A Test for Aliasing Using Bispectral Analysis

MELVIN J. HINICH and MURRAY A. WOLINSKY*

Aliasing is a signal-confounding problem that arises when a continuous-time signal is sampled at a rate slower than twice the highest frequency component of a Fourier series representation of the signal. Aliasing can be especially serious for social-science time series applications, since the sampling designs used to construct most social-science data bases are fixed by considerations other than the nature of the underlying continuous-time mechanisms. After collecting sampled data, it is of value to test the observations for the presence of aliasing. It is shown that the nature of the support set of a sampled band-limited stationary signal can be used to motivate an amended version of the Hinich bispectrum test for Gaussianity (Hinich 1982) as a test for aliasing.

KEY WORDS: Principal domain; Hypothesis test; Support set; Band limited; Bicovariance function.

1. INTRODUCTION

The two sinusoids $x(t) = \cos(2\pi f_o t)$ and $y(t) = \cos(4\pi f_o t)$ yield identical sampled values if they are sampled at times $t = n/f_o$, where n is an integer. Thus these two functions cannot be identified from a discrete-time sample of a time series that is sampled at the rate f_o. This is a simple example of the aliasing problem that can arise when a time series is sampled at equal time points. If the time between successive observations is τ (for a sampling rate of $1/\tau$), then each frequency component of the series for frequencies $-1/2\tau \le f \le 1/2\tau$ can be confounded by components whose frequencies are $f + n/\tau$, where n is a signed integer. This confounding by aliases will not occur if the following conditions hold: (a) the underlying continuous-time signal has no frequency content beyond a frequency f_o, which is called the *band limit*, and (b) the signal is sampled at a rate that is greater than or equal to the frequency $2f_o$—that is, if $1/\tau \ge 2f_o$.

Aliasing is usually avoided in engineering and physical-science applications by filtering the signal to eliminate its energy above a certain cutoff frequency and then sampling the filtered signal at or above twice the cutoff frequency. This type of sampling design is usually impossible for social-science applications, since the sampling rate is determined according to logistic and cost considerations without regard for the spectral characteristics of the underlying process.

Once a sampling process is used to collect data, it is of value to be able to test the observations for the presence of a significant amount of aliasing. We will show that an overlooked property of the principal domain of a discrete-time band-limited stationary signal can be used to motivate an amended version of the Hinich bispectrum test for Gaussianity (Hinich 1982) as a test for aliasing.

The bispectrum of a discrete-time signal is a periodic function in two frequency indices. There is a surprisingly persistent confusion between the statistics and engineering literature regarding the triangular form of the principal domain of the bispectrum of a time series that is a discrete-time sample of a continuous-time band-limited signal. Huber, Kleiner, Gasser, and Dummermuth (1971), Kim and Powers (1979), and Matsuoka and Ulrych (1984) postulated isosceles triangular domains that are a subset of the triangle given by Brillinger and Rosenblatt (1967a,b), Lii and Rosenblatt (1982), Subba Rao and Gabr (1980), Hinich (1982), and Subba Rao (1983).

We will now demonstrate that the set of positive support of the continuous-time bispectrum is a proper subset of the principal domain of the bispectrum of the sampled process. Huber et al. (1971) and others in the engineering literature are really referring to the *support set* rather than the *principal domain*, and that support set is the one of substantive interest. The extra triangle in Brillinger and Rosenblatt (1967a,b) is due to the frequencies that sum to 1 (2π in angular frequency units). We will show that the bispectrum must be *zero* in that extra triangle if the sampling rate is equal to or exceeds twice the highest frequency in which the spectrum is positive, the band limit f_o (see Fig. 1). This result is surprising when $1/\tau = 2f_o$, since the spectrum of the sampled signal is nonzero up to the folding frequency $1/2\tau$, assuming that the spectrum has no zero regions in its principal domain of $0 \le f \le 1/2\tau$. A proof of this result will be given after a review of the definition of a bispectrum.

2. CONTINUOUS–TIME BISPECTRA

To review the mathematics of bispectra, let $x(t)$ denote a real zero-mean stationary continuous-time stochastic process. Assume that all expected values, sums, and integrals used here exist. The *bicovariance function* of the process is $c(u, v) = Ex(t)x(t + u)x(t + v)$, which does not depend on t because the process is stationary. The double Fourier transform of $c(u, v)$,

$$B(f, g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(u, v)\exp[-i2\pi(fu + gv)]\, du\, dv, \tag{1}$$

is called the *bispectrum* of the process. Although this two-frequency index notation is standard, it hides the three-

* Melvin J. Hinich is Professor of Government and a member of the Research Faculty at Applied Research Laboratories, University of Texas, Austin, TX 78713-8029. Murray A. Wolinsky is a Graduate Research Assistant at Applied Research Laboratories, University of Texas at Austin. This work was supported by Office of Naval Research Contract N00014-86-K-0237. The authors thank David Brillinger for his idea of squaring the data if the process is Gaussian and the associate editor for his helpful comments.

Journal of the American Statistical Association
June 1988, Vol. 83, No. 402, Theory and Methods

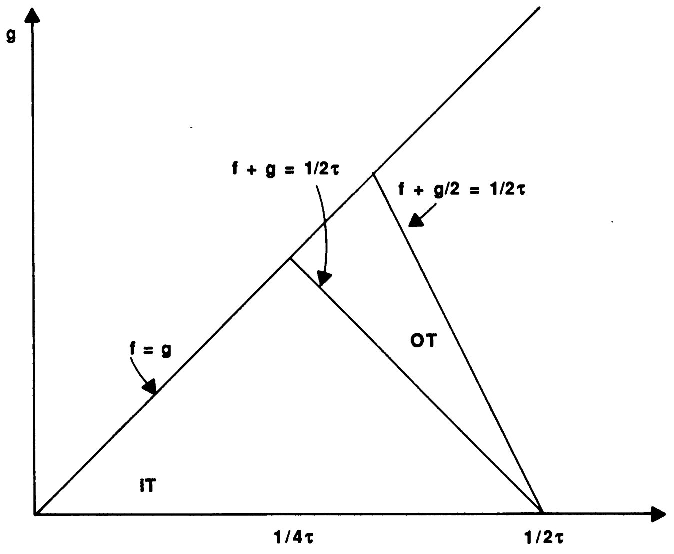

Figure 1. Discrete-Time Principal Domain.

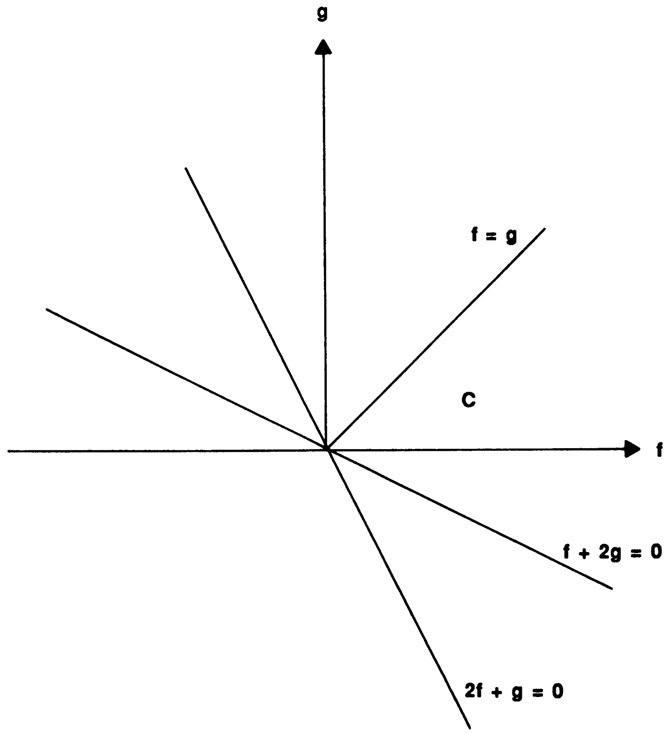

Figure 2. Symmetries of Bispectrum B(f, g).

frequency interaction that is so important for applications of bispectral estimation. To help the exposition, we now switch to the three-index notation used by Brillinger and Rosenblatt (1967a,b), namely $B(f, g, h)$, where $h = -f - g$. To motivate this notation, consider the following redefinition of the bispectrum as a Fourier transform of $c(u, v)$ for all t:

$$\begin{aligned} B(f, g, h) &= \iint Ex(t)x(t + u)x(t + v) \\ &\quad \times \exp\{-i2\pi[f(t + u) \\ &\quad + g(t + v) + ht]\}\, du\, dv \\ &= \iint c(u, v)\exp\{-i2\pi[fu + gv \\ &\quad + (f + g + h)t]\}\, du\, dv. \end{aligned} \tag{2}$$

For $B(f, g, h)$ to equal $B(f, g)$ regardless of t, then $f + g + h = 0$. The right side of expression (2) is invariant to permutations of the frequency indexes f, g, and $h = -f - g$. Thus the bispectrum's symmetry lines are $f = g$, $f = h$ ($2f = -g$), and $g = h$ ($2g = -f$). Another symmetry holds, since $c(u, v)$ is real; namely $B(-f, -g, -h) = B*(f, g, h)$, where $*$ denotes complex conjugation. This skew symmetry yields another three symmetry lines: $f = -g$, $f = -h$ ($g = 0$), and $g = -h$ ($f = 0$) (see Fig. 2). Thus the cone $C = \{f, g : 0 \le f, 0 \le g \le f\}$ is a principal domain of this continuous-time bispectrum in the (f, g) plane.

3. THE PRINCIPAL DOMAIN FOR DISCRETE TIME

Now suppose that the process is band limited at frequency f_o. Then there is no variance in the process for frequencies beyond f_o, and thus the bispectrum cuts off at $f = f_o$, $g = \pm f_o$, and $f + g = \pm f_o$. Then the continuous-time support set is the isosceles right triangle $\{f, g : 0 \le f \le f_o, g \le f, f + g = f_o\}$. The discrete-time principal domain, though, is a larger triangle if the process is sampled at the Nyquist frequency of $2f_o$.

The principal domain can be derived from Equation (2) in a straightforward manner (for a general treatment of polyspectra, see Rosenblatt 1983). Consider the discrete-time sequence $\{x(n\tau)\}$, where $\tau = 1/(2f_o)$. The bicovariance function of this sampled version of $\{x(t)\}$ is really an array, $\{c(j\tau, k\tau) : j, k = 0, \pm 1, \pm 2, \ldots\}$. Then the sampled-data bispectrum is defined, analogous with Equation (2), to be given by the Fourier transform in three indexes:

$$\begin{aligned} {}_\tau B(f, g, h) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c(j\tau, k\tau) \\ &\quad \times \exp\{-i2\pi[fj\tau + gk\tau + (f + g + h)\tau]\}, \end{aligned} \tag{3}$$

where now $f + g + h$ is not just constrained to be 0 but can be equal to n/τ for any signed integer n. Thus the sampling introduces an infinite set of parallel symmetry lines, $2f + g = n/\tau$ and $f + 2g = n/\tau$. The cone C is first cut by the symmetries $2f + g = 1/\tau$, and thus the principal domain of ${}_\tau B$ is the triangle $\{f, g : 0 \le f \le 1/2\tau, g \le f, 2f + g = 1/\tau\}$ in the cone C. We will now give an expression for B in terms of the underlying bispectrum for frequencies in the odd triangle OT $= \{f, g : g \le f, 1/2\tau \le f + g \le (1/\tau) - f\}$ adjoining the isosceles triangle IT $= \{f, g : g \le f, 0 \le f + g \le 1/2\tau\}$ (Fig. 1).

The discrete-time bispectrum ${}_\tau B$ is a periodic function of period $1/\tau$ in each of its three indexes. A special case of a formula in Brillinger and Rosenblatt (1967b, p. 190) states that for f, g, and h in the principal domain of ${}_\tau B$,

$$
{}_\tau B(f, g, h) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} B(f + k/\tau, g + m/\tau, h + n/\tau),
$$

$$
f + g + h + (k + m + n)/\tau = 0, \quad (4)
$$

where the signed integers are restricted to keep the indexes in B's principal domain. For example, if $h = -f - g$, then $k + m + n = 0$. If $h = (1/\tau) - f - g$, then $k + m + n = -1$. But B is band limited at f_o, so the sum is restricted to the k, m, and n such that $|f + k/\tau| \le f_o$, $|g + m/\tau| \le f_o$, and $|h + n/\tau| \le f_o$.

Now consider the case in which there is no aliasing. Take $\tau = 1/(2f_o)$. If f and g are in OT, then $f + g > 1/2\tau = f_o$, $f < f_o$, and $g < f_o$. Thus the key term in the sum to consider is $B(f, g, h - 1/\tau)$, where $h = -f - g + 1/\tau$. This, however, is $B(f, g, -f - g)$, which is 0 because $f + g > f_o$. All of the other terms are 0 for a similar reason. Thus ${}_\tau B$ is 0 for (f, g) in OT.

3.1 Testing for Aliasing

Let $S(f_j, f_k)$ denote a consistent estimator of ${}_\tau B(f_j, g_k)$ for a grid of equally spaced bifrequencies (f_j, g_k) given a sample of size N of $\{x(n\tau)\}$. This estimator can be computed by smoothing the sample bicovariance (Subba Rao 1983), by smoothing the sample bispectrum in the bifrequency domain (Hinich 1982), or by dividing the sample into pieces and averaging the piecewise sample bispectra and then smoothing in the bifrequency domain (Lii and Rosenblatt 1982).

Let B_N denote the bandwidth of the bispectral estimate for the smoothing method used. For simple piecewise averaging using L_N points in each piece of the data record, $B_N = 1/L_N$. Consistency requires that $B_N \to \infty$ as $N \to \infty$. To ensure that the variances of the real and imaginary parts of S go to 0 as $N \to \infty$, assume that $NB_N^2 \to \infty$. Hinich and Patterson (in press) found that setting the bandwidth in the range $1/\sqrt{N} < B_N$ gives results that conform to the large-sample results derived from asymptotic analysis.

Letting $S_x(f)$ denote the spectrum of $\{x(t)\}$, under some mixing conditions such as the ones given by Brillinger (1975) or Rosenblatt (1985), the large-sample distribution of the statistic

$$
2|S(f_j, g_k)|^2/[NB_N^2 S_x(f_j) S_x(g_k) S_x(f_j + g_k)]
$$

is a central chi-square with 2 df if the bispectrum is 0 in a region about (f_j, g_k). The statistics for (f_j, g_k) and (f_r, g_s) are asymptotically independent (for exact statements of the asymptotic properties alluded to here, see Brillinger and Rosenblatt 1967a; Rosenblatt 1985).

The obvious modification of this test for the problem on hand is to sum the chi-squared statistics only over the triangle OT. Under the null hypothesis that there is no aliasing for a sampling rate $1/\tau$, the distribution of the sum is approximately central chi-squared with $2K$ df, where K is the number of grid points in OT and $B_N > 1/\sqrt{N}$. Ashley, Hinich, and Patterson (1986) showed that the approximation used in the Hinich test is good for samples as small as $N = 256$, so there is no reason to doubt its application to this restricted sum test for aliasing. If one does not want to use the large-sample properties of the estimated bispectrum, then the Subba Rao and Gabr (1980) test can be modified in a similar way.

This bispectrum-based test has no power if $c(u, v) \equiv 0$ and thus the bispectrum is identically 0 for all bifrequencies. David Brillinger suggested (personal communication, 1987) squaring the observations so that the test will have power using the transformed series. A variety of transformations can be tried, provided that the mean of the transformed observations is 0 and that no new frequencies are introduced beyond those originally present. The latter constraint rules out squaring and may eliminate *all* nonlinear transformations.

The bispectral analysis programs that we use take at most a few seconds of central-processing-unit time on an IBM 3081 computer, even for $N = 10{,}000$ observations. One program operates on a PC microcomputer. The computational ease of such programs on modern computers makes it feasible to try a variety of transformations using a number of bandwidth values.

In a sense, the test has already been applied. The results of Hinich and Patterson (1985) show that many daily stock series have large peaks in various parts of the principal domain, including OT. Since stock prices change by the minute, it is not surprising that high-frequency components are aliased. All statistically significant terms in OT verify aliasing for the sampling interval of one day. The test was applied by Hinich and Patterson (in press) to three years of continuous trade-by-trade data for 15 stocks. These trade-by-trade data are essentially a continuous-time record of the stock prices, corrected for splits and dividends. The data are passed through an anti-aliasing low-pass filter and sampled at a rate equal to the bandwidth frequency cutoff for the anti-aliasing filter.

3.2 Analysis of a 10-Stock Price Series

To provide an example of this test with new data, the chi-squared alias statistic was completed by using daily rates of return from 10 randomly selected stocks for the sample period January 18, 1980–December 30, 1983. This period has $N = 1{,}000$ consecutive days. The bispectrum is computed for points in the principal domain, averaging in the frequency domain over a square whose sides have a resolution bandwidth of .03 day^{-1}. In other words, 900 raw sample bispectrum values are averaged to produce an estimate of the bispectrum for the square. The spectrum for each stock is averaged by using a truncated cosine kernel whose bandwidth is .237 day^{-1}. There are $K = 24$ center points in the alias triangle OT. We obtained similar results for our aliasing test, using different smoothing parameters for the bispectral and spectral estimates.

Table 1. Aliasing Test Statistics

	1/18/80–12/30/83		2/3/76–1/17/80	
Stock	χ^2_{48}	Z	χ^2_{48}	Z
American Airlines	74.6	5.4	56.2	3.8
Alberto Culver	95.2	7.0	60.5	4.2
CBS	43.9	2.5	79.3	5.7
Campbell Soup	103.5	7.5	48.2	3.0
El Paso Natural Gas	291.2	17.3	133.3	9.5
Swift & Co.	74.7	5.4	80.8	5.9
Federated Department Stores	77.4	5.6	44.0	2.5
Northern Gas	76.4	5.5	85.1	6.2
Indianapolis Power and Light	92.7	6.8	55.2	3.7
Merrill Lynch	54.6	3.6	134.7	9.6

NOTE: The χ^2_{48} statistic is approximately a central chi-squared variate with 48 df under H_0. The Z statistic is a normal approximation of the chi-squared with large df. It is N(0, 1) under H_0.

The results are presented in Table 1. The values for the chi-squared statistics with 48 df are given in column 1. Under the null hypothesis that there is no aliasing, the statistic is approximately chi-squared with 48 df. If there is aliasing, the distribution is approximately *noncentral* chi-squared with 48 df and a positive noncentrality parameter.

The second column presents the Gaussian approximation to a chi-squared random variable with large degrees of freedom. This statistic is approximately $N(0, 1)$ under the null hypothesis. Each stock series has a test statistic that is *not* consistent with the null hypothesis. The results overwhelmingly indicate that the data are aliased for the sampling interval of one day. The results suggest that the underlying continuous-time mechanism that generates the price series has a spectrum whose bandwidth far exceeds the .5 day^{-1} folding frequency for the measured rates of return.

To check on the stability of the results, the prior 1,000 trading-day period (Feb. 3, 1976–Jan. 17, 1980) was analyzed using the same bandwidth for smoothing. Once again each series has a test statistic that is not consistent with the null hypothesis.

These significant results are unlikely to be due to nonstationarity in the distribution of a *linear model*, since such structural nonstationarity will tend to attenuate the sample average of the same bispectral estimates (or equivalently, the sample bicovariances) relative to the sample average of the periodogram, using any of the equivalent smoothing methods. This attenuation is analogous to the reduction undetectability of a shift in the mean of a heteroscedastic time series. In other words, the power of the test is reduced if the joint distribution of the observed process varies within the sampling period.

For each series analyzed, the bispectrum in the isosceles triangle is similar in form to that of the odd triangle. The statistically significant bispectral values are concentrated at certain bifrequency pairs, as is the case for the bispectra presented by Hinich and Patterson (1985). The results are consistent with a hypothesis that the generating mechanism is nonlinear. The aliasing test statistics indicate that a higher sampling rate must be used to identify the structure of the nonlinear mechanism.

[*Received January 1987. Revised September 1987.*]

REFERENCES

Ashley, R., Hinich, M. J., and Patterson, D. (1986), "A Diagnostic Test for Nonlinear Serial Dependence in Time Series Fitting Errors," *Journal of Time Series Analysis*, 7, 165–178.

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

Brillinger, D., and Rosenblatt, M. (1967a), "Asymptotic Theory of *k*th Order Spectra," in *Spectral Analysis of Time Series*, ed. B. Harris, New York: John Wiley, pp. 153–188.

——— (1967b), "Computation and Interpretation of *k*th Order Spectra," in *Spectral Analysis of Time Series*, ed. B. Harris, New York: John Wiley, pp. 189–232.

Hinich, M. J. (1982), "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis*, 3, 169–176.

Hinich, M. J., and Patterson, D. M. (1985), "Evidence of Nonlinearity in Daily Stock Returns," *Journal of Business & Economic Statistics*, 3, 69–77.

——— (in press), "Evidence of Nonlinearity in the Trade-by-Trade Stock Market Return Generating Process," in *International Symposia in Economic Theory and Econometrics*, ed. W. Barnett, Cambridge, U.K.: Cambridge University Press.

Huber, P., Kleiner, B., Gasser, T., and Dummermuth, G. (1971), "Statistical Methods for Investigating Phase Relations," *IEEE Transactions on Audio and Electroacoustics*, 19, 78–86.

Kim, Y. C., and Powers, E. (1979), "Digital Bispectral Analysis and Its Applications to Nonlinear Wave Interactions," *IEEE Transactions on Plasma Science*, 7, 120–131.

Lii, K. S., and Rosenblatt, M. (1982), "Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non-Gaussian Linear Processes," *The Annals of Statistics*, 10, 1195–1208.

Matsuoka, T., and Ulrych, T. J. (1984), "Phase Estimation Using the Bispectrum," *Proceedings of the IEEE*, 72, 1403–1411.

Rosenblatt, M. (1983), "Cumulants and Cumulant Spectra," in *Handbook of Statistics* (Vol. 3), eds. D. Brillinger and P. Krishnaiah, Amsterdam: North-Holland, pp. 369–382.

——— (1985), *Stationary Sequences and Random Fields*, Boston: Birkhauser.

Subba Rao, T. (1983), "The Bispectral Analysis of Nonlinear Stationary Time Series With Reference to Bilinear Time Series Models," in *Handbook of Statistics* (Vol. 3), eds. D. Brillinger and P. Krishnaiah, Amsterdam: North-Holland, pp. 293–319.

Subba Rao, T., and Gabr, M. (1980), "A Test for Linearity of Stationary Time Series," *Journal of Time Series Analysis*, 1, 145–158.