Exact Simulation of Gaussian Time Series from Nonparametric Spectral Estimates with Application to Bootstrapping

Donald B. Percival[1] and William L. B. Constantine[2]

[1]Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195–5640, USA

[2]Insightful Corporation, 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109–9891, USA

Correspondence to Dr. Donald B. Percival, Applied Physics Laboratory, Box 355640, University of Washington, Seattle, Washington 98195–5640, USA; Phone: (206)–543–1368; Fax: (206)–543–6785; E-mail: `dbp@apl.washington.edu`

Abstract

The circulant embedding method for generating statistically exact simulations of time series from certain Gaussian distributed stationary processes is attractive because of its advantage in computational speed over a competitive method based upon the modified Cholesky decomposition. We demonstrate that the circulant embedding method can be used to generate simulations from stationary processes whose spectral density functions are dictated by a number of popular nonparametric estimators, including all direct spectral estimators (a special case being the periodogram), certain lag window spectral estimators, all forms of Welch's overlapped segment averaging spectral estimator and all basic multitaper spectral estimators. One application for this technique is to generate time series for bootstrapping various statistics. When used with bootstrapping, our proposed technique avoids some – but not all – of the pitfalls of previously proposed frequency domain methods for simulating time series.

Keywords: Circulant Embedding; Gaussian Process; Power Law Process; Surrogate Time Series; Time Series Analysis

1 Introduction

Suppose that we have a real-valued time series that can be considered to be a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of a zero mean Gaussian stationary process $\{X_t\}$ with an unknown spectral density function (SDF) given by $S_X(f), |f| \leq 1/2$, where f is a normalized Fourier frequency (i.e., the sampling interval between X_t and X_{t+1} is assumed to be unity). The problem of estimating $S_X(\cdot)$ based upon an observed time series is the subject of a number of textbooks (see, e.g., Bloomfield (2000), Jenkins and Watts (1968), Kay (1988), Koopmans (1995), Marple (1987), Percival and Walden (1993) or Priestley (1981)). The SDF estimators that are most widely used in practice fall into one of two categories, namely, parametric and nonparametric estimators. Parametric estimators presume a functional form for $S_X(\cdot)$ that is dictated by a certain number of parameters (usually assumed to be small compared to the sample size N). Estimation of these parameters in turn leads to an SDF estimator. Suppose, for example, that $S_X(\cdot)$ has the form of an SDF for a pth order stationary autoregressive (AR) process; i.e., we assume that $\{X_t\}$ can be expressed at least approximately as

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a zero mean Gaussian white noise process with variance σ_ϵ^2, while $\phi_1, \ldots, \phi_p$ are p fixed coefficients. Estimation of the $p+1$ parameters via, say, $\hat{\phi}_j$ and $\hat{\sigma}_\epsilon^2$ yields a parametric SDF estimator of the form

$$\hat{S}_X^{(ar)}(f) \equiv \frac{\hat{\sigma}_\epsilon^2}{\left| 1 - \sum_{j=1}^{p} \hat{\phi}_j e^{-i2\pi fj} \right|^2}.$$

Parametric estimators can also be based upon autoregressive/moving average (ARMA) processes (see, e.g., Kay (1988) or Marple (1987)).

By contrast, nonparametric estimators do not assume a functional form for the SDF, but are essentially based on a presumed theoretical relationship between the SDF and the

autocovariance sequence (ACVS) $\{s_{X,\tau}\}$ for $\{X_t\}$, namely,

$$S_X(f) = \sum_{\tau=-\infty}^{\infty} s_{X,\tau} e^{-i2\pi f\tau}, \tag{1}$$

where

$$s_{X,\tau} \equiv \operatorname{cov}\{X_t, X_{t+\tau}\} = E\{X_t X_{t+\tau}\}$$

(here $E\{X_t\}$ denotes the expected value of the random variable (RV) X_t, and the right-hand equality follows because of the assumption that $E\{X_t\} = 0$). Some well-known nonparametric estimators are direct spectral estimators (including the periodogram), lag window spectral estimators, Welch's overlapped segment averaging (WOSA) spectral estimator and multitaper spectral estimators (see §3 for precise definitions). All of these well-known estimators can be expressed as

$$\hat{S}_X^{(np)}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f\tau}, \tag{2}$$

where $\{\hat{s}_{X,\tau}^{(np)}\}$ is an estimator of the ACVS such that $\hat{s}_{X,-\tau}^{(np)} = \hat{s}_{X,\tau}^{(np)}$. A comparison of the above with Equation (1) indicates that these estimators implicitly assume $s_{X,\tau} = 0$ when $|\tau| \geq N$. A potential difficulty with lag window estimates is that they can be negative at certain frequencies, which is problematic since all theoretical SDFs must be nonnegative. This difficulty can be easily avoided with a proper choice of the lag window (and it cannot occur with direct, WOSA or multitaper spectral estimates). In what follows, we concentrate on nonnegative nonparametric estimators; i.e., $\hat{S}_X^{(np)}(f) \geq 0$ for all f.

Estimation of the SDF along with the assumption that $\{X_t\}$ is a Gaussian zero mean process yields a complete statistical model for a time series. One practical use for this model is to generate simulated time series whose statistical properties closely resemble the time series under study; i.e., we want to generate portions of realizations from a Gaussian zero mean process whose SDF is given by the estimated SDF. For this use, AR and ARMA parametric estimators have been advocated because the functional form of the approximating

2

models leads to an exact and computationally efficient method for generating realizations (Kay (1981)); here the qualifier 'exact' means that the statistical properties of the simulated series are in exact agreement with the estimated model for the time series under study (see Wood and Chan (1994) for additional discussion). The goal of this paper is to point out that, for all nonnegative nonparametric estimators that can be expressed in the form of Equation (2), it is possible to generate realizations exactly and efficiently using circulant embedding. To our knowledge, this interesting connection between commonly used nonnegative nonparametric estimators and circulant embedding has not been pointed out before in the literature.

The remainder of this paper is organized as follows. In §2 we review the circulant embedding method for simulating stationary Gaussian processes. This method is not fully general in that a nonnegativity condition must hold in order for it to work, and there are certain stationary processes for which this condition fails to be true. In §3, we demonstrate that this nonnegativity condition holds for all SDFs corresponding to nonnegative nonparametric estimators, after which we give some computational details for specific estimators. One application for our proposed simulation technique is in bootstrapping time series, which we discuss in §4. We make some concluding remarks in §5.

2   Circulant Embedding Approach to Simulation

The circulant embedding method for generating realizations of Gaussian stationary processes was introduced into the statistical literature in an appendix to Davies and Harte (1987) (and hence is sometimes called the 'Davies–Harte method'; see also Wood and Chan (1994)); however, the method predates this appendix, with Woods (1972) being an early reference in a two dimensional formulation. The basic idea behind the method is as follows.

Suppose that $\mathbf{U}$ is an M dimensional vector whose elements $U_0, U_1, \ldots, U_{M-1}$ are a portion of a zero mean Gaussian stationary process with ACVS $\{s_{U,\tau}\}$; i.e., $\mathbf{U}$ obeys a multivariate Gaussian distribution with a mean vector containing just zeros and with a covariance matrix whose (j,k)th element is given by $s_{U,j-k}$. Given this ACVS for lags 0 to M, we compute

$$S_k \equiv \sum_{\tau=0}^{M} s_{U,\tau} e^{-i2\pi f_k \tau} + \sum_{\tau=M+1}^{2M-1} s_{U,2M-\tau} e^{-i2\pi f_k \tau}, \quad k = 0, 1, \ldots, 2M-1, \tag{3}$$

where $f_k \equiv k/(2M)$; the above is the discrete Fourier transform (DFT) of

$$s_{U,0}, s_{U,1}, \ldots, s_{U,M-2}, s_{U,M-1}, s_{U,M}, s_{U,M-1}, s_{U,M-2}, \ldots, s_{U,1}. \tag{4}$$

The circulant embedding method requires that the S_k's be nonnegative. This condition is known to hold for certain stationary processes (see, e.g., Gneiting (2000) or Craigmile (2003)), but it need not hold for other such processes. Assuming that in fact $S_k \geq 0$, let $Z_0, Z_1, \ldots, Z_{2M-1}$ be a set of $2M$ independent standard Gaussian RVs (hence each Z_k has zero mean and unit variance). We now form the following complex-valued sequence:

$$\mathcal{V}_k \equiv \begin{cases} Z_0 \sqrt{S_0/(2M)}, & k = 0; \\ (Z_{2k-1} + iZ_{2k})\sqrt{S_k/(4M)}, & 1 \leq k < M; \\ Z_{2M-1}\sqrt{S_M/(2M)}, & k = M; \\ \mathcal{V}_{2M-k}^*, & M < k \leq 2M-1; \end{cases} \tag{5}$$

(in the above the asterisk denotes complex conjugation). We then use the DFT to construct the real-valued sequence

$$V_t = \sum_{k=0}^{2M-1} \mathcal{V}_k e^{-i2\pi f_k t}, \quad t = 0, \ldots, 2M-1. \tag{6}$$

By construction, a vector containing $V_0, V_1, \ldots, V_{M-1}$ has exactly the same multivariate Gaussian distribution as the vector $\mathbf{U}$; i.e., we can regard a realization of the first M values

of the V_t sequence as a realization of $U_0, U_1, \ldots, U_{M-1}$ (Davies and Harte (1987) and Wood and Chan (1994)).

In matrix terms, the circulant embedding method follows from the following three facts. First, if $\mathcal{C}$ is an $2M \times 2M$ circulant matrix (i.e., for $j = 1, \ldots, 2M - 1$, its jth row is obtained by circularly shifting its zeroth row to the right by j units), then we can write $\mathcal{C} = \mathcal{F}\Lambda\mathcal{F}^H$, where $\mathcal{F}$ is a $2M \times 2M$ unitary matrix whose (j, k)th element is given by $\exp(-i2\pi jk/2M)/\sqrt{(2M)}$, Λ is a diagonal matrix, and $\mathcal{F}^H$ is the Hermitian transpose of $\mathcal{F}$ (see, e.g., Barnett (1990), Equation 13.27). Second, if $\mathcal{C}$ is a real-valued covariance matrix, the diagonal elements of Λ are necessarily nonnegative. Finally, if $\mathbf{Z}$ is a suitably defined vector of uncorrelated complex-valued Gaussian RVs, then $\mathbf{W} = \mathcal{F}\Lambda^{1/2}\mathbf{Z}$ is a real-valued vector of RVs whose covariance matrix is $\mathcal{C}$ (equations similar to this one are commonly used for simulating Gaussian RVs with a specified covariance structure). The circulant embedding method follows by forming the circulant matrix $\mathcal{C}$ with the zeroth row given by (4). The elements of Λ are proportional to the DFT $\{S_k\}$ of this row. If these elements are all nonnegative, then $\mathcal{C}$ is a valid covariance matrix. The vector $\mathbf{W}$ is then essentially the DFT of suitably weighted uncorrelated elements of $\mathbf{Z}$, and its first M elements have exactly the same expected value and covariance as $U_0, U_1, \ldots, U_{M-1}$.

Dietrich and Newsam (1997) discuss an alternative formulation of the circulant embedding method that is more efficient computationally and somewhat easier to implement. Let $Z_0, Z_1, \ldots, Z_{4M-1}$ be a set of $4M$ independent standard Gaussian RVs, and redefine $\mathcal{V}_k$ in Equation (5) to be

$$\mathcal{V}_k = (Z_{2k} + iZ_{2k+1})\sqrt{S_k/2M}, \quad 0 \le k \le 2M - 1.$$

As before, construct V_t as the DFT of $\mathcal{V}_k$ via Equation (6). Let $\Re\{z\}$ and $\Im\{z\}$ denote the real and imaginary parts of the complex variable z. The sequences $\Re\{V_0\}, \ldots, \Re\{V_{M-1}\}$ and $\Im\{V_0\}, \ldots, \Im\{V_{M-1}\}$ are two independent realizations of $U_0, \ldots, U_{M-1}$.

Although neither variation on the circulant embedding method can be used with certain stationary processes, it is attractive computationally when it is applicable. Once the frequency domain weights S_k have been computed, we can make use of the computationally efficient fast Fourier transform algorithm to compute the DFT needed to generate one or two realizations. The computational burden of this algorithm is of order $M \cdot \log_2(M)$, which is much less than the order M^2 burden required by a competing exact method based upon the modified Cholesky decomposition (Dietrich and Newsam (1997)).

3   Simulations from Nonparametric SDF Estimates

Suppose now that we have a nonnegative nonparametric SDF estimate $\hat{S}_X^{(np)}(\cdot)$ given by Equation (2). Define $\hat{s}_{X,\tau}^{(np)} = 0$ when $|\tau| \geq N$. By Wold's theorem (see, e.g., Priestley (1981)), this estimate is the SDF for some stationary process, which we denote as $\{U_t\}$ to match up with our discussion in §2. If we let $s_{U,\tau} = \hat{s}_{X,\tau}^{(np)}$ in Equation (3) and if $M \geq N$, we obtain

$$
\begin{aligned}
S_k &= \sum_{\tau=0}^{M} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k \tau} + \sum_{\tau=M+1}^{2M-1} \hat{s}_{X,2M-\tau}^{(np)} e^{-i2\pi f_k \tau} \\
&= \sum_{\tau=0}^{M-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k \tau} + \sum_{\tau=1}^{M-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k (2M-\tau)} \quad \text{since } \hat{s}_{X,M}^{(np)} = 0 \\
&= \sum_{\tau=0}^{M-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k \tau} + \sum_{\tau=1}^{M-1} \hat{s}_{X,-\tau}^{(np)} e^{-i2\pi k} e^{-i2\pi f_k (-\tau)} \\
&= \sum_{\tau=-(M-1)}^{M-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k \tau} = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(np)} e^{-i2\pi f_k \tau} = \hat{S}_X^{(np)}(f_k),
\end{aligned}
$$

where we have made use of the facts that $\hat{s}_{X,-\tau}^{(np)} = \hat{s}_{X,\tau}^{(np)}$ for all τ and $\exp(-i2\pi k) = 1$ for all k. Since we are assuming that $\hat{S}_X^{(np)}(f) \geq 0$ for all f, it follows immediately that $S_k \geq 0$ for all k, and hence the circulant embedding method is applicable. Alternatively, we can argue that nonnegativity of the SDF estimate is equivalent to positive definiteness of the $\{\hat{s}_{X,\tau}^{(np)}\}$

sequence; however, $\{\hat{s}_{X,\tau}^{(np)}\}$ is finitely supported, i.e., $\hat{s}_{X,\tau}^{(np)} = 0$ when $|\tau| \geq N$, which implies the validity of the circulant embedding approach (Dietrich and Newsam (1997)). We can thus generate simulations for any length $M \geq N$ from a nonnegative nonparametric SDF estimate $\hat{S}_X^{(np)}(\cdot)$ by using the weights $S_k = \hat{S}_X^{(np)}(k/(2M))$.

Let us now consider several well-known nonparametric SDF estimators that are nonnegative and can be expressed in the form of Equation (2). We note that, if it is unrealistic to assume that the time series under study is a realization of a process whose mean value is known *a priori* (as is usually the case in practical applications), then we need to replace X_t in the equations that follow with the recentered values $X_t - \overline{X}$, where $\overline{X} \equiv \sum_{t=0}^{N-1} X_t/N$ is the sample mean of the series.

3.1 Direct Spectral Estimators

By definition a direct spectral estimator is given by

$$\hat{S}_X^{(d)}(f) = \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi ft} \right|^2,$$

where $\{h_t\}$ is a data taper normalized such that $\sum_{t=0}^{N-1} h_t^2 = 1$. This estimator is obviously nonnegative; moreover, it can be reexpressed as

$$\hat{S}_X^{(d)}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(d)} e^{-i2\pi f\tau},$$

where

$$\hat{s}_{X,\tau}^{(d)} \equiv \begin{cases} \sum_{t=0}^{N-|\tau|-1} h_t X_t h_{t+|\tau|} X_{t+|\tau|}, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N \end{cases}$$

(see, e.g., Percival and Walden (1993), Equations (207c) and (207d)). If we let $h_t = 1/\sqrt{N}$, we obtain the periodogram; i.e.,

$$\hat{S}_X^{(p)}(f) \equiv \frac{1}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i2\pi ft} \right|^2.$$

With the help of a fast Fourier transform algorithm, we can compute the required weights S_k efficiently by zero padding our original time series. Define $h_t X_t = 0$ for $t \geq N$, and, as before, let $f_k = k/(2M)$ for any $M \geq N$. We then compute the DFT of $h_t X_t$, $t = 0, \ldots, 2M - 1$, from which we obtain

$$\left| \sum_{t=0}^{2M-1} h_t X_t e^{-i2\pi f_k t} \right|^2 = \left| \sum_{t=0}^{N-1} h_t X_t e^{-i2\pi f_k t} \right|^2 = \hat{S}_X^{(d)}(f_k) = S_k. \tag{7}$$

3.2   Lag Window Spectral Estimators

By definition a lag window spectral estimator is given by

$$\hat{S}_X^{(lw)}(f) = \sum_{\tau=-(N-1)}^{N-1} w_\tau \hat{s}_{X,\tau}^{(d)} e^{-i2\pi f \tau},$$

where $\{w_\tau\}$ is called a lag window, and $\{\hat{s}_{X,\tau}^{(d)}\}$ is the ACVS estimator corresponding to some direct spectral estimator $\hat{S}_X^{(d)}(\cdot)$ (often taken to be the periodogram $\hat{S}_X^{(p)}(\cdot)$). The above is an example of Equation (2) if we define the corresponding ACVS estimator $\{\hat{s}_{X,\tau}^{(lw)}\}$ to be $\hat{s}_{X,\tau}^{(lw)} = w_\tau \hat{s}_{X,\tau}^{(d)}$. A lag window estimator can be reexpressed as

$$\hat{S}_X^{(lw)}(f) = \int_{-1/2}^{1/2} W(f - f') \hat{S}^{(d)}(f') \, df', \quad \text{where } W(f) \equiv \sum_{\tau=-(N-1)}^{N-1} w_\tau e^{-i2\pi f \tau}$$

(see, e.g., Percival and Walden (1993), Equation (238a)). The function $W(\cdot)$ is called a smoothing window. Since $\hat{S}^{(d)}(f) \geq 0$, a sufficient (but not necessary) condition for $\hat{S}_X^{(lw)}(\cdot)$ to be nonnegative is that $W(f) \geq 0$ for all f. Popular lag windows that are guaranteed to yield a nonnegative estimate include the Parzen, Papoulis and Daniell windows (see Priestley (1981) or Percival and Walden (1993)). The advantage of a lag window estimator is that its variance is typically smaller than that of the corresponding direct spectral estimator because of the smoothing across frequencies.

We can compute the weights S_k for a nonnegative lag window estimator as follows. We first use Equation (7) to compute $\hat{S}_X^{(d)}(f_k)$, $k = 0, \ldots, 2M - 1$. We then use an inverse DFT

to form

$$\tilde{s}_{X,\tau}^{(d)} \equiv \frac{1}{2M} \sum_{k=0}^{2M-1} \hat{S}_X^{(d)}(f_k)e^{i2\pi f_k \tau}, \quad \tau = 0,\ldots,2M-1$$

(we note that $\tilde{s}_{X,\tau}^{(d)} = \hat{s}_{X,\tau}^{(d)}$ for $\tau = 0,\ldots,N-1$). We then define the sequence

$$\tilde{w}_\tau = \begin{cases} w_\tau, & 0 \leq \tau < N; \\ 0, & N \leq \tau \leq 2M - N; \\ w_{\tau-2M}, & 2M - N < \tau \leq 2M - 1. \end{cases}$$

Finally we use a DFT to obtain the desired weights via

$$\sum_{\tau=0}^{2M-1} \tilde{w}_\tau \tilde{s}_{X,\tau}^{(d)} e^{-i2\pi f_k \tau} = \hat{S}_X^{(lw)}(f_k) = S_k.$$

3.3 Welch's Overlapped Segment Averaging Spectral Estimator

A WOSA spectral estimator consists of (i) segmenting the time series into, say, N_B different segments, with each segment containing $N_S < N$ values; (ii) forming a direct spectral estimator for each segment, with the same data taper $\{h_t\}$ being applied to each segment; and then (iii) averaging all of the direct spectral estimators together (Welch (1967)). An important aspect of this estimator is that the segments are allowed to overlap one another. Mathematically, we can write

$$\hat{S}_X^{(wosa)}(f) = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \hat{S}_{jN_O}^{(d)}(f), \text{ where } \hat{S}_l^{(d)}(f) \equiv \left| \sum_{t=0}^{N_S-1} h_t X_{t+l} e^{-i2\pi ft} \right|^2, \quad 0 \leq l \leq N - N_S;$$

(8)

here N_O is a positive integer that controls how much overlap there is between segments and that must satisfy both $N_O \leq N_S$ and $N_O(N_B - 1) = N - N_S$, while $\{h_t\}$ is a data taper appropriate for a series of length N_S (i.e., $\sum_{t=0}^{N_S-1} h_t^2 = 1$). Since the WOSA estimator is an average of direct spectral estimators, we can reexpress it in the form of Equation (2); i.e.,

$$\hat{S}_X^{(wosa)}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(wosa)} e^{-i2\pi f\tau},$$

9

where

$$
\hat{s}_{X,\tau}^{(wosa)} \equiv
\begin{cases}
\frac{1}{N_B} \sum_{j=0}^{N_B-1} \sum_{t=0}^{N_S-|\tau|-1} h_t X_{t+jN_O} h_{t+|\tau|} X_{t+jN_O+|\tau|}, & |\tau| \le N_S - 1; \\
0, & |\tau| \ge N_S.
\end{cases}
$$

It is clear from Equation (8) that a WOSA estimator must be nonnegative.

If we assume that the number of points N_S in each segment is an even number and if we use the Hanning data taper, i.e.,

$$
h_t = \left(\frac{2}{3(N_S+1)} \right)^{1/2} \left[1 - \cos \left(\frac{2\pi(t+1)}{N_S+1} \right) \right], \quad t = 0, \dots, N_S - 1,
$$

then a popular choice for N_O is to set it equal to $N_S/2$, which gives us segments that overlap by 50%. In this case the number of segments is given by $N_B = [2(N - N_S)/N_S] + 1$, where it is assumed that N_S is selected such that this expression yields an integer value for N_B.

By defining $h_t = 0$ for $t \ge N_S$, we can obtain the weights S_k by averaging the outcome of N_B applications of Equation (7):

$$
\frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{2M-1} h_t X_{t+jN_O} e^{-i2\pi f_k t} \right|^2 = \frac{1}{N_B} \sum_{j=0}^{N_B-1} \left| \sum_{t=0}^{N_S-1} h_t X_{t+jN_O} e^{-i2\pi f_k t} \right|^2 = \hat{S}_X^{(wosa)}(f_k) = S_k.
$$

3.4   Multitaper Spectral Estimators

A multitaper spectral estimator is given by

$$
\hat{S}_X^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{t=0}^{N-1} h_{k,t} X_t e^{-i2\pi ft} \right|^2,
$$

where $\{h_{k,t}\}$, $k = 0, \dots, K - 1$, is a set of K orthonormal data tapers; i.e.,

$$
\sum_{t=0}^{N-1} h_{k,t} h_{k',t} =
\begin{cases}
1, & \text{if } k = k'; \\
0, & \text{otherwise}
\end{cases}
$$

(see Thomson (1982), Percival and Walden (1993) and Riedel and Sidorenko (1995)). This estimator is always nonnegative and is an example of $\hat{S}_X^{(np)}(\cdot)$ of Equation (2) since it can

be reexpressed as

$$\hat{S}_X^{(mt)}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{s}_{X,\tau}^{(mt)} e^{-i2\pi f \tau},$$

where

$$\hat{s}_{X,\tau}^{(mt)} \equiv \begin{cases} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{t=0}^{N-|\tau|-1} h_{k,t} X_t h_{k,t+|\tau|} X_{t+|\tau|}, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N. \end{cases}$$

There are two families of multitapers in wide use. Thomson (1982) proposed use of tapers based upon discrete prolate spheroidal sequences, which require specialized techniques to compute (see Percival and Walden (1993) for details). Riedel and Sidorenko (1995) proposed a family of sinusoidal data tapers that work quite well in most practical applications and are approximated well by a simple formula.

The weights S_k can be computed efficiently in a manner similar to what is given in Equation (7) for direct spectral estimators. With $h_{k,t} X_t \equiv 0$ for $t \geq N$, we have

$$\frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{t=0}^{2M-1} h_{k,t} X_t e^{-i2\pi f_k t} \right|^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{t=0}^{N-1} h_{k,t} X_t e^{-i2\pi f_k t} \right|^2 = \hat{S}_X^{(mt)}(f_k) = S_k.$$

4   Bootstrapping Time Series

One potential use for our proposed methodology is in bootstrapping time series to assess the sampling variability in certain statistics. Frequency domain techniques for bootstrapping time series have been studied previously and are discussed in detail in Davison and Hinkley (1997), Berkowitz and Kilian (2000), Härdle et al. (2003), Lahiri (2003), and references therein. In addition, the form of frequency domain bootstrapping in which we are interested (i.e., creating simulated series by manipulating the DFT) is essentially the same as the method of surrogate data originally proposed by Theiler et al. (1992) for detecting nonlinear structure in a stationary time series (see also Schreiber and Schmitz (2000), Chan and Tong (2001), Krantz and Schreiber (2004) and references therein). In the literature cited above,

there are several different proposals for how to do the manipulations. We will focus on one proposed in Theiler et al. (1992) and another described by Davison and Hinkley (1997).

4.1 Alternative Frequency Domain Methods for Bootstrapping

The Theiler et al. frequency domain method consists of simply randomly reassigning the phases in the DFT via the following steps. Given a real-valued time series $X_0, \ldots, X_{N-1}$, we compute its DFT

$$\mathcal{X}_k = \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N}, \quad k = 0, \ldots, N-1.$$

Define $\theta_0 = 0$. Let θ_k, $k = 1, \ldots, \lfloor (N-1)/2 \rfloor$, be a random sample from a uniform distribution over the interval $[0, 2\pi]$, where $\lfloor x \rfloor$ refers to the largest integer that is less than or equal to x. Additionally, if N is even, let $\theta_{N/2}$ be a random deviate from a distribution that assigns a probability of $1/2$ to the points 0 and π. For $N/2 < k \leq N-1$, let $\theta_k = -\theta_{N-k}$. The simulated series is then given by the inverse DFT of $\{\mathcal{X}_k e^{i\theta_k}\}$:

$$\widetilde{X}_t = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{X}_k e^{i\theta_k} e^{i2\pi kt/N}, \quad t = 0, \ldots, N-1.$$

By construction, the series $\{\widetilde{X}_t\}$ is real-valued, and the sample means and the periodograms for $\{\widetilde{X}_t\}$ and $\{X_t\}$ are constrained to be identical.

The Davison–Hinkley frequency domain method again randomizes the phases in $\{\mathcal{X}_k\}$, but also includes an averaging operation that modifies the amplitudes as well. Let θ_k, $k = 1, \ldots, N-1$, be a random sample from a uniform $[0, 2\pi]$ distribution, and define $\mathcal{A}_k \equiv \mathcal{X}_k e^{i\theta_k}$. We then form

$$\widetilde{\mathcal{X}}_k = 2^{-1/2}(\mathcal{A}_k + \mathcal{A}_{N-k}^*), \quad k = 1, \ldots, N-1.$$

With $\widetilde{\mathcal{X}}_0 = \mathcal{X}_0$, the simulated series is now given by

$$\widetilde{X}_t = \frac{1}{N} \sum_{k=0}^{N-1} \widetilde{\mathcal{X}}_k e^{i2\pi kt/N}, \quad t = 0, \ldots, N-1.$$

12

By construction, the sample means for $\{\widetilde{X}_t\}$ and $\{X_t\}$ must be identical, but the periodograms need not be.

For the purposes of bootstrapping, both the Theiler et al. and Davison–Hinkley methods work reasonably well for certain processes and statistics, but can fail badly in some situations. Davison and Hinkley (1997) stress that their method and related frequency domain methods cannot be expected to work well unless (i) the original time series is well modeled by a stationary Gaussian process with 'short memory' (i.e., with an autocovariance sequence that does not decay to zero too slowly) and (ii) the statistic under study is a linear contrast (see, e.g., Scheffé (1959), p. 66; in particular, the statistic must be invariant if we force a change in the sample mean of the time series by adding a constant to all its original values). An additional concern about the Theiler et al. and Davison–Hinkley methods is that they yield simulated series with a circularity assumption built into them. One consequence of this defect is that $\widetilde{X}_0$ and $\widetilde{X}_{N-1}$ in any simulated series of length N will tend to be close to each other, even if X_0 and X_{N-1} are not.

By contrast, the circulant embedding approach has several characteristics that can overcome some – but not all – of the deficiencies in the Theiler et al. and Davison–Hinkley methods. First, because the circulant embedding approach creates simulated series $\{\widetilde{X}_t\}$ of length N from a DFT involving $2N$ or more terms, the resulting series has a sample mean and a periodogram that can vary from one realization to the next. In addition, the first and last parts of the simulated series are not artificially coupled together, and, as a consequence, the values for $\widetilde{X}_0$ and $\widetilde{X}_{N-1}$ are typically decoupled. We illustrate this decoupling in Figure 1. The top plot is a 2048 point series representing the tip velocity of a vertically mounted thin steel beam subjected to transverse sinusoidal excitation using a electromechanical shaker (Moon (1992); Constantine (1999)). The beam tip is mounted between two rare earth magnets, and the opposite tip is excited close to the beam's first

natural mode of vibration (approximately 10 cycles per second). The excitation amplitude was adjusted to produce (seemingly) chaotic motion. The bottom four plots show four typical simulated series generated from the chaotic beam series using, from top to bottom, (i) the Theiler et al. method, (ii) the Davison–Hinkley method, (iii) the circulant embedding method based upon a periodogram and (iv) this same method using a WOSA estimator formed from $N_B = 31$ segments. For the WOSA estimator there were $N_S = 128$ values in each segment; the segments overlapped by 50%; and the Hanning data taper was used on each segment. While the first and last points in the chaotic beam series are rather far apart, the corresponding points in the Theiler et al. and Davison–Hinkley series are close to one another. By contrast, the two series generated via circulant embedding have end points that do not match up particularly well. To look at this aspect of the simulations more quantitatively, we constructed a thousand simulated series for each method and computed the sample correlation coefficient $\hat{\rho}$ between the first point $\widetilde{X}_0$ and last point $\widetilde{X}_{N-1}$ in the simulated series. The values of $\hat{\rho}$ for each method are listed to the right of the example series plotted in Figure 1. As expected, the end points for the Theiler et al. and Davison–Hinkley method are highly correlated, whereas those for the two circulant embedding methods are practically uncorrelated.

4.2   Monte Carlo Comparison of Bootstrapping Methods

To explore how well the circulant embedding method works, we conducted Monte Carlo experiments to see how well series simulated by this method can capture the variance of various statistics. For some statistics such as the sample variance and the unit lag sample autocorrelation, we found the circulant embedding approach to be comparable to the Theiler et al. and Davison–Hinkley methods. Here we report on experiments regarding two simple statistics for which these latter two methods are – or might be – problematic. The first
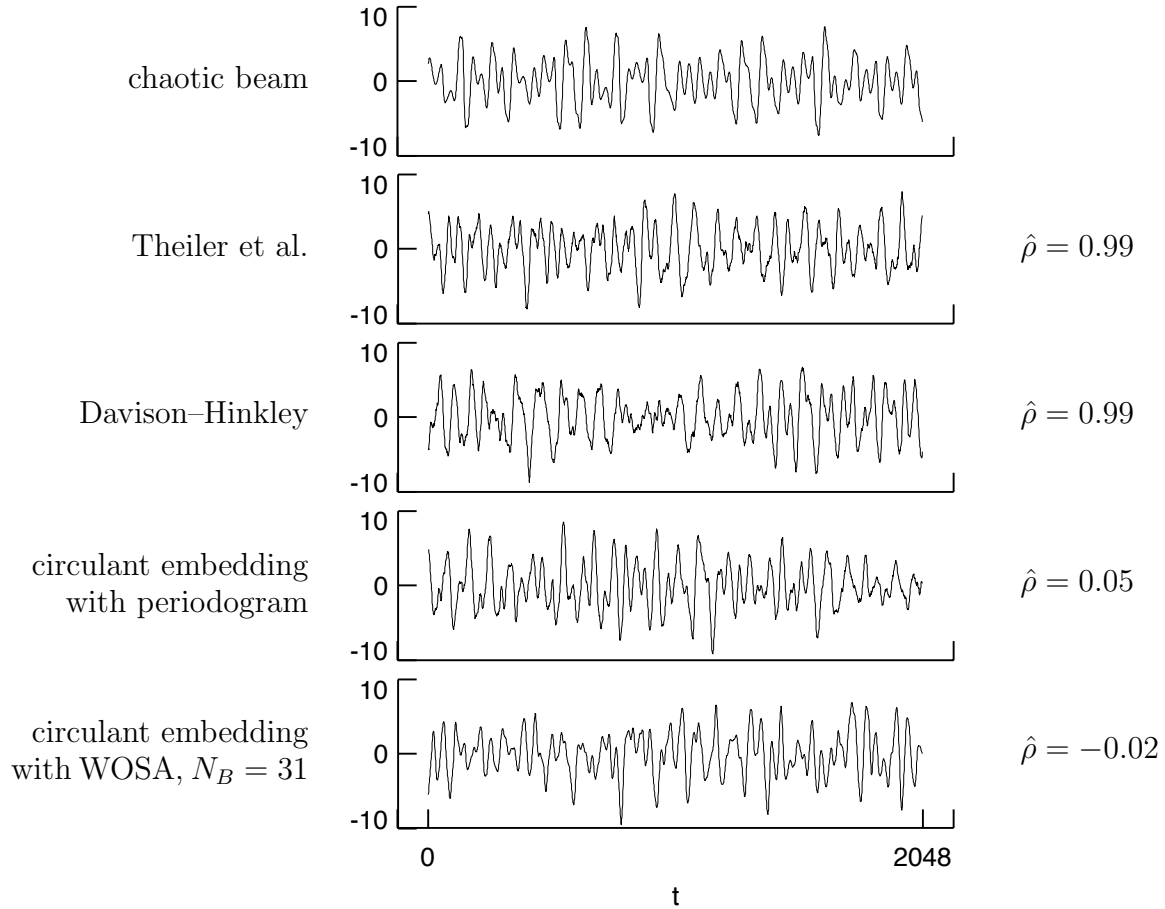
Figure 1: Time series of a beam undergoing apparently chaotic motion (top plot), along with associated simulated series generated using, from top to bottom, the Theiler et al. method, the Davison–Hinkley method, the circulant embedding method based upon the periodogram and the same method based upon a WOSA estimate with $N_B = 31$ segments (see text for further details).
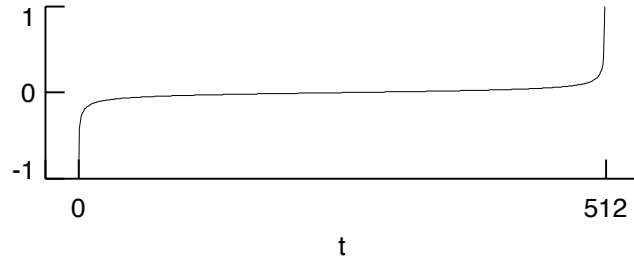
15

Figure 2: Weights $\{a_t\}$ for the Abelson–Tukey test statistic for $N = 512$ (see text for further details).

statistic is the sample mean $\overline{X}$. The rationale for considering $\overline{X}$ is that series generated by the Theiler et al. and Davison–Hinkley methods all have a sample mean equal to $\overline{X}$ and hence cannot be used to assess the variability in $\overline{X}$. We also considered the Abelson–Tukey test statistic, which is given by

$$\widehat{T} \equiv \sum_{t=0}^{N-1} a_t X_t, \text{ where } a_t = \left[t \left(1 - \frac{t}{N} \right) \right]^{1/2} - \left[(t+1) \left(1 - \frac{t+1}{N} \right) \right]^{1/2}.$$

The weights $\{a_t\}$ for this statistic are plotted in Figure 2 for the case $N = 512$. We chose to look at $\widehat{T}$ because Davison and Hinkley (1997) considered it previously in the context of bootstrapping. As discussed in their book, $\widehat{T}$ is an 'optimal' test for trend in a time series that operates by essentially contrasting points at the beginning of the series with those at the end. In this context, the fact that the Theiler et al. and Davison–Hinkley methods generate series whose end points are highly correlated might lead us to question how well their bootstrapped samples can assess the sampling variability in $\widehat{T}$.

We used the following six models as test cases in our Monte Carlo experiments:

(a) the AR(1) model $X_t = 0.9X_{t-1} + \epsilon_t$, which is one of the models used by Chan and Tong (2001) in their study of Fourier-based methods for generating surrogate series;

(b) the AR(2) model $X_t = 0.75X_{t-1} - 0.5X_{t-2} + \epsilon_t$, which has been used as a test case in the engineering literature;

16

(c) the AR(2) model $X_t = 1.14X_{t-1} - 0.31X_{t-2} + \epsilon_t$, which Davison and Hinkley (1997) determined to be a good model for a time series related to the flow of the Rio Negro (we henceforth refer to this model as 'Rio');

(d) the AR(4) model $X_t = .7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + \epsilon_t$, which has been used in the engineering literature as a challenging test for spectral analysis;

(e) a fractionally differenced (FD) process with long memory parameter $\delta = 0.2$, which is close to a model used to describe atmospheric pressure fluctuations over the North Pacific (Percival et al. (2001)); and

(f) an FD process with $\delta = 0.45$, which is close to a model used to describe yearly variations in a historical record of minima of the Nile River (Beran (1994)).

For all six models the variance of the driving white noise $\{\epsilon_t\}$ was set so that the process variance was unity. Figure 3 shows the SDFs for each model (left-hand column). It should be noted that, while the SDFs for models (a) to (d) are continuous and bounded over $f \in [-1/2, 1/2]$, those for the FD processes are approximately proportional to $|f|^{-2\delta}$ at low frequencies and hence have a singularity at $f = 0$. To the right of each SDF plot, there is a series of length $N = 512$ simulated from the corresponding model using the circulant embedding approach. For each series, we used the same set of 1024 random Gaussian deviates to make it easier to contrast the correlation properties of the different models.

We conducted the Monte Carlo experiments in the following manner (the results of the experiments are summarized in Table 1). For each model, we generated 250 simulated series, each of length $N = 512$, using the circulant embedding method based upon the ACVS corresponding to the known SDF. We computed $\overline{X}$ and $\widehat{T}$ for each simulated series and then computed the sample variances based upon the 250 values for each statistic. The square roots of these variances (i.e., the sample standard deviations) are shown in the bottom rows
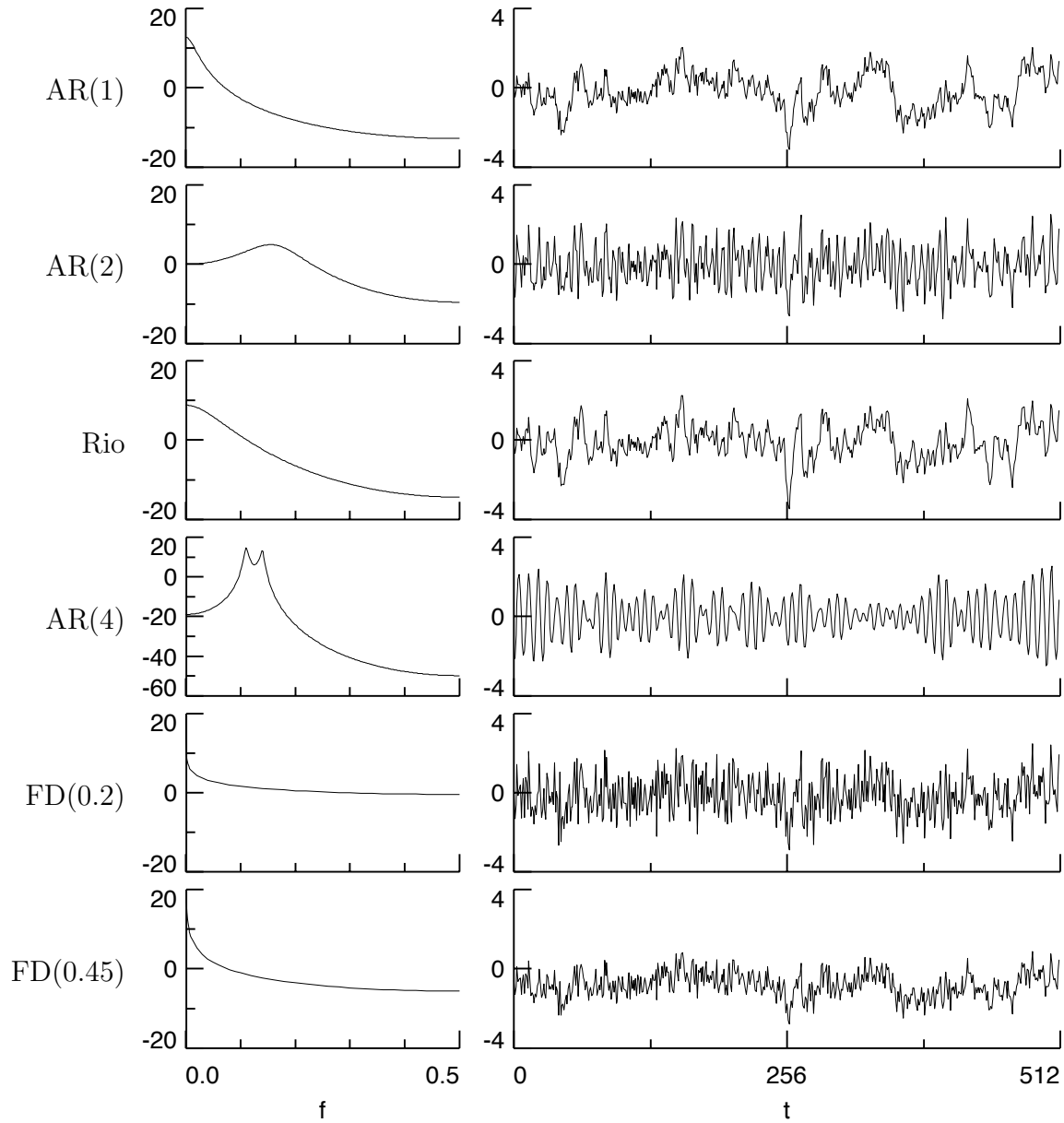
17

Figure 3: The spectral density functions (SDFs) for six models used in the Monte Carlo experiments (left-hand column), along with one realization of length $N = 512$ from each model generated by the circulant embedding method using the same set of $2N = 1024$ standard Gaussian random deviates (right-hand). The SDFs are expressed in decibels; i.e., $10 \cdot \log_{10}(S_X(f))$ is plotted versus f.

of Table 1 (marked as 'Monte Carlo'). We take these values to be a measure of the true variability in $\overline{X}$ and $\widehat{T}$. For each simulated series, we then generated 100 bootstrap samples for each of five implementations of the circulant embedding method, now based upon the ACVS corresponding to an estimated SDF. The five estimated SDFs were the periodogram and four WOSA estimates, with the latter using a Hanning data taper on segments of size $N_S = 256$, 128, 64 and 32 with a 50% overlap, yielding a total of, respectively, $N_B = 3$, 7, 15 and 31 segments. For each of the five implementations, we computed both $\overline{X}$ and $\widehat{T}$ for each of the 100 bootstrap estimates and then used these in turn to compute sample variances (we also did the same for $\widehat{T}$ using the Theiler et al. and Davison–Hinkley methods). Finally, we averaged these bootstrap sample variances for the 250 simulated series. The square roots of these averages are reported in Table 1. (In fact, we replicated the entire experiment multiple times so that we could determine how much variability there was in the numbers reported in the table. Generally, the entries in the table changed by at most one in the last recorded digit when the entire experiment was replicated.)

We can draw the following conclusions from our Monte Carlo experiments. With regard to the sample mean $\overline{X}$, circulant embedding with the periodogram tends to systematically underestimate the actually variability; however, this scheme with the WOSA estimators with more than $N_B = 3$ segments does reasonably well for the 'short memory' processes (AR(1), AR(2), Rio and AR(4)), but it badly underestimates the variability in the two FD (long memory) processes. For $\widehat{T}$, the five circulant embedding implementations give quite similar results. They all capture the actual variability in $\widehat{T}$ reasonably well for short memory processes. In comparison with the Theiler et al. and Davison–Hinkley methods for these processes, they are either competitive (AR(2) case) or offer a modest improvement (AR(1), Rio and AR(4) cases). None of the methods reported in Table 1 works well with the FD process with $\delta = 0.45$.

$$\overline{X}$$

method\model	AR(1)	AR(2)	Rio	AR(4)	FD(0.2)	FD(0.45)
periodogram	0.11	0.025	0.06	0.0057	0.057	0.09
WOSA, $N_B = 3$	0.16	0.037	0.10	0.0060	0.081	0.12
WOSA, $N_B = 7$	0.17	0.041	0.11	0.0061	0.083	0.12
WOSA, $N_B = 15$	0.16	0.043	0.11	0.0062	0.080	0.11
WOSA, $N_B = 31$	0.15	0.044	0.11	0.0066	0.073	0.09
Monte Carlo	**0.19**	**0.045**	**0.12**	**0.0061**	**0.150**	**0.70**

$$\widehat{T}$$

method\model	AR(1)	AR(2)	Rio	AR(4)	FD(0.2)	FD(0.45)
periodogram	5.5	2.0	4.1	1.57	3.1	4.3
WOSA, $N_B = 3$	5.6	2.0	4.0	1.56	3.0	4.1
WOSA, $N_B = 7$	5.6	2.1	4.2	1.56	3.0	3.9
WOSA, $N_B = 15$	5.5	2.1	4.2	1.56	2.9	3.5
WOSA, $N_B = 31$	5.1	2.1	4.1	1.56	2.8	3.1
Theiler et al.	4.8	2.1	3.7	1.71	2.8	3.4
Davison & Hinkley	4.7	2.1	3.7	1.72	2.8	3.4
Monte Carlo	**6.1**	**2.1**	**4.4**	**1.61**	**3.4**	**5.9**

Table 1: Standard deviations for the sample mean $\overline{X}$ and Abelson–Tukey test statistic $\widehat{T}$ computed via Monte Carlo experiments. The bottom rows ('Monte Carlo') are measures of the actual standard deviations for each statistic, while the other rows indicate the average standard deviation determined using various bootstrapping methods.

The fact that the circulant embedding approach does not lead to a significant improvement for the long memory FD processes is undoubtedly because nonparametric SDF estimates are not reliable at frequencies close to zero for these processes. From a time domain perspective, this failure is because these SDF estimators have a corresponding ACVS that is finitely supported, whereas the ACVS for an FD process decays slowly to zero. In the frequency domain, these facts translate into the SDF estimates being continuous at $f = 0$ whereas the true SDF is discontinuous. As a result, the WOSA estimator with a large number of segments N_B tends to have small variability – but large biases – close to zero frequency, whereas the periodogram has small bias but large variability in that region. For the short memory processes, the WOSA estimator can do well at estimating the SDF over all frequencies, which seems to be the key to its superior performance over the periodogram (tests to date indicate that lag window and multitaper estimators give results comparable to those reported in Table 1 for the WOSA estimator if the former are adjusted to have comparable statistical properties).

5   Conclusions

We have demonstrated that it is always possible to use the exact circulant embedding method for simulating a time series from a Gaussian stationary process whose SDF is dictated by an SDF estimate produced by some of the most widely used nonnegative nonparametric estimators. Since one of the uses for spectral analysis is to generate simulated series for a phenomenon of interest, this methodology places nonparametric estimators on an equal footing with commonly used parametric estimators, for which exact simulations can be obtained using the underlying parametric model; however, it should be noted that, if the parametric model has relatively few parameters compared to the sample size N, then the

parametric method can generate individual realizations faster than the circulant embedding method (for autoregressive models of order p, the computational burden required for each simulation is $O(Np)$ as compared to $O(N \cdot \log_2(N))$ for circulant embedding). Nonetheless, because the circulant embedding method is based upon the DFT and hence can be implemented efficiently using a fast Fourier transform algorithm, individual simulations can be computed rapidly. For example, using C code called from `S-Plus` running under Red Hat Linux on a 2.6 gigahertz Pentium 4 processor, we were able to generate individual series of length $N = 2048$ similar to the ones in the bottom two rows of Figure 1 in approximately 5.5 milliseconds.

We have also shown that simulations produced by circulant embedding can be used – with some caution – for bootstrapping certain statistics and that this method performs well for statistics such as the sample mean for which other proposed frequency domain simulation techniques automatically fail. The main cautionary notes are that (i) the time series under study must not exhibit marked departures from Gaussianity (a point as emphasized by Davison and Hinkley (1997)) and (ii) the series cannot exhibit long memory. In addition, our proposed method seems to offer some modest improvements for deducing the sampling properties of statistics that are handled reasonably well by other frequency domain simulation techniques. Full exploration of the bootstrapping potential of our proposed method is a topic for future research.

6   Software Availability

The circulant embedding approach described in this paper has been implemented in the `S-Plus` language. Readers should contact the authors if they are interested in obtaining this software.

Acknowledgments

References

Barnett, S. (1990). *Matrices: Methods and Applications.* Clarendon Press, Oxford.

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48.

Bloomfield, P. (2000). *Fourier Analysis of Time Series: An Introduction (Second Edition).* Wiley, New York.

Chan, K.-S. and Tong, H. H. (2001). *Chaos: A Statistical Perspective.* Springer–Verlag, New York.

Constantine, W. L. B. (1999). *Wavelet Techniques for Chaotic and Fractal Dynamics.* PhD thesis, Mechanical Engineering Department, University of Washington.

Craigmile, P. F. (2003). Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24:505–511.

Davies, R. B. and Harte, D. S. (1987). Tests for the Hurst effect. *Biometrika*, 74:95–102.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge, UK.

Dietrich, C. R. and Newsam, G. N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107.

Gneiting, T. (2000). Power-law correlations, related models for long-range dependence, and their simulation. *Journal of Applied Probability*, 37:1104–1109.

Härdle, W., Horowitz, J., and Kreiss, J.-P. (2003). Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral Analysis and its Applications*. Holden–Day, San Francisco.

Kay, S. M. (1981). Efficient generation of colored noise. *Proceedings of the IEEE*, 69:480–481.

Kay, S. M. (1988). *Modern Spectral Estimation: Theory and Application*. Prentice-Hall, Englewood Cliffs, New Jersey.

Koopmans, L. H. (1995). *The Spectral Analysis of Time Series (Second Edition)*. Academic Press, San Diego.

Krantz, H. and Schreiber, T. (2004). *Nonlinear Time Series Analysis (Second Edition)*. Cambridge University Press, Cambridge, UK.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York.

Marple, S. L. (1987). *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, New Jersey.

Moon, F. C. (1992). *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*. John Wiley and Sons, Inc, New York.

Percival, D. B., Overland, J. E., and Mofjeld, H. O. (2001). Interpretation of North Pacific Variability as a Short- and Long-Memory Process. *Journal of Climate*, 14(24):4545–4559.

Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.* Cambridge University Press, Cambridge, UK.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

Riedel, K. S. and Sidorenko, A. (1995). Minimum biased multitaper spectral estimation. *IEEE Transactions on Signal Processing*, 43:188–195.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley & Sons, Inc., New York.

Schreiber, T. and Schmitz, A. (2000). Surrogate time series. *Physica D*, 142:346–382.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70:1055–1096.

Welch, P. D. (1967). The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73.

Wood, A. T. A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3:409–432.

Woods, J. W. (1972). Two-dimentional discrete Markovian fields. *IEEE Transactions on Information Theory*, 18(2):232–240.