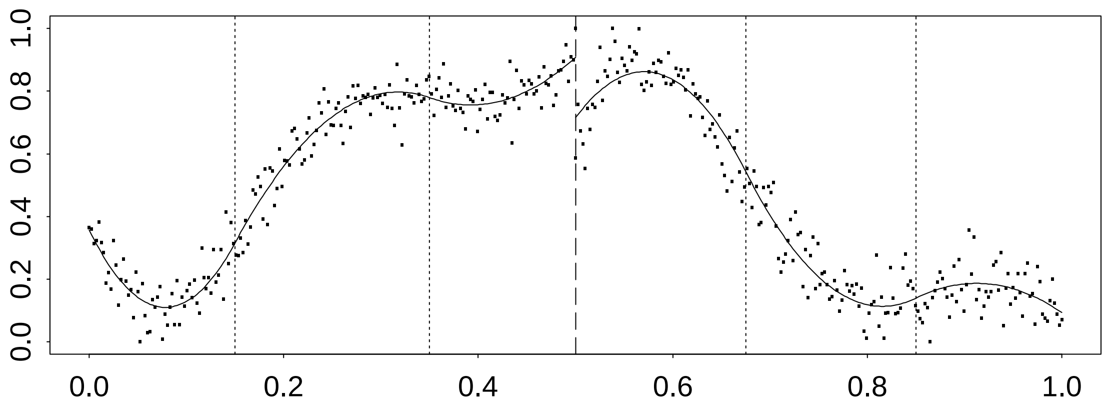

Figure 2: A series of disconnected regression splines that consists of two quadratic regression splines. The vertical broken line $x = 0.5$ indicates the location of the break point, while the vertical dotted lines indicate the locations of the knots.

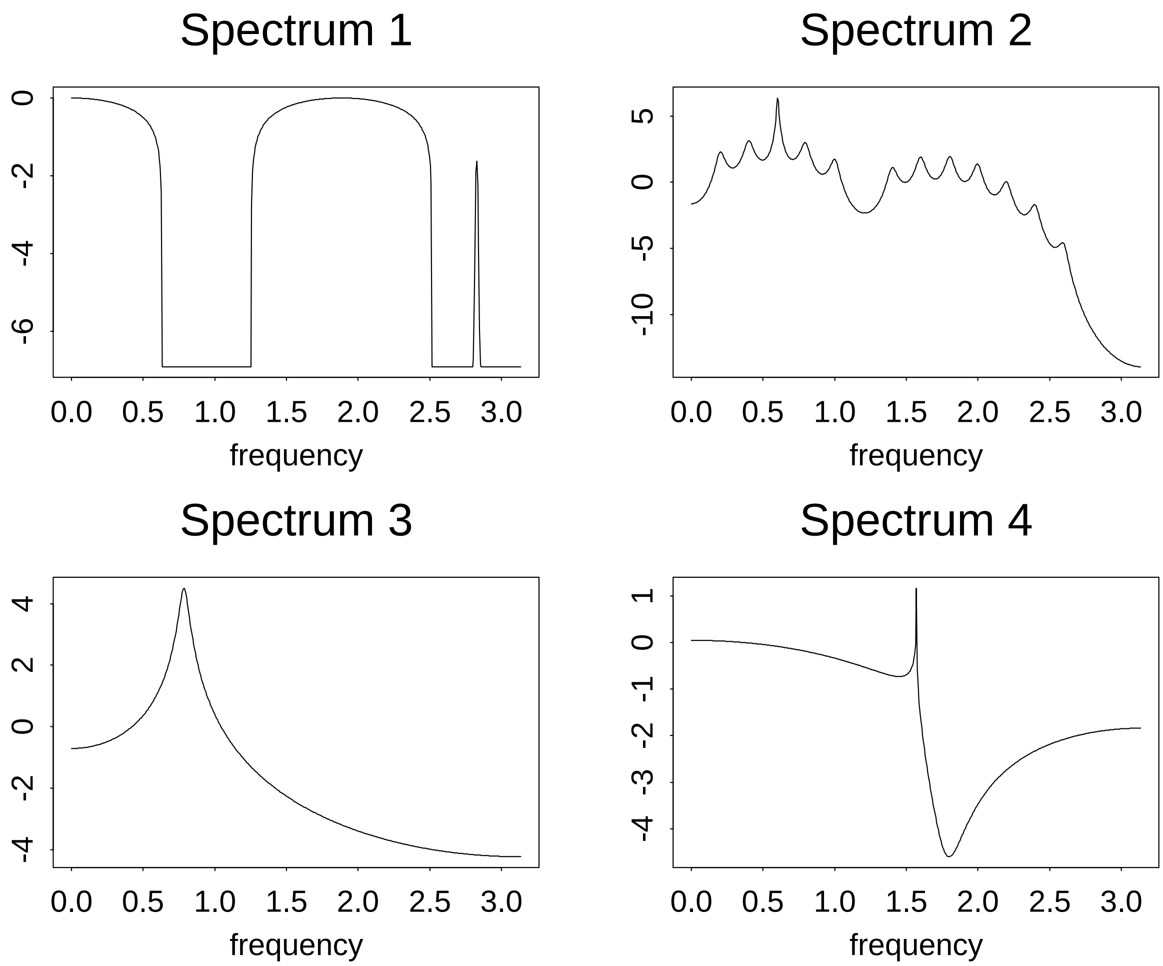

Figure 3: Log of the four testing spectra used in the simulation study.

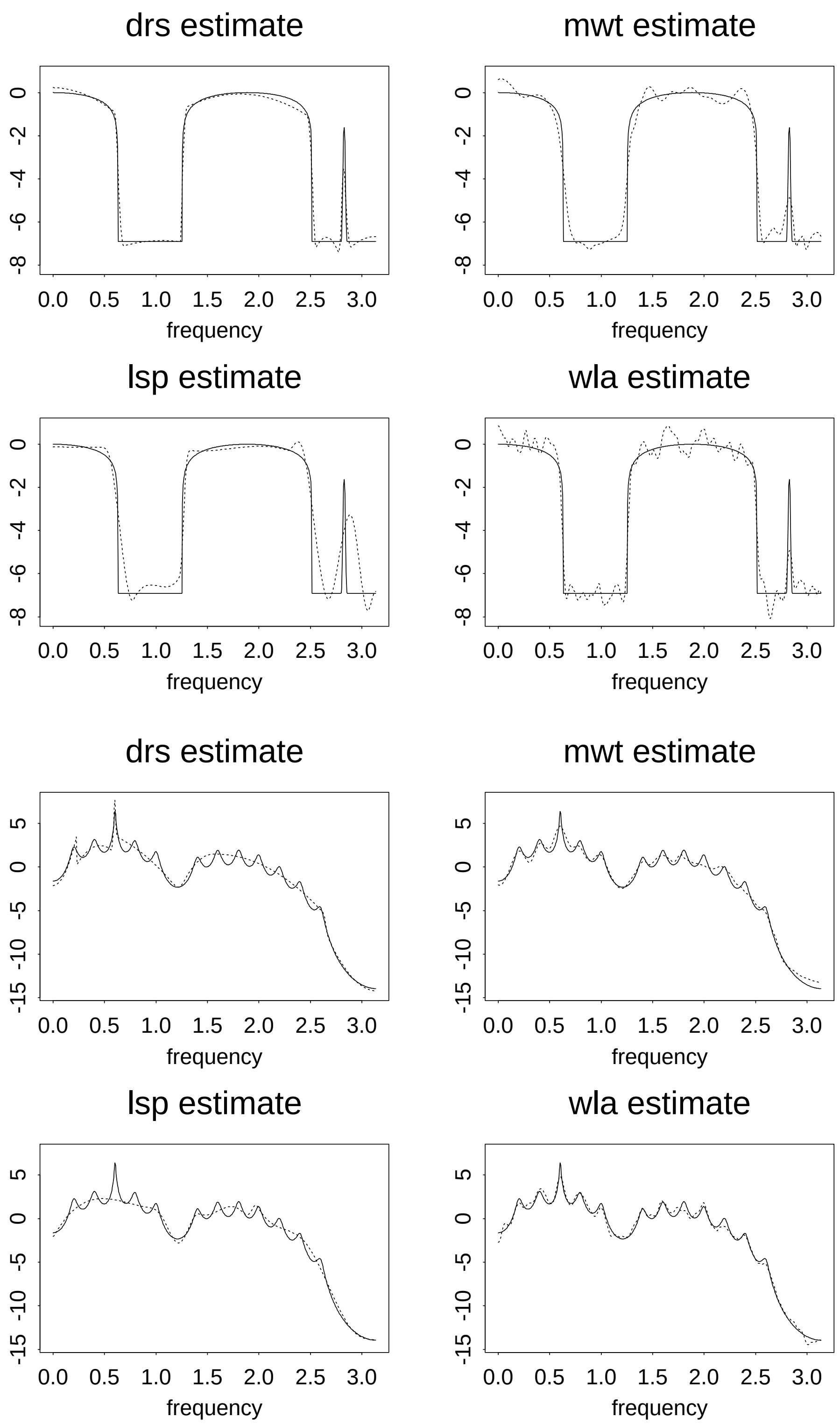

Figure 4: Plots of various log spectrum estimates (dotted lines) when $2n = 1024$ and the true log spectra for Testing Spectra 1 (top two rows) and 2 (bottom two rows).

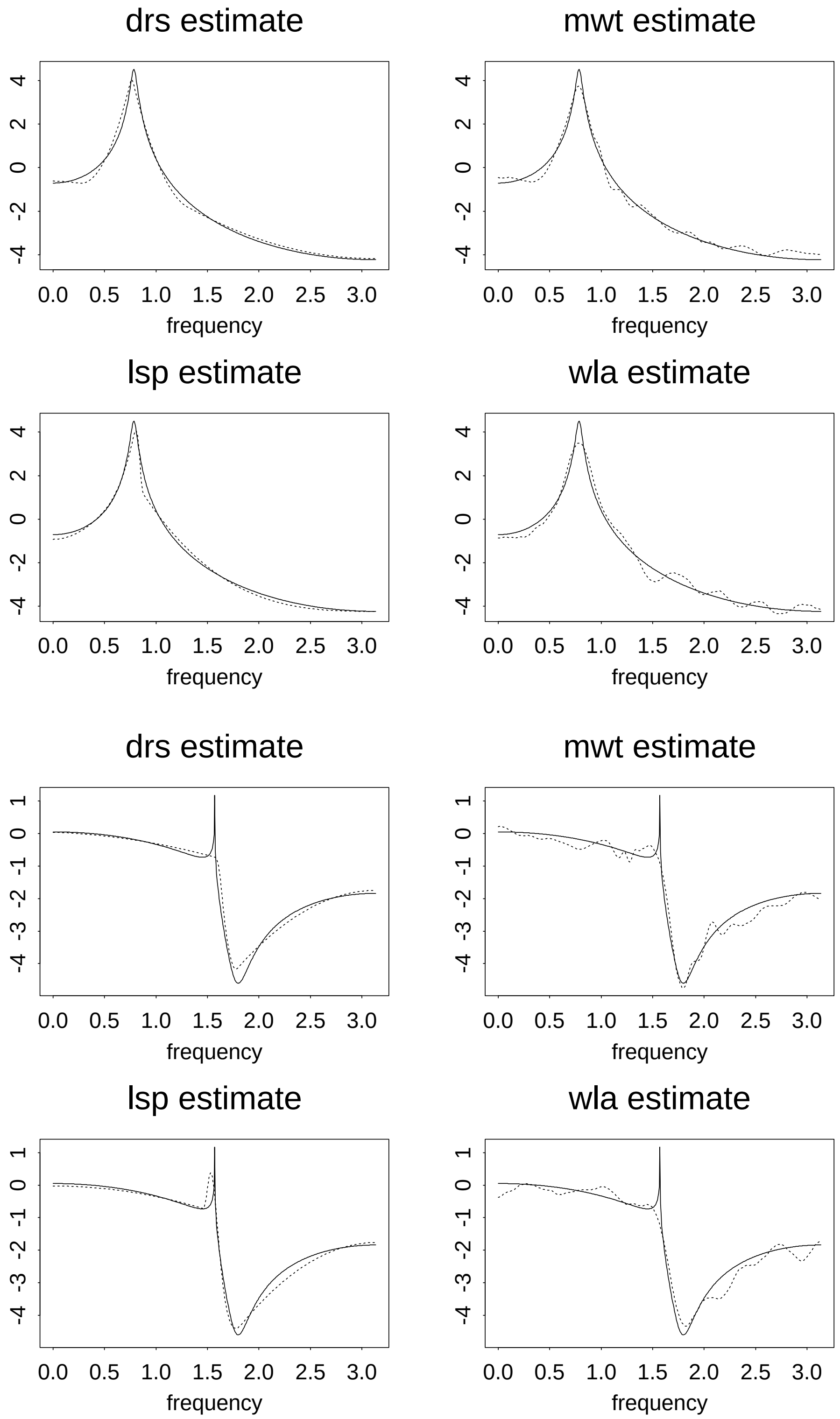

Figure 5: Plots of various log spectrum estimates (dotted lines) when $2n = 1024$ and the true log spectra for Testing Spectra 3 (top two rows) and 4 (bottom two rows).

Spectrum	$2n$	drs mean	drs s.d.	drs rank	mwt mean	mwt s.d.	mwt rank	lsp mean	lsp s.d.	lsp rank	wla mean	wla s.d.	wla rank
1	256	0.6107	0.1980	1	1.8381	0.2817	4	1.2969	0.3877	3	0.7101	0.1395	2
	512	0.3496	0.0995	1	0.8723	0.1036	3	0.9739	0.2426	4	0.4732	0.0764	2
	1024	0.1711	0.0611	1	0.4175	0.0359	3	0.9586	0.3041	4	0.3161	0.0406	2
2	256	0.6577	0.1155	2	2.2963	1.2482	4	0.7253	0.1550	3	0.4763	0.1438	1
	512	0.4954	0.0832	2.5	1.0074	0.6138	4	0.5155	0.0778	2.5	0.2729	0.0538	1
	1024	0.3425	0.0502	2.5	0.3348	0.1976	2.5	0.4484	0.0700	4	0.1623	0.0313	1
3	256	0.1439	0.0578	2	0.1634	0.0542	2	0.1767	0.0939	2	0.2300	0.1413	4
	512	0.0840	0.0265	1.5	0.1035	0.0258	3	0.0871	0.0806	1.5	0.1371	0.0436	4
	1024	0.0448	0.0185	1.5	0.0580	0.0145	3	0.0398	0.0190	1.5	0.0808	0.0233	4
4	256	0.1166	0.0467	1	0.2020	0.0727	3	0.1648	0.0862	2	0.2544	0.1600	4
	512	0.0654	0.0235	1	0.1072	0.0256	3	0.0853	0.0391	2	0.1387	0.0478	4
	1024	0.0425	0.0132	1.5	0.0627	0.0136	3	0.0432	0.0189	1.5	0.0784	0.0246	4

Table 1: Means and Standard Deviations (s.d.) of the MSE values under different combinations of Spectrum, sample size ($2n$) and log spectrum estimation procedures. Also listed are the relative Fisher's Least Significant Difference Test rankings (rank). The average rankings for drs, mwt, lsp and wla are 1.54, 3.13, 2.58 and 2.75 respectively.

Spectrum	$2n$	drs with 75% of $2n$ mean	s.d.	rank	mwt mean	s.d.	rank	lsp mean	s.d.	rank	wla mean	s.d.	rank
	256	0.8824	0.2872	2	1.8648	0.2923	4	1.3172	0.4063	3	0.7184	0.1561	1
1	512	0.3911	0.1011	1	0.8387	0.1135	3	1.0028	0.3351	4	0.4892	0.0778	2
	1024	0.2087	0.0741	1	0.4196	0.0427	3	1.0083	0.3508	4	0.3048	0.0386	2

Table 2: Similar to Table 7 but for the MSE values obtained from the second experiment described in Section 6.4. Notice that drs only used 75% of the observations. The average rankings for drs, mwt, lsp and wla are 1.33, 3.33, 3.67 and 1.67 respectively.

$2n$	drs*	mwt	lsp	wla
256	4	2	2	2
512	12	2	3	2
1024	31	2	4	2

Table 3: Average computation times, in seconds, for the four procedures. All timings were taken from a Sparc Ultra–60 machine. *Note that the time quoted for drs is for one iteration, with 300 chromosomes in each generation. Typically the number of iterations for drs to finish is between 20 and 60.