Instantaneous Frequency Estimation Using Stochastic Calculus and Bootstrapping

A. Abutaleb

Systems and Bioengineering Department, School of Engineering, Cairo University, 12613 Giza, Egypt
Email: ataleb@mcit.gov.eg

Stochastic calculus methods are used to estimate the instantaneous frequency of a signal. The frequency is modeled as a polynomial in time. It is assumed that the phase has a Brownian-motion component. Using stochastic calculus, one is able to develop a stochastic differential equation that relates the observations to instantaneous frequency. Pseudo-maximum likelihood estimates are obtained through Girsanov theory and the Radon-Nikodym derivative. Bootstrapping is used to find the bias and the confidence interval of the estimates of the instantaneous frequency. An approximate expression for the Cramér-Rao lower bound is derived. An example is given, and a comparison to existing methods is provided.

Keywords and phrases: bootstrapping, Ito calculus, instantaneous frequency, time-varying frequency, Girsanov theory.

1. INTRODUCTION

Over the past several years there has been a marked increase in the application of coherent signal and image processing. Coherent processing requires an accurate estimate of the phase (Blackledge [1]). Examples where coherent processing is required include synthetic aperture radar (SAR), synthetic aperture sonar, adaptive beam forming, acoustic imaging, projection and diffraction tomography, adaptive optics, magnetic resonance imaging (MRI), and inverse synthetic aperture radar (ISAR) (see Goldstein et al. [2], Song et al. [3], Ghiglia and Pritt [4]). Phase estimation also has many applications in the areas of radar (Wehner [5]) and communications (Li [6]) and others. The instantaneous frequency (IF) is defined as the derivative of the phase.

There are mainly two approaches for the estimation of the IF of the time-varying frequency signals; parametric and nonparametric (Chen and Ling [7]). In the nonparametric approach, if the signal has frequency contents that are rapidly changing over time, one of the most popular approaches is to use short-time Fourier transform (STFT). In this approach, the signal is divided into overlapping segments. The fast Fourier transform (FFT) is then applied to each segment. By observing the evolution of the frequency components over time, one should be able to estimate both the frequency components and their magnitudes (Wehner [5]). This approach suffers from the poor resolution of the FFT, and the large amount of needed data to get reasonable results. Nevertheless, the STFT gives us some idea about the complexity of the data and suggests probable models for the IF.

The Wigner-Ville distribution has proven to be the most suitable nonparametric approach to handle instantaneous frequency estimation (IEEE [8]). Several modifications have been suggested, that concentrate on the choice of the window (see Katkovnik and Stankovic [9], Barkat [10]). Still, a major requirement of the Wigner-Ville distribution is the large size of the data, a situation that is not available in some applications.

In the parametric approach, the phase is modeled as an autoregressive (AR) process, moving average (MA) process, or a polynomial (Benidir and Ouldali [11]). The coefficients of the model are estimated adaptively using the least mean square algorithm (LMS), the recursive least square (RLS) algorithm, phase-locked loop (PLL), or others such as hidden Markov models (see Goto et al. [12], IEEE [8]). For large data sizes, however, the accuracy of the parametric methods was shown to be less than the nonparametric methods. Obviously if the parametric model is accurate, then it should outperform any nonparametric method. The parametric methods on the other hand work better for small-size data (Boashash [13]).

In many applications in communications, the signal could be modeled as higher-order polynomial phase signals with constant amplitudes. Conventional power spectrum methods are ineffective in handling these kinds of signals. For example, the Wigner-Ville distribution (WVD) can handle polynomial phases with a maximum order of two. The polynomial phase transform was introduced to tackle this problem (see Peleg and Porat [14], Peleg and Friedlander [15], Porat and Friedlander [16]). Another suboptimal

algorithm was introduced to estimate the polynomial phase (Golden and Friedlander [17]). The major problems with these two algorithms are the required large number of data, and the needed, relatively, high SNR (30 dB and above) to obtain good estimates.

In this paper, we introduce a new parametric approach that is based on the stochastic (Ito) calculus. Concepts from stochastic or Ito calculus (Kloeden and Platen [18]) are used to set up the phase estimation problem. Assuming that the phase could be modeled as a polynomial in time with a scaled additive Brownian motion, which is a common model for random media transmission, one could develop a stochastic differential equation (SDE) for the observations. Other models for the phase or the amplitude could be tackled with the same approach. For this model of the phase, the SDE is a linear equation in the unknown coefficients of the polynomial phase but nonlinear in the other parameters. This way a nonlinear estimation problem has been transformed into a partially linear estimation problem. Using the Girsanov theory, and assuming complete knowledge of the other parameters of the signal, one is able to find an exact expression for the likelihood function of the observations. This is actually a pseudo-likelihood function. Maximizing this pseudo-likelihood function with respect to the unknown coefficients of the phase, one is able to find a closed form solution for these coefficients. Simulation-based methods (McFadden [19]) are used to estimate the other parameters of the signal.

Moreover, approximate statistical properties of the pseudo-maximum likelihood estimates of the coefficients of the phase were obtained in a straightforward fashion. Approximate expressions for the Cramér-Rao lower bounds for the variances of the estimates were easily obtained; another advantage of using Ito calculus. While many estimation problems could be cast in a discrete time setting, there is no equivalent in the discrete time to Ito's lemma; a fundamental result in stochastic calculus. Thus, one has to use continuous time models in order to take advantage of the Ito calculus properties. "It should be added that in some cases it is easier, at first, to study the continuous analog of problems formulated for discrete time, and use the results obtained in the solution of the latter" (Lipster and Shiryayev [20, Volume I]). In some cases, however, using stochastic calculus would be cumbersome and the discrete methods of time series analysis would be much easier to use.

In the given examples, the required SNR to get good estimates of the phase is more than 10 dB. Using the bootstrapping method, one was able to obtain good results at lower SNR. Bootstrap is a method that handles small samples and yields good estimates for the bias and the confidence intervals for the estimates (see Efron and Tibshirani [21], Zaman [22], Politis [23], Zoubir and Boashash [24]). It was not meant to be a method to improve the accuracy of the estimates. Using the peak of the histogram, generated from the bootstrapping method for each estimated parameter, as the desired estimate of the unknown parameter, one is able to get good estimates at lower SNR. This was observed in this report and in some other applications (see Souza and Neto [25], Abutaleb [26]). Through the bootstrapping method, one is

able to generate a large number of samples from the measured small sample data. Bootstrapping amounts to treating the observed samples as if they exactly represent the whole population. In nonparametric bootstrapping, drawing from this sample space, at random with replacement, will generate as many samples as desired. In parametric bootstrapping, a model is used to generate as many samples as desired. The bootstrapping method found many applications in the area of statistical signal processing (Zoubir and Boashash [24]).

This paper is divided as follows: Section 2 describes the mathematical formulation of the phase estimation problem and an approximate maximum likelihood method is described. We also describe in this section the method of stochastic annealing. In Section 3, the proposed stochastic-calculus-based approach is introduced and the statistical properties of the estimates are given. A discussion of the bootstrapping method is also given. In Section 4, we present simulation results and comparison to the approximate maximum likelihood method of Section 2. In Section 5, summary and conclusions are given.

2. PROBLEM FORMULATION

The real signal, $z(t)$, with one phase component could be modeled as follows:

$$z(t) = A(t) \sin [\phi(t)] + \varepsilon(t), \tag{1}$$

where

$$\phi(t) = 2\pi \sum_{i=1}^{I} f_i t^i + \gamma B(t), \tag{2}$$

I is the order of the polynomial, $A(t)$ is the amplitude, $\phi(t)$ is the phase, $B(t)$ is a Brownian motion, γ is an unknown constant, and $\varepsilon(t)$ is an additive white noise process of unknown variance σ^2. The one phase component could represent the radar tracking of just one target, that is, maneuvering. It could also represent the tracking of one target in a sequence of images.

For constant phase component and polynomial frequency we have

$$\phi(t) = 2\pi \sum_{i=0}^{I} f_i t^i + \gamma B(t). \tag{3}$$

In this paper we will be concerned with the situation where $A(t) = 1$, and the error/noise term $\varepsilon(t)$ is imbedded in the random part of the phase, that is, we are interested in the model

$$z(t) = \sin [\phi(t)]. \tag{4}$$

The case where the amplitude is not equal to 1 could be treated in the same context. We basically estimate the amplitude first by taking, for example, the average of the max and the min of the signal. Other models for the phase

such as the Ornstein-Uhlenbeck process, and other models for the amplitude could also be tackled with the proposed stochastic-calculus-based approach. This will be the subject of other reports (Abutaleb [27]). The case where we have more than one stochastic IF is currently under investigation using what is called the Malliavin calculus (see Abutaleb and Papaioannou [28], Nualart [29]).

2.1. Approximate maximum likelihood method

To estimate the unknown coefficients, f_i, one could ignore the Brownian motion term, $\gamma B(t)$, and use nonlinear least square methods (Goto et al. [12]), the genetic algorithms (Abutaleb [30]), or stochastic annealing (Kloeden and Platen [18]). Specifically, in the least square approach, the instantaneous phase $\phi_{LS}(t)$ is assumed to be pure polynomial in the frequency, that is,

$$z(t) = \sin\left[\phi_{LS}(t)\right] + \varepsilon(t), \tag{5}$$

where

$$\phi_{LS}(t) = 2\pi \sum_{i=1}^{I} f_i t^i. \tag{6}$$

One could then find the unknowns by minimizing the sum of squared error, $\{z(t) - \sin[\phi_{LS}(t)]\}^2$. While this approach is approximate, it gives an indication about the estimates of the unknown parameters that could be used later on to guide the search area with numerical methods. The above approach is a simplification to the true problem since the correct expression should be

$$z(t) = \sin\left[\phi(t)\right] = \sin\left[\phi_{LS}(t)\right] \cos\left[\gamma B(t)\right] \\ + \cos\left[\phi_{LS}(t)\right] \sin\left[\gamma B(t)\right], \tag{7}$$

and when $\gamma \to 0$, we get the approximate expression

$$z(t) = \sin\left[\phi(t)\right] \approx \sin\left[\phi_{LS}(t)\right] + \cos\left[\phi_{LS}(t)\right]\left[\gamma B(t)\right], \tag{8}$$

that is,

$$z(t) \approx \sin\left[\phi_{LS}(t)\right] + \varepsilon(t), \tag{9}$$

where $\varepsilon(t) \approx \cos[\phi_{LS}(t)][\gamma B(t)]$, that is, the error is function of the unknown parameters.

In discrete time, one could write (8) as

$$z(k) = \sin\left[\phi(k)\right] \approx \sin\left[\phi_{LS}(k)\right] + \cos\left[\phi_{LS}(k)\right]\left[\gamma\sqrt{\Delta} \sum_{j=0}^{k} w(j)\right], \tag{10}$$

where $\phi_{LS}(k) = 2\pi \sum_{i=1}^{I} f_i(k\Delta)^i$, $B(t) = \sqrt{\Delta} \sum_{j=0}^{k} w(j)$, Δ is the sampling interval, and $w(j)$ is Gaussian with zero mean and variance 1.

Rearranging (10) we get

$$\frac{\{z(k) - \sin\left[\phi_{LS}(k)\right]\}}{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k)\right]} \approx \left[\sum_{j=0}^{k} w(j)\right], \tag{11}$$

and for time $(k+1)$ we get

$$\frac{\{z(k+1) - \sin\left[\phi_{LS}(k+1)\right]\}}{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k+1)\right]} \approx \left[\sum_{j=0}^{k+1} w(j)\right]. \tag{12}$$

Subtracting the expressions at k and at $(k+1)$ we get

$$-\frac{\{z(k) - \sin\left[\phi_{LS}(k)\right]\}}{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k)\right]} + \frac{\{z(k+1) - \sin\left[\phi_{LS}(k+1)\right]\}}{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k+1)\right]} \\ \approx w(k+1). \tag{13}$$

Rearranging (13) we get

$$z(k+1) = \sin\left[\phi_{LS}(k+1)\right] \\ + \frac{\{z(k) - \sin\left[\phi_{LS}(k)\right]\}}{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k)\right]} \{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k+1)\right]\} \\ + \{\gamma\sqrt{\Delta} \cos\left[\phi_{LS}(k+1)\right]\} w(k+1). \tag{14}$$

The conditional probability density function (pdf) $f[z(k+1)|z(k)]$ is thus given by

$$f\left[z(k+1)|z(k)\right]$$

$$= \frac{\exp-(1/2)\{z(k+1) - \sin[\phi_{LS}(k+1)] - (\{z(k) - \sin[\phi_{LS}(k)]\}/\gamma\sqrt{\Delta}\cos[\phi_{LS}(k)])\{\gamma\sqrt{\Delta}\cos[\phi_{LS}(k+1)]\}\}^2/\gamma^2\Delta\cos^2[\phi_{LS}(k+1)]}{\sqrt{2\pi\gamma^2\Delta\cos^2\left[\phi_{LS}(k+1)\right]}}. \tag{15}$$

The likelihood function, which is the joint pdf of the observations, is thus given by

$$f[z(N), z(N-1), \ldots, z(1)] = f[z(1)] \prod_{k=1}^{N-1} f[z(k+1)|z(k)], \quad (16)$$

and the log likelihood function λ, ignoring the initial conditions, is given by

$$\lambda = \log\{f[z(N), z(N-1), \ldots, z(1)]\}$$
$$= \sum_{k=1}^{N-1} \log\{f[z(k+1)|z(k)]\}. \quad (17)$$

Substituting for the expression in (15) into (17), we get

$$\lambda = \log\{f[z(N), z(N-1), \ldots]\}$$
$$= -\frac{1}{2} \sum_{k=1}^{N-1} \log\{2\pi\gamma^2 \Delta \cos^2[\phi_{LS}(k+1)]\}$$
$$- \frac{1}{2} \sum_{k=1}^{N-1} \frac{\{z(k+1) - \sin[\phi_{LS}(k+1)] - (\{z(k) - \sin[\phi_{LS}(k)]\}/\gamma\sqrt{\Delta}\cos[\phi_{LS}(k)])\{\gamma\sqrt{\Delta}\cos[\phi_{LS}(k+1)]\}\}^2}{\gamma^2 \Delta \cos^2[\phi_{LS}(k+1)]}. \quad (18)$$

Maximizing the log likelihood function of (18) with respect to the unknown coefficients, f_i and γ, yields the desired estimates.

Due to the nonlinearities in the log likelihood function, one would expect the existence of several local minima and we should use global optimization methods. Thus, it is proposed that the frequencies and γ be estimated through the stochastic annealing method, which is known to have good numerical properties. Other methods such as genetic algorithms could also be used.

2.2. Parameter estimation through stochastic annealing

In the stochastic annealing method, for each unknown parameter we develop a stochastic equation. The solution of which is the maximum likelihood estimate of the parameter. This is explained in what is to follow.

The stochastic differential equation (SDE) of the unknown parameters is given by

$$d\hat{\underline{f}}(t) = -\underline{\nabla}\lambda[\hat{\underline{f}}(t), \hat{\gamma}(t)] + \sigma(t)d\underline{B}_f(t) \quad (19)$$

which has the discrete form

$$\hat{\underline{f}}(k+1) = \hat{\underline{f}}(k) - \underline{\nabla}\lambda[\hat{\underline{f}}(k), \hat{\gamma}(k)] + \sigma(k)\underline{\Delta B}_f(k), \quad (20)$$

where $\underline{B}_f(t)$ is a vector Brownian motion, $\underline{\Delta B}_f(k)$ is a vector white Gaussian noise. Each element has zero mean and variance Δ, Δ is the sampling interval. $\underline{\nabla}\lambda(\hat{\underline{f}}(k), \hat{\gamma}(k))$ is the gradient of $\lambda(\hat{\underline{f}}, \hat{\gamma})$ with respect to $\hat{\underline{f}}$.

A similar equation for γ is used:

$$d\hat{\gamma}(t) = -\nabla\lambda[\hat{\underline{f}}(t), \hat{\gamma}(t)] + \sigma(t)dB_\gamma(t), \quad (21)$$

where $B_\gamma(t)$ is Brownian motion, and $\nabla\lambda[\hat{\underline{f}}(t), \hat{\gamma}(t)]$ is the gradient of $\lambda(\underline{f}, \gamma)$ with respect to γ, and is defined as

$$\nabla\lambda[\hat{\gamma}(k)] = \frac{\lambda[\hat{\underline{f}}(k), \hat{\gamma}(k)] - \lambda[\hat{\underline{f}}(k), \hat{\gamma}(k-1)]}{\hat{\gamma}(k) - \hat{\gamma}(k-1)}. \quad (22)$$

The variance is defined as

$$\sigma(t) = \frac{c}{\sqrt{\log(t+2)}}, \quad c > 0. \quad (23)$$

The choice above for $\sigma(t)$ is shown to give quick convergence (Kloeden and Platen [18]). Other choices for $\sigma(t)$ are also possible.

2.3. The Wigner-Ville distribution

The most common nonparametric method to estimate the instantaneous frequency is through the Wigner-Ville distribution (WVD) and its modifications. The WVD(t, ω) at instant t and frequency ω of a signal $z(t)$ and its complex conjugate $z^*(t)$ is defined as (Chen and Ling [7])

$$\text{WVD}(t, \omega) = \int dt' z\left(t + \frac{t'}{2}\right) z^*\left(t - \frac{t'}{2}\right) \exp(-j\omega t') \quad (24)$$

which has the discrete form (Boashash [13])

$$\text{WVD}(n, k) = \sum_{m=-M}^{M} z\left(n + \frac{m}{2}\right) z^*\left(n - \frac{m}{2}\right) \exp\left(-j2\pi m \frac{k}{M}\right), \quad (25)$$

where $M = (N-1)/2$, and N is the number of data points.

For a sampling interval Δ, the discrete WVD is given as (Boashash and Black [31])

$$
\begin{aligned}
&\mathrm{WVD}(n\Delta, f) \\
&= 2\Delta \sum_{m=-M}^{M} z(nT + m\Delta) z^*(n - m\Delta) \exp(-j4\pi f m\Delta),
\end{aligned}
\tag{26}
$$

where $M = (N - 1)/2$, and N is the number of data points. Both formulae (25) and (26) are used in the literature.

The instantaneous frequency $f(n)$ derived via the first moment of the discrete WVD is given by

$$
f(n) = \frac{M}{2\pi} \sum_{k=0}^{M-1} \mathrm{WVD}(n, k) \exp\left(j2\pi \frac{k}{M}\right) \bmod 2\pi.
\tag{27}
$$

In Section 4, the instantaneous frequency of (27) is compared to the true values. It is shown that they are far apart. Thus, the WVD approach was not used for further comparisons.

2.4. The Hilbert transform and the analytic function

A given real signal $z(t)$ can be used to construct a complex waveform, $z_a(t)$, as

$$
z_a(t) = \frac{1}{\sqrt{2}} [z(t) + j\hat{z}(t)],
\tag{28}
$$

where $\hat{z}(t)$ is the Hilbert transform of $z(t)$ and is given as (Cohen [32])

$$
\hat{z}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{z(\tau)}{t - \tau} d\tau.
\tag{29}
$$

The complex waveform, $z_a(t)$, is the analytic representation of $z(t)$ and could be calculated either in the time domain using a finite impulse response (FIR) filter or in the frequency domain as follows (Boashash [33]).

(1) Take the N-point discrete Fourier transform of $z(n)$ to get $\mathrm{DFT}[z(n)]$, $n = 0, \ldots, N - 1$.
(2) Define

$$
\mathrm{DFT}[z_a(k)] = \begin{cases} \mathrm{DFT}[z(k)], & k = 0, \dfrac{N}{2}, \\[2mm] 2\,\mathrm{DFT}[z(k)], & k = 1, \ldots, \dfrac{N}{2} - 1, \\[2mm] 0, & k = \dfrac{N}{2} + 1, \ldots, N - 1. \end{cases}
\tag{30}
$$

(3) Calculate $z_a(n) = \mathrm{IDFT}\{\mathrm{DFT}[z_a(k)]\}$, $n = 0, \ldots, N - 1$, where IDFT is the inverse Fourier transform.

The phase angle is then calculated and fitted to the assumed model of the phase. The phase angle, ϕ_{Analytic}, is defined as

$$
\phi_{\mathrm{Analytic}} = \arctan\left\{ \frac{\mathrm{Imaginary}[z_a(t)]}{\mathrm{Real}[z_a(t)]} \right\}.
\tag{31}
$$

In Section 4, the analytic phase, ϕ_{Analytic} of (31), is estimated and compared to the true value. It is shown that they are far apart. Thus, the analytic signal approach was not used.

3. ITO CALCULUS FOR PHASE ESTIMATION: THE PROPOSED APPROACH

In this section, we introduce the stochastic-calculus-based approach for phase estimation. It is assumed that the randomness in the signal is due to the Brownian-motion component in the phase. Other models of the signal could also be treated in the same way.

3.1. Derivation of the SDE for the observation

As in Section 2, the received signal is modeled as

$$
z(t) = \sin\left(2\pi \sum_{i=1}^{I} f_i t^i + \gamma B(t)\right),
\tag{32}
$$

where it is assumed that the additive error is imbedded in the random component (Wiener process or Brownian motion) $B(t)$.

Let $x(t) = B(t)$, thus

$$
dx(t) = dB(t),
\tag{33}
$$

where $dx(t)$ is the stochastic differential for $x(t)$, and $dB(t)$ is what is known as white Gaussian noise with zero mean and variance dt. We need to find a stochastic differential equation (SDE) for $dz(t)$. In order to do that, we first introduce the general form for SDE and for nonlinear transformation.

The general form for a stochastic differential equation is

$$
dx(t) = a(t, x(t))dt + b(t, x(t))dB(t).
\tag{34}
$$

In our case, (33), $a(t, x(t)) = 0$, and $b(t, x(t)) = 1$.

Let $U_1(t, x(t))$ be a nonlinear transformation of $x(t)$, that is, $z(t) = U_1(t, x(t))$, using Ito lemma (Kloeden and Platen [18]) we obtain

$$
\begin{aligned}
dz(t) = &\left[\frac{\partial U_1}{\partial t} + a(t, x(t)) \frac{\partial U_1}{\partial x} + \frac{1}{2} b^2(t, x(t)) \frac{\partial^2 U_1}{\partial x^2} \right] dt \\
&+ b(t, x(t)) \frac{\partial U_1}{\partial x} dB(t).
\end{aligned}
\tag{35}
$$

In our case,

$$z(t) = U_1(t, x(t)) = \sin\left(2\pi \sum_{i=1}^{I} f_i t^i + \gamma x(t)\right), \quad (36)$$

$$x(t) = B(t), \qquad a(t, x(t)) = 0, \qquad b(t, x(t)) = 1,$$

$$\frac{\partial U_1}{\partial t} = 2\pi \left[\sum_{i=1}^{I-1}(it^{i-1} f_i)\right]\cos\left(2\pi \sum_{i=1}^{I} f_i t^i + \gamma x(t)\right),$$

$$\frac{\partial U_1}{\partial x} = \gamma \cos\left(2\pi \sum_{i=1}^{I} f_i t^i + \gamma x(t)\right), \quad (37)$$

$$\frac{\partial^2 U_1}{\partial x^2} = -\gamma^2 \sin\left(2\pi \sum_{i=1}^{I} f_i t^i + \gamma x(t)\right).$$

Substituting the expressions of (37) into (35) we get

$$dz(t) = \left\{2\pi \sum_{i=1}^{I}(it^{i-1} f_i)\sqrt{(1 - z^2(t))} - \frac{1}{2}\gamma^2 z(t)\right\}dt$$
$$+ \gamma\sqrt{(1 - z^2(t))}\,dB(t), \quad (38)$$

where we chose $\cos x = \sqrt{(1 - \sin^2 x)}$. Nevertheless, one has to choose the right sign of the $\cos(x)$ depending on which quadrant lies x.

Let

$$\underline{f} = \begin{bmatrix} f_1 & f_2 & \cdots & f_I \end{bmatrix}^T, \quad (39)$$

and let

$$\underline{a}(t) = 2\pi \begin{bmatrix} 1 & 2t & \cdots & It^{I-1} \end{bmatrix}^T. \quad (40)$$

Then (38) could be written compactly as

$$dz(t) = \left\{\underline{a}^T(t)\underline{f}\sqrt{(1 - z^2(t))} - \frac{1}{2}\gamma^2 z(t)\right\}dt$$
$$+ \gamma\sqrt{(1 - z^2(t))}\,dB(t). \quad (41)$$

The stochastic process $z(t)$ has the following SDE form:

$$dz(t) = c(t, z(t))dt + e(t, z(t))dB(t), \quad (42)$$

where

$$c(t, z(t)) = \left\{\underline{a}^T(t)\underline{f}\sqrt{(1 - z^2(t))} - \frac{1}{2}\gamma^2 z(t)\right\},$$
$$e(t, z(t)) = \gamma\sqrt{(1 - z^2(t))}. \quad (43)$$

Notice that the SDE of $z(t)$ turned out to be linear in the unknown frequency components, $\underline{f}$. This is one advantage of using Ito calculus. Other forms of the phase may not yield this nice linear property however. One could stop here and derive an approximate expression for the likelihood function of the observations as we did in Section 2. Instead we go through the rigorous derivation of the exact likelihood function.

3.2. Pseudo-maximum likelihood estimation

3.2.1. The pseudo-maximum likelihood estimates of the unknown parameters

Let the unknown parameters be defined as the vector $\underline{\theta} = [\underline{f}^T \gamma]^T$, and then the SDE of the observations, $dz(t)$, is rewritten as

$$dz(t) = c(t, z(t); \underline{\theta})dt + e(t, z(t); \gamma)dB(t). \quad (44)$$

According to the Girsanov theory and for a known γ, the Radon-Nikodym derivative which is the likelihood function, L, is given by (see Lipster and Shiryayev [20], Oksendal [34], Prakasa Rao [35])

$$L(t, z(t); \underline{\theta})$$
$$= \exp\left\{\frac{-1}{2}\int_0^t \frac{c^2[s, z(s); \underline{\theta}]ds}{e^2[s, z(s); \gamma]} + \int_0^t \frac{c[s, z(s); \underline{\theta}]dz(s)}{e^2[s, z(s); \gamma]}\right\}, \quad (45)$$

and the log likelihood function $\lambda = \ln L$ becomes

$$\lambda = \ln L(t, z(t); \underline{\theta})$$
$$= \frac{-1}{2}\int_0^t \frac{c^2[s, z(s); \underline{\theta}]ds}{e^2[s, z(s); \gamma]} + \int_0^t \frac{c[s, z(s); \underline{\theta}]dz(s)}{e^2[s, z(s); \gamma]}. \quad (46)$$

Substituting the expressions for $c[t, z(t); \underline{\theta}]$ and $e[t, z(t); \gamma]$ into (46), we get the log likelihood function as

$$\lambda = \ln L(t, z(t); \underline{\theta})$$
$$= \frac{-1}{2}\int_0^t \frac{\left\{\underline{a}^T(s)\underline{f}\sqrt{(1 - z^2(s))} - (1/2)\gamma^2 z(s)\right\}^2 ds}{\gamma^2[1 - z^2(s)]}$$
$$+ \int_0^t \frac{\left\{\underline{a}^T(s)\underline{f}\sqrt{(1 - z^2(s))} - (1/2)\gamma^2 z(s)\right\}dz(s)}{\gamma^2[1 - z^2(s)]}. \quad (47)$$

Equation (45) is a true likelihood function as long as γ is known. And since γ will be replaced by its estimate $\hat{\gamma}$, (45) represents a pseudo-likelihood function. Maximizing the log of the pseudo-likelihood function $\lambda = \ln L$ with respect to the unknown vector $\underline{f}$ and assuming estimate $\hat{\gamma}$, one could get the pseudo-maximum likelihood estimate of $\underline{f}$ as follows:

$$\frac{\partial \lambda}{\partial \underline{f}} = \underline{0}$$

$$= -\int_0^t \frac{\left\{\sqrt{(1 - z^2(s))}\underline{a}(s)\right\}\left\{\underline{a}^T(s)\hat{\underline{f}}\sqrt{(1 - z^2(s))} - (1/2)\hat{\gamma}^2 z(s)\right\}ds}{\hat{\gamma}^2[1 - z^2(s)]}$$

$$+ \int_0^t \frac{\left\{\sqrt{(1 - z^2(s))}\underline{a}(s)\right\}dz(s)}{\hat{\gamma}^2[1 - z^2(s)]}. \quad (48)$$

This is reduced to

$$\frac{1}{\hat{\gamma}^2}\left\{\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right\}\hat{\underline{f}}$$
$$= \int_0^t \frac{\underline{a}(s)dz(s)}{\hat{\gamma}^2\left[\sqrt{(1-z^2(s))}\right]} + \frac{1}{2}\int_0^t \frac{\underline{a}(s)z(s)ds}{\left[\sqrt{(1-z^2(s))}\right]}. \tag{49}$$

Finally the pseudo-maximum likelihood estimate of $\underline{f}$ is given as

$$\hat{\underline{f}} = \left\{\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right\}^{-1}\int_0^t \frac{\underline{a}(s)dz(s)}{\left[\sqrt{(1-z^2(s))}\right]}$$
$$+ \frac{\hat{\gamma}^2}{2}\left\{\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right\}^{-1}\int_0^t \frac{\underline{a}(s)z(s)ds}{\left[\sqrt{(1-z^2(s))}\right]}. \tag{50a}$$

Substituting for $dz(s)$ in (50) we get

$$\hat{\underline{f}} = \underline{f} + \frac{\hat{\gamma}\int_0^t \underline{a}(s)dB(s)}{\left\{\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right\}^{-1}}. \tag{50b}$$

To estimate the unknown parameter γ, one has to go through another route. This is explained in Section 3.2.3.

3.2.2. Another derivation for the pseudo-maximum likelihood estimates

In the previous subsection, the pseudo-maximum likelihood estimates were obtained using the observed data sequence, $z(t)$. If one is able to find an initial guess for γ, using, for example, the approximated method of Section 2, one could apply a transformation to the data and derive an easier method to find the pseudo-maximum likelihood estimate of the vector $\underline{f}$. This approach, however, is tricky. It is always better to work with the original data. The simulations showed the accuracy of the approach presented in Section 2.2 compared to the one presented here. We present the derivation any way because we will need some of its results when calculating the Cramér-Rao lower bound.

To find the pseudo-maximum likelihood estimates of the unknown parameters, it will be easier if the coefficient of the Brownian motion is unity. Thus, we need to find another transformation $y(t) = U_2(t, z(t))$ such that

$$dy(t) = g(t, y(t))dt + dB(t). \tag{51}$$

We pick $U_2(t, z(t))$ such that

$$\frac{\partial U_2}{\partial z} = \frac{1}{e(t,z(t))} = \frac{1}{\gamma\sqrt{(1-z^2(t))}} = \frac{1}{\gamma}\left[1-z^2(t)\right]^{-1/2},$$
$$\frac{\partial U_2}{\partial t} = 0, \tag{52}$$

and thus,

$$\frac{\partial^2 U_2}{\partial z^2} = \left[\frac{1}{\gamma}z(t)\right]\left[1-z^2(t)\right]^{-3/2}. \tag{53}$$

Integrating $\partial U_2/\partial z$ with respect to $z(t)$ we get

$$y(t) = U_2(t, z(t)) = \frac{1}{\gamma}\sin^{-1}(z(t)), \tag{54}$$

that is,

$$z(t) = \sin(\gamma y(t)), \tag{55}$$

where we have chosen to set the constant of integration to zero.

The expression for $g(t, y(t))$ is given as follows:

$$g(t, y(t))|_{z(t)=\sin(\gamma y(t))} = \frac{\partial U_2}{\partial t} + c(t, z(t))\frac{\partial U_2}{\partial z} + \frac{1}{2}e^2(t, z(t))\frac{\partial^2 U_2}{\partial z^2}. \tag{56}$$

Substituting (52) and (53) into (56) we get

$$g(t, z(t))$$
$$= \left\{\underline{a}^T(t)\underline{f}\sqrt{[1-z^2(t)]} - \frac{1}{2}\gamma^2 z(t)\right\}\left\{\frac{1}{\gamma}\left[1-z^2(t)\right]^{-1/2}\right\}$$
$$+ \frac{1}{2}\gamma^2\left[1-z^2(t)\right]\left[\frac{1}{\gamma}z(t)\right]\left[1-z^2(t)\right]^{-3/2} \tag{57}$$

which is reduced to

$$g(t, z(t)) = \frac{1}{\gamma}\left[\underline{a}^T(t)\underline{f}\right]. \tag{58}$$

Substituting (58) into (51), we get the SDE for $dy(t)$ as follows:

$$dy(t) = \frac{1}{\gamma}\left[\underline{a}^T(t)\underline{f}\right]dt + dB(t). \tag{59}$$

This form of $dy(t)$ has the coefficients of $dB(t)$ unity. This will make the derivation of the Cramér-Rao lower bound easier and tractable (Oksendal [34]). Remember that $\underline{f} = [f_1 \quad f_2 \quad \cdots \quad f_I]^T$ and $\underline{a}(t) = 2\pi[1 \quad 2t \quad \cdots \quad It^{I-1}]^T$.

According to the Girsanov theory and for a given γ, the Radon-Nikodym derivative, which is the pseudo-likelihood function, L, is given by (Oksendal [34])

$$L = \exp\left\{\frac{-1}{2}\int_0^t\left[\frac{1}{\gamma}\underline{a}^T(s)\underline{f}\right]^2 ds + \int_0^t\left[\frac{1}{\gamma}\underline{a}^T(s)\underline{f}\right]dy(s)\right\}, \tag{60}$$

and the log of the pseudo-likelihood function $\lambda = \log L$ is given as

$$\lambda = \log L = \frac{-1}{2} \int_0^t \left[\frac{1}{\gamma} \underline{a}^T(s) \underline{f} \right]^2 ds + \int_0^t \left[\frac{1}{\gamma} \underline{a}^T(s) \underline{f} \right] dy(s). \tag{61}$$

Maximizing the log of the pseudo-likelihood function $\lambda = \log L$ with respect to the unknowns $\underline{f}$, one could get the pseudo-maximum likelihood estimate of the unknowns. Remember that in this expression, however, we use the sequence of transformed data $y(t)$, not the observed data $z(t)$. Thus, an initial guess for γ must be available.

3.2.3. Simulation-based estimation of γ

Equation (45) is a true likelihood function as long as the diffusion term, $e(t, z(t); \gamma)$, is completely known, that is, γ is known. In practice, we do not know γ. As a matter of fact it is one of the unknowns that we need to estimate. There are many methods to estimate γ. A popular, yet not accurate, method to estimate γ (Yoshida [36]) is through the observation that

$$[dz(t)]^2 = \gamma^2 [1 - z^2(t)] dt. \tag{62}$$

One could use this equation or its modifications to find an estimate for γ. Instead we use the more accurate simulation-based method of McFadden [19] to find $\hat{\gamma}$. The following steps describe the algorithm. (1) For a given estimate $\hat{\underline{f}}$, we simulate a sequence of data points through (4) and we calculate the sum of the squared differences between the simulated and the observed data; this is the error criterion. (2) We change γ; using the method of stochastic annealing of Section 2.2, till the sum of squared error is minimized. (3) Use the maximum likelihood method of Section 3.1 with the estimated $\hat{\gamma}$ to find a new estimate for $\hat{\underline{f}}$. (4) Go to step (1). (5) Stop when there is no more reduction in the sum of the squared error. Specifically for an estimated $\hat{\underline{f}}$, and for a given $\hat{\gamma}$, simulate the phase $\hat{\phi}(t)$ given by

$$\hat{\phi}(t) = 2\pi \sum_{i=0}^{I} \hat{f}_i t^i + \hat{\gamma} B(t). \tag{63}$$

Thus, the simulated signal $\hat{z}(t)$ is given by

$$\hat{z}(t) = \sin \left[\hat{\phi}(t) \right]. \tag{64}$$

The error criterion is defined as

$$\text{optimization criterion for } \gamma = \sum_t \left[z(t) - \hat{z}(t) \right]^2. \tag{65}$$

3.2.4. Pseudo-maximum likelihood estimate for $\underline{f}$

The pseudo-maximum likelihood estimates of the coefficients of the polynomial phase are obtained as

$$\lambda = \log L = \frac{-1}{2} \int_0^t \left[\frac{1}{\gamma} \underline{a}^T(s) \underline{f} \right]^2 ds + \int_0^t \left[\frac{1}{\gamma} \underline{a}^T(s) \underline{f} \right] dy(s),$$

$$\frac{\partial \lambda}{\partial \underline{f}} \Big|_{\gamma = \hat{\gamma}, \underline{f} = \hat{\underline{f}}} = \underline{0} = -\int_0^t \frac{1}{\gamma} \underline{a}(s) \left[\frac{1}{\gamma} \underline{a}^T(s) \underline{f} \right] ds + \frac{1}{\gamma} \int_0^t \underline{a}(s) dy(s). \tag{66}$$

Rearranging we get

$$\left[\int_0^t \underline{a}(s) \underline{a}^T(s) ds \right] \hat{\underline{f}} = \hat{\gamma} \int_0^t \underline{a}(s) dy(s), \tag{67}$$

that is,

$$\hat{\underline{f}} = \hat{\gamma} \left[\int_0^t \underline{a}(s) \underline{a}^T(s) ds \right]^{-1} \int_0^t \underline{a}(s) dy(s) \tag{68}$$

which is a closed form solution for estimates of the coefficients of the polynomial phase. Remember that $\underline{a}(t) = 2\pi [1 \ (2t) \ \cdots \ (It^{I-1})]^T$, $z(t) = \sin[\gamma y(t)]$, and it is $z(t)$ that we observe. Thus, in (68) we substitute for $y(t) = \{\sin^{-1}[z(t)]\}/\hat{\gamma}$.

While (68) seems to be easier to evaluate than (50), we recommend not using it due to the discontinuities that might appear in calculating $y(t) = \{\sin^{-1}[z(t)]\}/\hat{\gamma}$. The source of the discontinuities is the expression $\sin^{-1}[z(t)]$.

3.3. Statistical properties of the frequency estimates and Cramér-Rao lower bound

To find the expected value, the bias, and the Cramér-Rao lower bound on the variance of the estimates $\hat{\underline{f}}$, we assume that γ is known and deterministic. This way a closed form expression could be found. Otherwise the expressions would be complicated and no insight would be gained.

3.3.1. The mean of $\hat{\underline{f}}$

The following is a derivation of an approximate expression for the mean of $\hat{\underline{f}}$. Substitute for the expression of $dy(s)$, of (59), into (68). This yields

$$\hat{\underline{f}} = \left[\int_0^t \underline{a}(s) \underline{a}^T(s) ds \right]^{-1} \left\{ \left[\int_0^t \underline{a}(s) \underline{a}^T(s) ds \right] \underline{f} + \gamma \int_0^t \underline{a}(s) dB(s) \right\}, \tag{69}$$

that is,

$$\hat{\underline{f}}(t) = \underline{f} + \gamma A^{-1}(t) \int_0^t \underline{a}(s) dB(s), \tag{70}$$

where $A(t) = [\int_0^t \underline{a}(s)\underline{a}^T(s)ds]$ and as before $\underline{a}(t) = 2\pi[1 \ (2t) \ \cdots \ (It^{I-1})]^T$. Notice that we got the same expression for $\hat{\underline{f}}(t)$ in (50b).

Taking the expected value of both sides of (70), and keeping in mind that under mild conditions $E\{\int_0^t \underline{a}(s)dB(s)\} = \underline{0}$, we get

$$E\{\hat{\underline{f}}\} = \underline{f}, \tag{71}$$

that is, the estimate is unbiased. This is true given that the value of γ is known and deterministic. In reality, one has to estimate γ. In this case, the asymptotic properties of $\hat{\underline{f}}$ will be complex to derive and one might end up with a biased estimate.

### 3.3.2.	The variance of $\hat{\underline{f}}$

The following is a derivation of an approximate expression for the variance of $\hat{\underline{f}}$. Since

$$\hat{\underline{f}}(t) = \underline{f} + \gamma A^{-1}(t) \int_0^t \underline{a}(s)dB(s), \tag{72}$$

then the variance of $\hat{\underline{f}}$ is given as

$$E\left\{[\hat{\underline{f}}(t) - \underline{f}]\left[[\hat{\underline{f}}(t) - \underline{f}]^T\right]\right\}$$
$$= \gamma^2 E\left\{\left[A^{-1}(t)\int_0^t \underline{a}(s)dB(s)\right]\left[\int_0^t \underline{a}^T(s)dB(s)A^{-T}(t)\right]\right\}, \tag{73}$$

where $A^{-T} = (A^{-1})^T$. The variance is reduced to

$$E\left\{[\hat{\underline{f}}(t) - \underline{f}]\left[[\hat{\underline{f}}(t) - \underline{f}]^T\right]\right\}$$
$$= \gamma^2 A^{-1}(t)E\left\{\int_0^t \int_0^t \underline{a}(s)\underline{a}^T(u)dB(s)dB(u)\right\}A^{-T}(t), \tag{74}$$

and since $dB(s)dB(u) = ds\delta(s-u)$, then

$$E\left\{[\hat{\underline{f}}(t) - \underline{f}]\left[[\hat{\underline{f}}(t) - \underline{f}]^T\right]\right\}$$
$$= \gamma^2 A^{-1}(t)\left[\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right]A^{-T}(t) \tag{75}$$
$$= \gamma^2 A^{-1}(t)A(t)A^{-T}(t),$$

that is,

$$E\left\{[\hat{\underline{f}}(t) - \underline{f}]\left[[\hat{\underline{f}}(t) - \underline{f}]^T\right]\right\} = \gamma^2 A^{-T}(t). \tag{76}$$

This is true given that the value of γ is known and deterministic. In reality, one has to estimate γ. In this case, the variance of $\hat{\underline{f}}$ will be complex to derive.

### 3.3.3.	An approximate Cramér-Rao lower bound on the variance of the frequency estimates

It is always useful to compare the variance of the estimates to the best possible minimum variance. This is the Cramér-Rao lower bound which is derived in the next part. Again, one is able to get a closed form expression given that γ is known and deterministic. Otherwise the expressions would be much more complicated and no closed form solutions could be obtained.

Since as before,

$$dy(t) = g(t, y(t))dt + dB(t), \tag{77}$$

where $g(t, y(t)) = \{(1/\gamma)\underline{a}^T(t)\underline{f}\}$, then according to Lipster and Shiryayev [20, Chapter 7], the Cramér-Rao lower bound for the unbiased estimate of $\hat{\underline{f}}$ given γ is

$$E\{(\underline{f} - \hat{\underline{f}})(\underline{f} - \hat{\underline{f}})^T\} \geq \left[E\left\{\int_0^t \left[\frac{\partial g}{\partial \underline{f}}\right]\left[\frac{\partial g}{\partial \underline{f}}\right]^T ds\right\}\right]^{-1}$$
$$= \left[E\left\{\frac{1}{\gamma^2}\int_0^t \underline{a}(s)\underline{a}^T(s)ds\right\}\right]^{-1}. \tag{78}$$

And since as before $A(t) = A^T(t) = [\int_0^t \underline{a}(s)\underline{a}^T(s)ds]$, and $\underline{a}(t) = 2\pi[1 \ (2t) \ \cdots \ (It^{I-1})]^T$, substituting $A(t)$ in the above equation we get

$$E\{(\underline{f} - \hat{\underline{f}})(\underline{f} - \hat{\underline{f}})^T\} \geq \{\gamma^2 A^{-1}(t)\}, \tag{79}$$

that is, the approximate Cramér-Rao lower bound is equal to $\{\gamma^2 A^{-1}(t)\}$.

Comparing (79) and (76), one concludes that, given γ, the variance of the frequency estimate is exactly the Cramér-Rao lower bound; a nice result indeed. This result is expected since the frequency estimate is the maximum likelihood estimate.

### 3.3.4.	Sensitivity of the estimate with respect to γ

We can find the derivative of $\hat{\underline{f}}$ with respect to $\hat{\gamma}$, $\partial\hat{\underline{f}}/\partial\hat{\gamma}$, assuming that $z(t)$ is completely known deterministic quantity. The other approach is to find the derivative of the expected value of $\hat{\underline{f}}$ with respect to $\hat{\gamma}$, $\partial E\{\hat{\underline{f}}\}/\partial\hat{\gamma}$, and replace $z(t)$ with its stochastic solution, that is, $z(t) = z(t, W(t))$. Fortunately, for a given $\hat{\gamma}$, $\hat{\underline{f}}$ is unbiased (see (50b) and (71)). Thus, $E[\hat{\underline{f}}]$ is independent of the estimate of $\hat{\gamma}$. The variance, however, is dependent on the estimated values of $\hat{\gamma}$ as shown in (79). Thus, an accurate estimate of $\hat{\gamma}$ is needed.

### 3.4.	Numerical solution of the maximization problem

The log likelihood function, λ of (47), is a true log likelihood as long as γ is known. λ has to be maximized with respect to the unknowns, $\underline{f}$. Since λ has many integral expressions, one has to use numerical solutions to the integral equation.

(1) Discretize the form of the pseudo-log-likelihood function λ of (47) using the formulae in (80) and (81).

(2) For a given γ, use (50) to find $\hat{f}$.

(3) For a given estimated $\hat{f}$, use the simulation-based method and the stochastic annealing method to find an estimate $\hat{\gamma}$.

(4) Repeat until there is no change in the estimates.

ALGORITHM 1: Pseudo-maximum likelihood estimate of the parameters.

The pseudo-log-likelihood function is discretized using, for example, Euler approximation and the following formulae:

$$\int_0^t h(s, z(s))\, ds = \sum_{i=0}^{N-1} h(i, z(i))\Delta_i, \tag{80}$$

where $z(i)$ is the value of $z(t)$ at instant i, $h(i, z(i))$ is the value of $h(s, z(s))$ at instant i, and Δ_i is the sampling interval between the ith sample and the $(i+1)$th sample. Δ_i is usually constant. Moreover

$$\int_0^t h(s, z(s))\, dz(s) = \sum_{i=0}^{N-1} h(i, z(i))\Delta z(i), \tag{81}$$

where $\Delta z(i) = z(i+1) - z(i)$. Other more accurate but computationally intensive discretization methods are also available (Kloeden and Platen [18]).

3.5. Algorithm 1

3.5.1. Proposed Algorithm 1

The steps described in the previous section are enough to yield the pseudo-maximum likelihood estimates of the unknown parameters. We now put these steps in a form of algorithm as shown in Algorithm 1.

Algorithm 1 yields accurate estimates of the parameters as long as the signal-to-noise ratio is more than 10 dB. It was observed that using the method of bootstrapping and using the peak of the histogram generated by this method, one is able to get reasonable results for SNR values as low as 0 db. While the method of bootstrapping is used to find an estimate for the bias and an estimate for the confidence interval around each estimated parameter, our observation is that it could also be used to improve the accuracy of the estimated parameters. This is explained in the bootstrapping subsection.

3.5.2. Convergence of the proposed Algorithm 1

One is always interested in using an algorithm that will yield an estimate that converges to the true unknown parameter. Algorithm 1 has this property. In essence, under mild regularity conditions and assuming that the observed process, $z(t)$, is ergodic, and assuming that we have good initial guesses of $\hat{\gamma}$ and $\hat{f}$, then the proposed steps will yield estimates that are consistent as the number of data points increases to infinity. For more mathematical details and proofs, for a similar problem, and for more on the necessary conditions for consistency see [36].

3.6. Bootstrapping and parameter estimation

The bootstrapping methods depend on the notion of a bootstrap sample. The samples are drawn or generated, at random, from the data. Each set of samples has the same size as the original data size. For each set, the model parameters are estimated. From these estimates, a histogram is obtained. If the samples are drawn at random, we have nonparametric bootstrapping. If the samples are generated at random, from a model for the data, we have what is called parametric bootstrapping (see Zaman [22], Efron and Tibshirani [21]). We will be more concerned with the parametric bootstrap.

3.6.1. Parametric bootstrap

In this approach, by maximizing the pseudo-log-likelihood function, and using the simulation-based method, the estimates of the parameters, $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_I$, and $\hat{\gamma}$ are obtained from the original data. These estimates are called plug-in estimates. The estimates of the Brownian-motion component, $\hat{B}(t)$, are obtained as follows:

$$\begin{aligned}
z(t) &= \sin\left[2\pi \sum_{i=1}^{I} \hat{f}_i t^i + \hat{\gamma}\hat{B}(t)\right] \\
&= \sin\left[2\pi \sum_{i=1}^{I} \hat{f}_i t^i\right] \cos\left[\hat{\gamma}\hat{B}(t)\right] \\
&\quad + \cos\left[2\pi \sum_{i=1}^{I} \hat{f}_i t^i\right] \sin\left[\hat{\gamma}\hat{B}(t)\right],
\end{aligned} \tag{82}$$

and since $\cos x = \sqrt{1 - (\sin x)^2}$, then the equation for $z(t)$ could be changed to

$$\begin{aligned}
z(t) &= \sin\left[2\pi \sum_{i=1}^{I} \hat{f}_i t^i\right] \sqrt{1 - \left\{\sin\left[\hat{\gamma}\hat{B}(t)\right]\right\}^2} \\
&\quad + \cos\left[2\pi \sum_{i=1}^{I} \hat{f}_i t^i\right] \sin\left[\hat{\gamma}\hat{B}(t)\right]
\end{aligned} \tag{83}$$

which is an equation in the unknown $\hat{B}(t)$. Solving, we get two values for $\hat{B}(t)$. Any one of them could be used for studying the statistical properties of $\hat{B}(t)$. We should expect $\hat{B}(t)$ to be Brownian motion if the signal model is correct and if the pseudo-maximum likelihood estimates are close to the true values. Otherwise, one has to reconsider the signal model. If it is Brownian motion as expected, we use random number generator to simulate the jth Brownian motion $B^{(j)}(t)$. We then generate the random samples for bootstrapping using the plug-in estimates and according to the equation

$$z^{(j)}(t) = \sin\left(2\pi \sum_{i=1}^{I} \hat{f}_i t^i + \hat{\gamma}B^{(j)}(t)\right), \tag{84}$$

where $z^{(j)}(t)$ is the jth set of observations that were generated by the jth simulated Brownian motion $B^{(j)}(t)$.

The set of the new observations, $z^{(j)}(t)$, represents the data set that will be used in the pseudo-maximum likelihood (47) to obtain the jth estimate of the parameters as explained in Section 3.2. This results in an estimate for each of the unknown parameters. We then use random number generator to simulate the $(j + 1)$th Brownian motion $B^{(j+1)}(t)$. Another set of observations, $z^{(j+1)}(t)$, is generated according to (84), and another bootstrap estimate is obtained and so forth for as many times as we want. After we stop, a histogram is calculated for each parameter from the bootstrap estimates. From this histogram, the final parameter estimate is the value at the peak of the histogram. One could use the average values of the bootstrap estimates as the final estimate of the parameters. In our case, the peak of the histogram proved to be a more reliable estimate than the plug-in estimate.

3.6.2. Model selection and computational complexity

In some applications one is not able to determine the order of the polynomial that represents the phase. In more complicated scenarios, one is faced with several models for the signal, and we have to determine which model is more representative of the data. For a given model, order selection could be based on something like Akaike information criterion (AIC) or another criterion (Zaman [22]). Model selection should be based on physical understanding of the source of the signal. There are, however, several tests to differentiate between one model and another (Zaman [22]).

It is clear that the proposed approach is computationally intensive. The numerical integration and the simulation-based method are time consuming. The bootstrapping method is known to require many iterations. The CPU time needed for the proposed approach is more than 100 times more than the approximate maximum likelihood method. The CPU time is mainly needed by the bootstrapping method. Parallel processing will drastically reduce the required time.

4. SIMULATIONS

The proposed stochastic-calculus-based approach was tested against the approximate maximum likelihood method of Section 2. In the example, a third-order polynomial with SNR gradually decreased from SNR = 20 dB to reach SNR = 0 dB. The SNR is defined as

$$\text{SNR} = 10 \log \left(\frac{\text{SP}}{\text{NP}} \right), \tag{85}$$

where SP is signal power and NP is noise power defined as follows:

$$\text{SP} = \left(\sin \left(2\pi \sum_{i=0}^{I} f_i t^i \right) \right)^2, \tag{86}$$

and the noise is defined as

$$\text{NP} = \left[\sin \left(2\pi \sum_{i=0}^{I} f_i t^i \right) - \sin \left(2\pi \sum_{i=0}^{I} f_i t^i + \gamma B(t) \right) \right]^2. \tag{87}$$

The stochastic-calculus-based approach had good performance for SNR as low as 10 dB. Adding the bootstrap approach, one is able to obtain good results even at values of the SNR close to 0 dB. The quality of the results are measured by comparing the variances of the estimates using stochastic calculus to the approximate maximum likelihood estimates and to their corresponding approximate Cramér-Rao lower bounds.
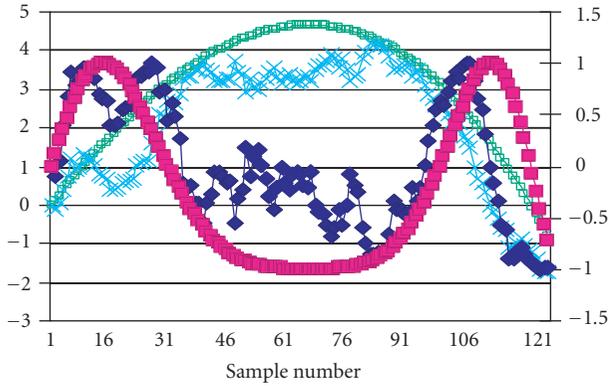
Example 1 (third-order polynomial). In this example, 128 points of the noisy measured signal, $z(0), z(1), \ldots, z(127)$, were simulated according to the equation

$$z(t) = \left[\sin \left(2\pi \sum_{i=1}^{3} f_i t^i + \gamma B(t) \right) \right]. \tag{88}$$

The coefficients were taken as $f_1 = 2$, $f_2 = -1$, $f_3 = -0.5$, that is, $I = 3$. The sampling interval $\Delta = 0.01$. The variance is set to have SNR values 20 dB ($\gamma = 0.1$), 10 dB ($\gamma = 0.3$), and 0 dB ($\gamma = 1.8$). The values of the parameter γ were changed till we got the desired SNR. The reported SNR is actually the average value of many simulations. A typical set of data is shown in Figure 1a where we plot the phase without Brownian motion $[2\pi \sum_{i=1}^{3} f_i t^i]$, and the phase with Brownian motion $[2\pi \sum_{i=1}^{3} f_i t^i + \gamma B(t)]$, with $\gamma \approx 1.8$. This yields SNR = 0 dB, where SNR is as defined in (85), (86), and (87).
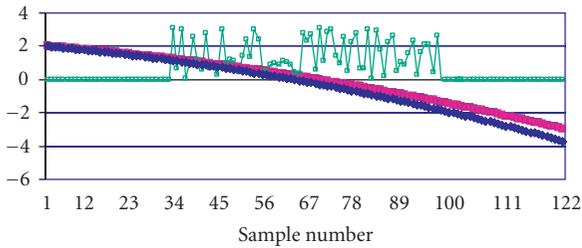
In Figure 1b, we show the instantaneous frequency (IF) as estimated using WVD, as described in Section 2, for a sample of a simulated signal with SNR = 0 dB. The simulated signal is as given above. We also show the true noise free IF and the IF estimated through the stochastic-calculus-based method and bootstrapping. Notice the poor performance of the WVD. Obviously, this performance could be improved if we use a proper window or other modifications to the WVD. Thus, we will not pursue the comparison to the WVD in this paper. Instead, the comparison will be limited to the approximated maximum likelihood method and the stochastic-calculus-based method.

Similarly, in Figure 1c we show the phase angle as estimated using Hilbert transform, as described in Section 2, for a sample of a simulated signal with SNR = 0 dB. We also show the true noise free phase. Notice the poor performance of the Hilbert-transform-based method. Obviously, this performance could be improved if we use a proper window or other modifications to the Hilbert transform. Thus, we will not pursue the comparison to the Hilbert-transform-based method in this paper. Instead, the comparison will be limited to the approximated maximum likelihood method and the stochastic-calculus-based method.
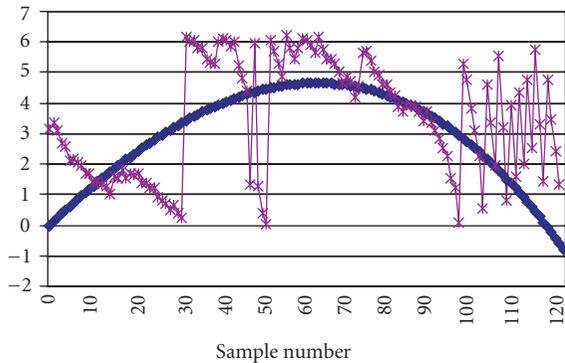
(a)

(b)

(c)

FIGURE 1: (a) Third-order polynomial phase without and with Brownian motion, and their sines, SNR = 0 dB. (b) Third-order polynomial phase true IF estimate using Wigner-Ville distribution, and IF estimate using stochastic-calculus-based method with bootstrapping, SNR = 0 dB. (c) Third-order polynomial phase and the estimated phase using Hilbert transform (adjusted between 0 and 2π), SNR = 0 dB.

(1) Discretize the form of the log likelihood function λ of (47) using the formulae (80) and (81).
(2) For a given γ, use (50) to find $\hat{f}$.
(3) For a given estimated $\hat{f}$, use the simulation-based method and the stochastic annealing method to find an estimate $\hat{\gamma}$.
(4) Repeat until there is no change in the estimates.
(5) Generate a bootstrap sample of observations using (84).
(6) Repeat steps (1)–(5) a large number of times, say 100.
(7) Find the histogram of the estimates and take the desired estimate of the parameter as the value where the histogram is maximum.
(8) Generate, through bootstrapping, a new set of observations and repeat steps 1–7 and stop after, say $M = 100$ iterations.

ALGORITHM 2: Pseudo-maximum likelihood estimate of the parameters with bootstrapping.

4.1. Proposed Algorithm 2

In Section 3.4 we gave an outline of how to find the pseudo-maximum likelihood estimates of the parameters. This did not include the bootstrapping segment. The proposed Algorithm 2 covers this gap.

One could use the average values of the bootstrap estimates as the final estimate of the unknown parameters. The peak, however, seems to yield a more accurate estimate of the unknown parameter. As the number of simulations increases, we get more accurate results. In our case, however, 100 simulations seems to be enough to get good results.

The histograms of the bootstrap estimates of the three parameters, using the stochastic-calculus-based method, are shown in Figure 2. The approximate Cramér-Rao lower bound for the variances, the variances of the estimates of the coefficients of the polynomial phase using stochastic-calculus-based method, and the approximated maximum likelihood method are shown in Figure 3.

The mean square error (MSE) between the estimated parameters and the true parameters is taken as a measure of performance. The MSE for the ith coefficient is defined as

$$\text{MSE}_i = \frac{1}{M} \sum_{m=1}^{M} \left(\hat{a}_i^m - a_i \right)^2, \qquad (89)$$

where M is the total number of the bootstrap simulations and is equal to 100, $\hat{a}_i^m$ is the mth estimate of the ith coefficient which is the peak of the histogram of this coefficient.

In Table 1, we present the log value of the approximated Cramér-Rao lower bound (CRB) and the log of the MSE values for the stochastic-calculus-based method with bootstrapping for SNR value of 20 dB, 10 dB, and 0 dB. Notice how close they are.

4.2. Effect of bootstrapping

In Table 2, we present the MSE values for SNR = 0 dB. Both methods, approximate maximum likelihood method
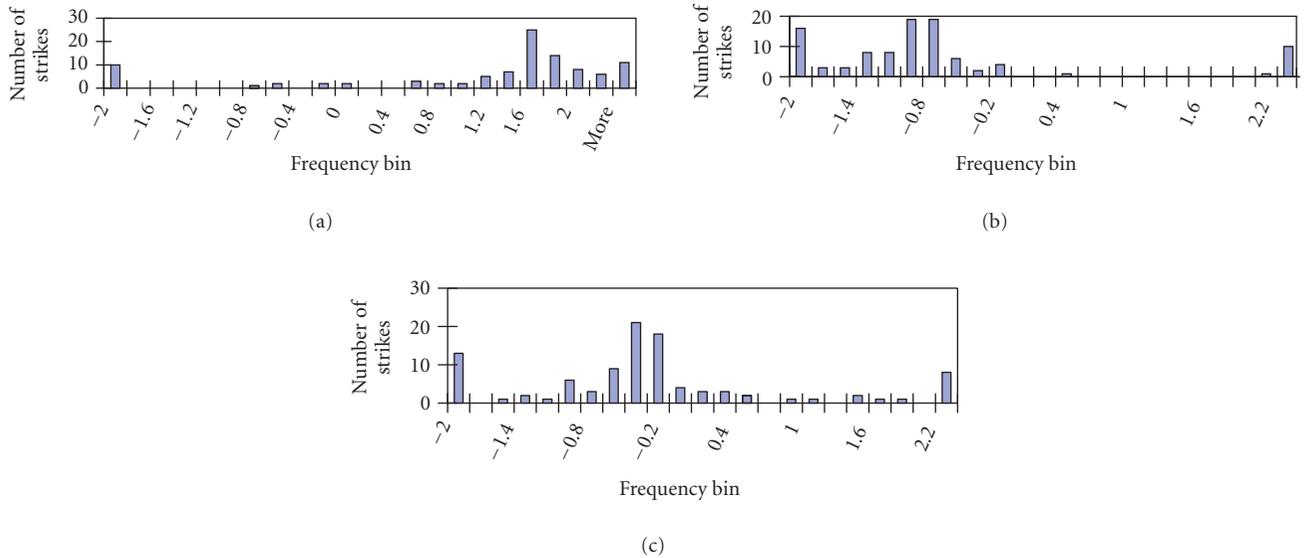
(a)

(b)

(c)

FIGURE 2: (a) Histogram of estimated $\hat{f}_1$ using stochastic-calculus-based method, the maximum value of the histogram at 1.6, SNR = 0 dB. (b) Histogram of estimated $\hat{f}_2$ using stochastic-calculus-based method, the maximum value of the histogram at -1, SNR = 0 dB. (c) Histogram of estimated $\hat{f}_3$ using stochastic-calculus-based method, the maximum value of the histogram at -0.4, SNR = 0 dB.

with bootstrap and the proposed stochastic-calculus-based method with bootstrap, are compared. Notice that using the bootstrapping, as expected, lowers the MSE of the estimates.

Notice that the accuracy of the proposed bootstrapping and stochastic-calculus-based method is much higher than the approximate maximum likelihood method. The measure of accuracy is the mean square error of the estimate.

4.3. Could we use bootstrapping to improve the performance?

Bootstrap is a method that handles small samples and yields good estimates for the bias and the confidence intervals for the plug-in estimates (see Efron and Tibshirani [21], Zaman [22], Politis [23], Zoubir and Boashash [24]). It is not meant to be a method to improve the accuracy of the estimates. Using the peak of the histogram, generated from the bootstrapping method for each estimated parameter, as the desired estimate of the unknown parameter, however, one is able to get good estimates at lower SNR. This was observed in this report and in some other applications (see Souza and Neto [25], Abutaleb [26]). This could be explained from the fact that the bootstrapping samples are almost the same as Monte Carlo simulations. Thus, instead of having just one data set for the estimation, one has as many samples as needed. More important, bootstrapping provides a reasonably good approximation to the density of the parameter estimator. This is apparent from the histogram of each parameter of Figure 2. If the distribution of each parameter was Gaussian, then one should take the mean of the histogram to be the best estimate of the unknown parame-

ter. But as we notice from (71) and Section 3.3, the distribution of the estimates, $\hat{\underline{f}}$, is not Gaussian. This is true since the estimate of $\hat{\gamma}$ comes into this equation. Thus, one might expect the density of $\hat{\underline{f}}$ to be skewed. This is exactly what we observe in the histograms of Figure 2. Choosing the peak of the histogram, as an estimate for the unknown parameter, is the reason for the improved accuracy and the lowering of the MSE. More discussion on this subject could be found in [22, Chapters 12 and 14]. Again this is not a mathematical explanation for this observation. More theoretical analysis is needed to explain why the peak yielded better estimates. This is currently under investigation.

Table 3 lists, for SNR = 0 dB and for a typical group of bootstrapping estimates, the true value of the unknown parameter, the plug-in estimate, the bias, the average, the variance, and the peak of the histogram for each parameter. An estimate of the bias is obtained from the bootstrap estimates. It is defined as follows (Efron and Tibshirani [21, Chapter 10]): bias is equal to the average of the bootstrap estimates of the unknown parameter—the plug-in estimate of the unknown parameter.

5. SUMMARY

In this paper, the formulation of the instantaneous frequency estimation problem using concepts in stochastic calculus was introduced, and applied to the estimation of a single time-varying phase represented by a polynomial with additive Brownian motion. A stochastic differential equation was developed for the observed signal. This resulted in a linear
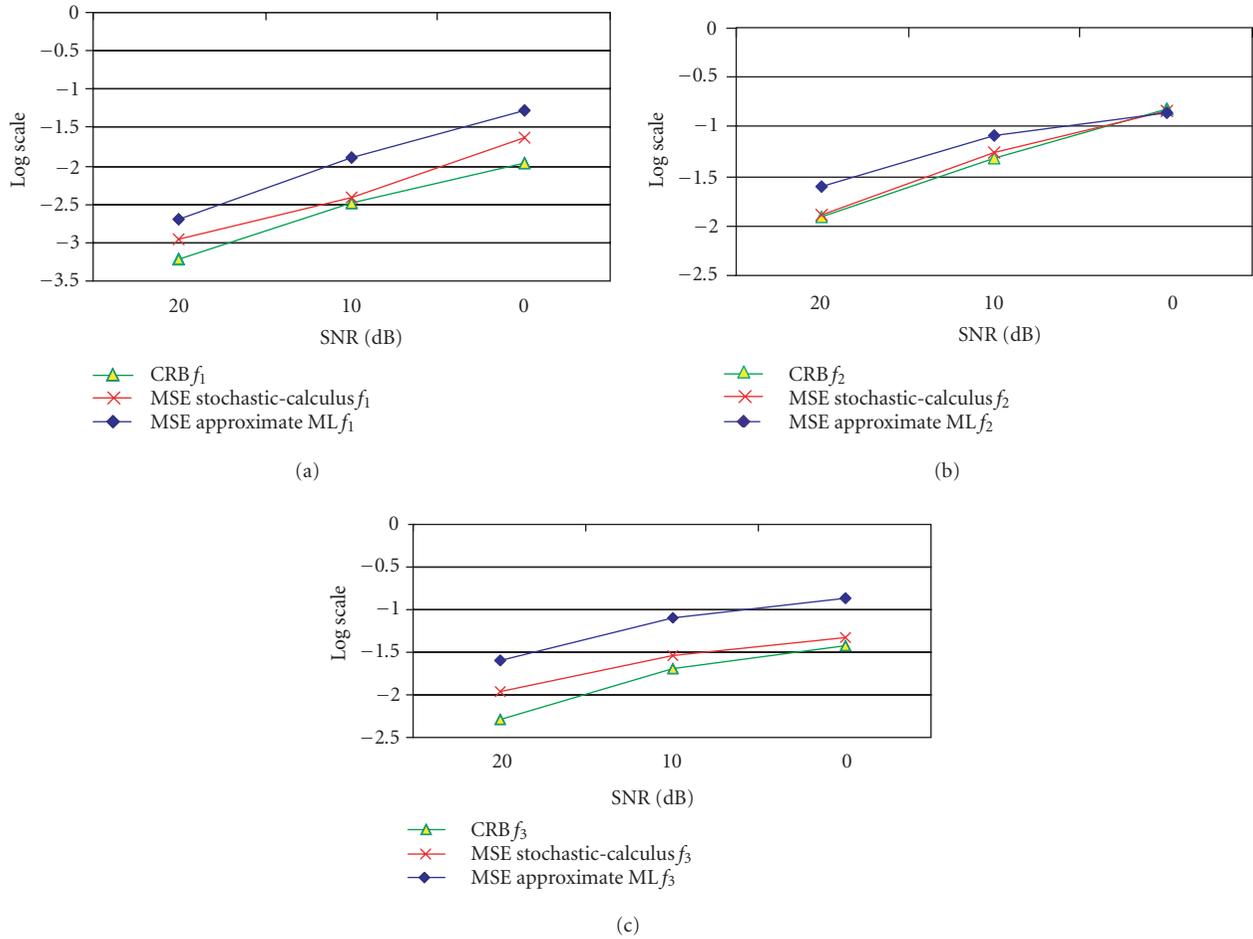
(a)

(b)

(c)

FIGURE 3: (a) Approximate CRB and MSE of the stochastic-calculus-based estimate of $\hat{f}_1$. (b) Approximate CRB and MSE of the stochastic-calculus-based estimate of $\hat{f}_2$. (c) Approximate CRB and MSE of the stochastic-calculus-based estimate of $\hat{f}_3$.

TABLE 1: Log of the approximate CRB and the MSE of the stochastic-calculus-based method with bootstrapping. SNR values vary from 20 dB down to 0 dB.

Coefficient	CRB (SNR = 20 dB)	CRB (SNR = 10 dB)	CRB (SNR = 0 dB)	MSE of stochastic calculus with bootstrapping (SNR = 20 dB)	MSE of stochastic calculus with bootstrapping (SNR = 10 dB)	MSE of stochastic calculus with bootstrapping (SNR = 0 dB)
f_1	−3.2	−2.4	−1.9	−2.9	−2.4	−1.6
f_2	−1.9	−1.3	−0.8	−1.9	−1.2	−0.8
f_3	−2.3	−1.7	−1.4	−1.9	−1.5	−1.3

equation in some of the unknown parameters of the phase; an advantage of using Ito calculus. A pseudo-likelihood function was obtained using the Girsanov theory. It was shown that the pseudo-likelihood function is quadratic in some of the unknown coefficients of the phase. This made it possible to find an exact expression for the estimates of the parameters of the polynomial phase; another advantage of the Ito calculus. The statistical properties of the estimates were also derived in a straightforward manner and an approximate expression for the Cramér-Rao lower bound was ob-

tained. A simulation-based method was used to, numerically, obtain estimates of the rest of the unknown parameters. Bootstrapping method was introduced and was used with stochastic annealing. This way, good estimates were obtained; the mean square error of each estimate is close to the approximate Cramér-Rao lower bound. Simulations proved the superior performance of the proposed approach, and a comparison was given with the performance of the Wigner-Ville distribution and the Hilbert-transform-based results.

TABLE 2: Mean square error of the estimates using the proposed stochastic-calculus-based method with bootstrapping, the approximate maximum likelihood method with bootstrapping, and the approximate Cramér-Rao lower bound, SNR = 0 dB. All values are in log scale.

Coefficient	MSE (proposed bootstrapping and stochastic-calculus-based method)	MSE (stochastic-calculus-based method)	MSE (bootstrapping and approximate maximum likelihood method)	MSE (approximate maximum likelihood method)	Approximate Cramér-Rao lower bound
f_1	-1.6	-1.1	-1.2	-0.9	-1.9
f_2	-0.8	-0.6	-0.2	0	-0.8
f_3	-1.3	-0.9	-0.8	-0.2	-1.4

TABLE 3: Some statistics for the histogram of Figure 2, SNR = 0 dB.

Coefficient	True value	Plug-in estimate	Peak of the histogram	Average of the histogram	Bias of the average	Variance of the average
f_1	2	2.3	1.6	2.2	0	0.007
f_2	-1	-1.5	-1.0	-1.3	0.195	0.052
f_3	-0.5	-0.5	-0.4	-0.6	-0.1	0.013

A more general model for the received signal, where we have more than one component each with random phase or random amplitude, is currently under investigation.

REFERENCES

[1] J. Blackledge, *Quantitative Coherent Imaging*, Academic Press, London, UK, 1989.

[2] R. Goldstein, H. Zebker, and C. Werner, "Satellite radar interferometry: two-dimensional phase unwrapping," *Radio Science*, vol. 23, no. 4, pp. 713–720, 1988.

[3] S. Moon-Ho Song, S. Napel, N. J. Pelc, and G. H. Glover, "Phase unwrapping of MR phase images using poisson equation," *IEEE Trans. Image Processing*, vol. 4, no. 5, pp. 667–676, 1995.

[4] D. Ghiglia and M. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, John Wiley & Sons, New York, NY, USA, 1998.

[5] D. Wehner, *High Resolution Radar*, Artech House, Norwood, Mass, USA, 1995.

[6] T. Li, "Multipath time delay estimation using regression stepwise procedure," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 191–195, 1998.

[7] V. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, Artech House, Boston, Mass, USA, 2002.

[8] *Proc. IEEE*, vol. 84, no. 9, September 1996, Special Issue on Instantaneous Frequency Estimation.

[9] V. Katkovnik and L. Stankovic, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2315–2325, 1998.

[10] B. Barkat, "Instantaneous frequency estimation of nonlinear frequency modulated signals in the presence of multiplicative and additive noise," *IEEE Trans. Signal Processing*, vol. 49, no. 10, pp. 2214–2222, 2001.

[11] M. Benidir and A. Ouldali, "Polynomial phase signal analysis based on the polynomial derivatives decomposition," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1954–1965, 1999.

[12] S. Goto, M. Nakamura, and K. Uosaki, "On line spectral estimation of nonstationary time series based on AR model parameter estimation and order selection with a forgetting factor," *IEEE Trans. Signal Processing*, vol. 43, no. 6, pp. 1519–1522, 1995.

[13] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," *Proc. IEEE*, vol. 80, no. 4, pp. 520–568, 1992.

[14] S. Peleg and B. Porat, "Estimation and classification of polynomial phase signals," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 422–430, 1991.

[15] S. Peleg and B. Friedlander, "The discrete polynomial phase transform," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 1901–1914, 1995.

[16] B. Porat and B. Friedlander, "Asymptotic statistical analysis of the high-order ambiguity function for parameter estimation of polynomial phase signals," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 995–1001, 1996.

[17] S. Golden and B. Friedlander, "Estimation and statistical analysis of exponential polynomial signals," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3127–3130, 1998.

[18] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, NY, USA, 1992.

[19] D. McFadden, "A Method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, vol. 57, no. 5, pp. 995–1026, 1989.

[20] R. Lipster and A. Shiryayev, *Statistics of Random Processes, Vols. I and II*, Springer-Verlag, New York, NY, USA, 1978.

[21] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, NY, USA, 1994.

[22] A. Zaman, *Statistical Foundations for Econometric Techniques*, Academic Press, New York, NY, USA, 1996.

[23] D. Politis, "Computer-intensive methods in statistical analysis," *IEEE Signal Processing Mag.*, vol. 15, no. 1, pp. 39–55, 1998.

[24] A. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Mag.*, vol. 15, no. 1, pp. 56–76, 1998.

[25] R. Souza and A. Neto, "A bootstrap simulation study in ARMA (p, q) structures," *Journal of Forecasting*, vol. 15, no. 4, pp. 343–353, 1996.

[26] A. Abutaleb, "Number theory and bootstrapping for phase unwrapping," *IEEE Trans. Circuits Syst. I*, vol. 49, no. 5, pp. 632–638, 2002.

[27] A. Abutaleb, "Instantaneous frequency estimation when the amplitude is a stochastic process using stochastic calculus and bootstrapping," *Circuits, Systems, and Signal Processing*, vol. 24, no. 1, pp. 35–52, 2005.

[28] A. Abutaleb and M. Papaioannou, "Malliavin calculus for the estimation of the US Dollar/EURO exchange rate when the volatility is stochastic," in *Global Information Technology and Competitive Financial Alliances*, Y. Kurihara, Ed., Elsevier Sciences, Amsterdam, the Netherlands, 2005.

[29] D. Nualart, *The Malliavin Calculus and Related Topics*, Springer-Verlag, New York, NY, USA, 1995.

[30] A. Abutaleb, "A genetic algorithm for the maximum likelihood estimation of the parameters of sinusoids in noisy environment," *Circuits, Systems, and Signal Processing*, vol. 16, no. 1, pp. 69–81, 1997.

[31] B. Boashash and P. Black, "An efficient real-time implementation of the Wigner-Ville distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 11, pp. 1611–1618, 1987.

[32] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, New York, NY, USA, 1995.

[33] B. Boashash, "Note on the use of the Wigner distribution for time-frequency signal analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 9, pp. 1518–1521, 1988.

[34] B. Oksendal, *Stochastic Differential Equations*, Springer-Verlag, New York, NY, USA, 1998.

[35] B. Prakasa Rao, *Semimartingales and Their Statistical Inference*, vol. 83 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Washington, DC, USA, 1999.

[36] N. Yoshida, "Estimation for diffusion processes from discrete observation," *Journal of Multivariate Analysis*, vol. 41, no. 2, pp. 220–242, 1992.

A. Abutaleb is currently a Professor at the Systems and Bioengineering Department, School of Engineering, Cairo University, Giza, Egypt. From 1988 till 1991 he was a staff member at Lincoln Laboratory, MIT, Lexington, Mass, USA, in the Radar Data Analysis Division. From 1986 till 1988, he was an Assistant Professor at the School of Engineering, Temple University, Philadelphia, Pa, USA. From 1984 till 1986, he was an Assistant Professor at the School of Graduate Studies, the Medical University of South Carolina, Charleston, South Carolina, USA. He received his Ph.D. and M.S. degrees from the School of Engineering, the University of Pennsylvania, in 1984 and 1980,respectively, the Diploma in electrical engineering from the School of Engineering, MIT, in 1979,and the B.S. degree from the Electrical Engineering Department, School of Engineering, Cairo University, in 1976. His major areas of interest are statistical signal processing, stochastic calculus and its applications in signal processing and control, communications, economics and forecasting, applied econometrics, image processing, radar data analysis, and number theory and its applications. He has published over 100 papers in technical fields since 1980, and is currently preparing a book entitled *Data Analysis and Signal Processing: Econometrics and Number Theory Perspective*. He is a Consultant to several national and international organizations.