Determination of the Nyquist Frequency for Unequally Spaced Data

Melvin J. Hinich and Warren E. Weber*

Introduction

When a power spectrum is estimated from data sampled at discrete time points, the problem of aliasing may be encountered. If the data are sampled using the sampling interval τ, any frequency ω_0 outside the principal domain $-\pi/\tau \leqslant \omega \leqslant \pi/\tau$ has an alias inside the interval. Therefore, the variance component of a frequency outside this range cannot be distinguished from the variance component of a frequency within this range. The frequency π/τ is known as the Nyquist or folding frequency.

In this paper, we derive the Nyquist frequency for data that is unevenly spaced on a grid of integer multiples of a time unit, which is the case for any sampling method using real systems. Aliasing can be mathematically avoided if the time between observations can be an irrational number, as is the case for the random sampling models of random processes discussed by Shapiro and Silverman (1960) and Masry (1978).

The Nyquist Frequency of Unequally Spaced Data

Let τ be a time interval (for example, a day) and define the set of times T to be $T = \{t = m\tau | m \epsilon M\}$, where M is the set of non-negative integers $\{0,1,\ldots,n-1\}$. Let $\{y(t)\}$ be a time series defined over the times in the set

---

*The authors are both Professors of Economics, Virginia Polytechnic Institute and State University. We are indebted to VPI & SU for financial support. Professor Hinich's research was supported by the Office of Naval Research (Statistics and Probability Program) under contract.

T. Taking the Fourier transform of $\{y(t)\}$ we obtain the Fourier coefficients of $\{y(t)\}$ as

$$A_y(\omega_m) = \frac{1}{n} \sum_{t\varepsilon T} y(t)\exp(-i\omega_m t) \qquad m\varepsilon M, \tag{1}$$

where the

$$\omega_m = 2\pi m/n\tau,$$

are the Fourier frequencies for the times in the set T. Thus, by taking the inverse transform of (1) we can express $\{y(t)\}$ as

$$y(t) = \sum_{m\varepsilon M} A_y(\omega_m)\exp(i\omega_m t), \qquad t\varepsilon T. \tag{2}$$

Equation (2) is the Fourier series representation of $\{y(t)\}$.

First, we consider the aliasing problem with equally spaced data in order to acquaint the reader with our method of determining the Nyquist frequency.[1] Suppose that instead of being able to observe $\{y(t)\}$ at all times T, we are able to sample it only at intervals $\nu\tau$, where ν is a positive integer greater than unity. Define the set $T_\nu = \{t(j) = j\nu\tau | j\varepsilon J\}$, where J is the set of non-negative integers $\{0,1,\ldots,n^*-1\}$, as this set of equally spaced sample times. For convenience, let $n^*\nu=n$. Using this notation, (2) can be written as

$$y[t(j)] = \sum_{m\varepsilon J} \sum_{\ell=0}^{\nu-1} \exp[i(\omega_{m+n^*\ell})t(j)]A_y(\omega_{m+n^*\ell}), \qquad t(j)\varepsilon T_\nu. \tag{3}$$

---

[1]For an alternative determination of the Nyquist frequency for evenly spaced data see Bloomfield [1976, pp. 26-7]. Our analysis does not require that the data be from a random process.

The aliasing problem with equally spaced data occurs because for the sample times in the set T_ν, $\exp[i\omega_m t(j)]$ is a periodic function of ω_m with period $2\pi/\nu\tau$. This is demonstrated as follows:

$$\exp[i(\omega_m + 2\pi/\nu\tau)t(j)] = \exp[i\omega_m t(j)]\exp(i2\pi j) = \exp[i\omega_m t(j)], \qquad (4)$$

for all ω_m and any $t(j)\varepsilon T_\nu$. Further, (4) implies that

$$\exp[i\omega_{m+n^*\ell} t(j)] = \exp[i\omega_m t(j)] \qquad (5)$$

for any positive integer ℓ. Substituting (5) into (3), we obtain

$$y[t(j)] = \sum_{m\varepsilon J} \exp[i\omega_m t(j)] \sum_{\ell=0}^{\nu-1} A_y(\omega_{m+n^*\ell}). \qquad (6)$$

Since only the terms $\exp(i\omega_m t)$, $m\varepsilon J$ appear in (6), the Fourier series representation of $y[t(j)]$ will be strictly in terms of frequencies in the interval $-\pi/\nu\tau \leqslant \omega_m \leqslant \pi/\nu\tau$ for the times T_ν. Thus, if $\{y(t)\}$ is bandlimited to frequencies in the interval $-\pi/\nu\tau \leqslant \omega_m \leqslant \pi/\nu\tau$ (that is, if $A_y(\omega_m) = 0$, for all $m\varepsilon M - J$), the Fourier series representation of $y[t(j)]$ for times in the set T_ν will be the same as (2) and no aliasing will occur. The frequency $\pi/\nu\tau$ is the Nyquist (or folding) frequency.

The aliasing problem arises when $\{y(t)\}$ has cyclic components at frequencies outside the interval $-\pi/\nu\tau \leqslant \omega \leqslant \pi/\nu\tau$; i.e., the aliasing problem arises when $A_y(\omega_m) \neq 0$ for some $m \varepsilon M - J$. Suppose, for example, that $\{y(t)\}$ has a cyclic component at the frequency $m' = m + n^*$, $m \varepsilon M$, so that $A_y(\omega_{m'}) \neq 0$. Since the Fourier coefficients for the frequency ω_m in (6) is $\sum_{\ell=0}^{\nu-1} A_y(\omega+n^*\ell)$, the cyclic component at frequency $\omega_{m'}$ will appear as a cyclic component at the frequency ω_m for the times in the set T_ν. Further, if $\{y(t)\}$ also contains a cyclic component at the frequency ω_m, it will not be possible to distinguish $A_y(\omega_m)$ and $A_y(\omega_{m'})$ in (6). Thus, the frequencies ω_m and $\omega_{m'}$ will be aliased since the cycle of frequency $\omega_{m'}$ cannot be distinguished from the cycle of frequency ω_m for times in the set T_ν. Since the summation to obtain the Fourier coefficient of the frequency ω_m is from zero to $\nu-1$, equation (6) indicates that the same distinguishability problem will occur for cyclic components of frequency ω_m and cyclic components of all frequencies $\omega_{m+n*\ell}$, where ℓ is a positive integer. It is for this reason that the frequencies $\omega_{m+n*\ell}$, ℓ integer, are said to be aliases, and the frequency ω_m is called the principal alias.[2]

We now consider the aliasing problem with instantaneously recorded data when the data are sampled at unevenly spaced intervals. Let H denote a set of p non-negative integers including zero with largest common divisor η, and for convenience let n-1 be the largest integer in H. Further, let h(j) denote the j-th element of H and define $t(j) = h(j)\tau$. Denote the set of unequally spaced sample times by T_η, so that $T_\eta = \{t(1),\ldots,t(p)\}$.

Proceeding as we did in the case of evenly spaced data, we begin by noting that for any time $t(j)\varepsilon T_\eta$, $\exp[i\omega_m t(j)]$ is a periodic function of ω_m with period $2\pi/h(j)\tau$. This occurs since

$$\begin{aligned}\exp[i(\omega_m + 2\pi/t(j))t(j)] &= \exp[i\omega_m t(j)]\exp(i2\pi)\\ &= \exp[i\omega_m t(j)].\end{aligned}$$

---

[2]The aliasing problem can also be considered in linear regression terms. Suppose we were to estimate (4) over the times T_ν using the $\exp(i\omega_m t)$ as the regressors. Equation (3) implies that a regressor $\exp(i\omega_k t)$ whose frequency ω_k is outside the interval $-\pi/\nu\tau \leqslant \omega \leqslant \pi/\nu\tau$ will be collinear with $\exp(i\omega_k t)$ whose frequency ω_k is in the interval. Thus, the effects of these two frequencies cannot be distinguished because they will have exactly the same coefficients when either one of them is included in the regression.

Suppose, however, that $h(j)$ is divisible by $b(j)$, i.e., $h(j)/b(j) = r(j)$ is an integer. Then for all ω_m and any $t(j) \varepsilon T_\eta$

$$\exp[i(\omega_m + 2\pi/b(j)\tau)t(j)] = \exp[i\omega_m t(j)]\ \exp[i2\pi r(j)]$$

$$= \exp[i\omega_m t(j)],$$

which means that $\exp[i\omega_m t(j)]$ goes through $r(j)$ complete cycles when ω_m goes from $\overset{o}{\omega}_m$ to $\overset{o}{\omega}_m + 2\pi/b(j)\tau$. Thus, in this case $\exp[i\omega_m t(j)]$ is a periodic function of ω_m with period $2\pi/b(j)\tau$.

Now let β be a common divisor of the integers $h(1),\ldots,h(p)$. Each $\exp[i\omega_m h(j)\tau]$ goes through an integer number of cycles in the period $\overset{o}{\omega}_m \leqslant \omega \leqslant \overset{o}{\omega}_m + 2\pi/\beta\tau$ for any $\overset{o}{\omega}_m$. The shortest period in which <u>all</u> the $\exp[i\omega_m t(j)]$ will repeat is $2\pi/\eta\tau$, where η is the <u>greatest common divisor</u> of the integers $h \varepsilon H$. Thus, for the unevenly spaced times in the set T_η

$$\exp[i\omega_m t(j)] = \exp[i\omega_{m+d\ell} t(j)], \tag{7}$$

when ℓ is any positive integer and $d = (n-1)/\eta$. Substituting (7) into (2) and combining terms, we obtain

$$y[t(j)] = \sum_{m \varepsilon J^*} \exp[i\omega_m t(j)] \sum_{\ell=0}^{\eta-1} A_y(\omega_{m+d\ell}), \tag{8}$$

where $J^* = \{0,1,\ldots,d-1\}$, as the equivalent to (6) for unevenly spaced data. Equation (8) demonstrates that for the unevenly spaced times in the set T_η, $\{y(t)\}$ can be expressed solely in terms of frequencies in the interval $-\pi/\eta\tau \leqslant \omega \leqslant \pi/\eta\tau$. Thus, the Nyquist frequency is $\pi/\eta\tau$.

References

Bloomfield, Peter J., Fourier Analysis of Time Series: An Introduction, John Wiley & Sons, New York, 1976.

Masry, E., "Alias-free sampling: an alternative conceptualization and its applications," IEEE Trans. Inform. Theory, 24, 317-324 (1978).

Shapiro, H. S. and R. A. Silverman, "Alias-free sampling of random noise," J. Soc. Indust. Appl. Math., 8, 225-248 (1960).

November, 1979

Aliasing and Unequally Spaced Observations: Or How To Distinguish Weekly Cycles in Monthly Data

Melvin J. Hinich and Warren E. Weber*

When Fourier analysis is applied to data sampled at discrete time points, the problem of aliasing may be encountered. If the data are sampled using the sampling interval τ, any frequency ω_o outside the interval $-\pi/\tau \leq \omega \leq \pi/\tau$ has an alias inside the interval. Therefore, the effects of a frequency outside this range cannot be distinguished from the effects of a frequency within this range. The frequency π/τ is known as the Nyquist or folding frequency.

An example will serve to illustrate the aliasing problem. Consider the case of "monthly" data sampled at intervals of 30 days. The Nyquist frequency for such data will be $\pi/30$ radians/day, which corresponds to a period of 60 days. The aliasing problem with such data is, then, that any frequency not in the interval $-\pi/30 \leq \omega \leq \pi/30$ will have an alias in this interval. To put it another way, any cycle with a period of less than 60 days will have an alias at a period of more than 60 days.

In particular, suppose that there exists a 7 day weekly cycle in these data. The frequency for such a cycle is $2\pi/7 > \pi/30$. To find its alias we subtract integer multiples of $2\pi/30$ until we obtain a frequency ω' such that $-\pi/30 \leq \omega' \leq \pi/30$. Thus, the alias of $2\pi/7$ is

$$\omega' = 2\pi/7 - 4(2\pi/30) = 2\pi/105 \text{ rads/day.}$$

---

*The authors are both Professors of Economics, Virginia Polytechnic Institute and State University. We are indebted to Francis Ahking for performing the computations presented in this paper. Professor Hinich's research is supported by the Office of Naval Research, Statistics and Probability Program, under contract.

The period of the alias is 105 days, so that 7 day cycles cannot be distinguished from 105 day cycles in data sampled at 30 day intervals.

However, all months are not of equal length in terms of days. This raises the question of what is the Nyquist frequency in the case of unevenly spaced data. In this paper, we derive the Nyquist frequency for the case of unevenly spaced sample observations. We then show how the unequal spacing of sample observations raises the Nyquist frequency and mitigates some of the aliasing problems. In this way, we demonstrate how Fourier analysis can be used to study cycles with periods as short as two days in monthly data.

The paper proceeds as follows. In the first section we present an analysis of the problem of aliasing for the case of data sampled at evenly spaced intervals and then for the case of data sampled at unevenly spaced intervals. In this section we also present a method for determining the Nyquist frequency for data sampled at uneven intervals. In the second section we apply our method to monthly data for commercial bank cash assets. In the third section we present some evidence on the problem of leakage in periodogram analysis of unequally spaced data.

I. Sampling Periodic Functions at Unequally Spaced Times

A function of time f(t) is periodic with period p if f(t + p) = f(t) for all time values t. If f(t) is continuously differentiable, it has the following Fourier series representation:

$$f(t) = a_0 + 2 \sum_{k=1}^{\infty} a_k \cos \omega_k t + 2 \sum_{k=1}^{\infty} b_k \sin \omega_k t \tag{1}$$

where $\omega_1 = 2\pi/p$ is called the fundamental (angular) frequency, $\omega_k = k\omega_1$ is called the (k-1)st harmonic of the fundamental, and

$$\begin{aligned} a_k &= \frac{1}{p} \int_0^p f(t) \cos \omega_k t \, dt \\ b_k &= \frac{1}{p} \int_0^p f(t) \sin \omega_k t \, dt \end{aligned} \tag{2}$$

are the Fourier coefficients. For many periodic functions, the a_k and b_k are very small for k greater than a relatively small integer. Let us then approximate f(t) by the finite sum $a_0 + 2 \sum_{k=1}^{M} a_k \cos\omega_k t + 2 \sum_{k=1}^{M} b_k \sin\omega_k t$ where, once again, $\omega_k = 2\pi k/p$. Switching to complex variable notation to simplify our explanation of aliasing,

$$f(t) \approx \sum_{k=-M}^{M} A_k e^{i\omega_k t} \tag{3}$$

where $$A_0 = a_0 \tag{4}$$

and $$A_k = a_k + ib_k \quad \text{for} \quad |k| = 1, \ldots, M.$$

When the function f(t) is real,

$$A_k = A^*_{-k} \qquad k > 0 \tag{5}$$

where * denotes the complex conjugate. This result follows by direct substitution into (2) and (4).

Suppose we observe $y(t) = f(t) + \varepsilon(t)$ at times $t_1, \ldots, t_n$, where $\varepsilon(t)$ is a stationary zero mean noise process. Since the origin of the time scale is arbitrary, let us use the convenient convention that zero time is at the time of the first observation, i.e. $t_1 = 0$. Ignoring the approximation error in (3), it follows that

$$y(t_j) = \sum_{k=-M}^{M} A_k e^{i\omega_k t_j} + \varepsilon(t_j). \tag{6}$$

Thus the $y(t_j)$ satisfy a linear model whose 2M independent variables are $\{e^{i\omega_k t_j} \mid k=\pm 1, \ldots, \pm M\}$.[1] The intercept is $A_0 = a_0$. If the data are equally spaced in time, e.g. $t_j = jN\tau$ for some integer N and interval $\tau > 0$, then these independent variables are <u>orthogonal</u> if the sampling period $nN\tau$ is an integer multiple of the period p (and $N\tau < p$).[2] If the spacing is unequal, the regressors are in general non-orthogonal.

An example will provide the rationale for including N in the above discussion. Suppose that we allow τ to be a day, but the analysis is conducted using weekly data. In this case, $N = 7$, the Nyquist frequency is $\pi/7$ days, and the cycle corresponding to Nyquist frequency is 14 days. Obviously, we would obtain the same results if we had allowed τ to be one week and set N=1. However, allowing the sampling frequency and the sampling interval to be different will be useful below.

First, consider the aliasing problem when $t_j = jN\tau$. In this case $e^{i\omega t_j}$ is a periodic function of ω with period $2\pi/N\tau$ since for all ω

$$e^{i(\omega + 2\pi/N\tau)t_j} = e^{i\omega t_j} e^{i2\pi j}$$
$$= e^{i\omega t_j} . \tag{7}$$

Thus the sequence $\{e^{i\omega t_j} | j = 1,\ldots,n\}$ is identical to the sequence $\{e^{i(\omega + 2\pi/N\tau)t_j} | j = 1,\ldots,n\}$. This implies that a regressor $e^{i\omega_k t_j}$ whose frequency ω_k is outside the interval $-\pi/N\tau \leq \omega \leq \pi/N\tau$ will be collinear with $e^{i\tilde{\omega}_k t_j}$ whose frequency $\tilde{\omega}_k$ is in the interval. The frequency $\tilde{\omega}_k$ is called the alias of ω_k.

This collinearity due to aliasing is no problem if the frequency ω_k is <u>known</u> by the investigator, and the noise $\{\varepsilon(t_j)\}$ is approximately white. For example, suppose that we know that there is a 7 day cycle in data sampled every thirty days. As long as there is no cycle at the 105 day alias, regressing the data on $e^{i(2\pi t_j/7)}$ and its harmonics yields unbiased estimates of the amplitudes of the cycles. However, if there is a cycle at 105 days, it <u>confounds</u> the 7 day cycle. Further even if there is no coherent 105 day cycle, the 7 day cycle will be obscurred if the noise spectrum has a broad peak at the 105 day frequency, i.e., the t-statistics for $\cos 2\pi t_j/7$ and $\sin 2\pi t_j/105$ may be insignificant due to the noise.[3]

The 105 day cycle, however, is the alias of the weekly cycle only for a thirty day sampling interval. Since many months have thirty one days and February is short, in general, monthly data will not be equally spaced. Let us now examine the aliasing problem for unequally spaced data. It will turn out that the weekly cycle is not aliased if the sampling intervals vary among 28, 30, and 31.

To consider the aliasing problem with unequally spaced data, suppose that we sample f(t) at times $t_1 = m_1\tau, \ldots, t_n = m_n\tau$, where each m_j is an integer and $m_1 = 0$. For each j, $e^{i\omega t_j}$ is a periodic function of ω with period $\frac{2\pi}{m_j\tau}$, since for all ω

$$e^{i(\omega + \frac{2\pi}{t_j})t_j} = e^{i\omega t_j} e^{i2\pi}$$

$$= e^{i\omega t_j} .$$

Suppose that m_j is divisible by d_j, i.e., $m_j/d_j = r_j$ is an integer. Then for all ω

$$e^{i(\omega + \frac{2\pi}{d_j\tau})t_j} = e^{i\omega t_j} e^{i2\pi r_j}$$

$$= e^{i\omega t_j} ,$$

which means that $e^{i\omega t_j}$ goes through r_j complete cycles when ω goes from ω_o to $\omega_o + 2\pi/d_j\tau$.

Now let D be a common divisor of the integers $m_1,\ldots,m_n$. Each $e^{i\omega m_j\tau}$ goes through an integer number of cycles in the period $\omega_o \leq \omega \leq \omega_o + 2\pi/D\tau$ for any ω_o. Thus, the sequence $\{e^{i\omega t_j} | j=1,\ldots,n\}$ is identical to the sequence $\{e^{i(\omega + \frac{2\pi}{D\tau})t_j} | j=1,\ldots,n\}$. The shortest period where all the $e^{i\omega t_j}$ repeat is $\frac{2\pi}{N\tau}$, where N is the greatest common divisor of $m_1,\ldots,m_n$. Then following the reasoning above, the Nyquist frequency becomes $\pi/N\tau$. Equally spaced observations then is just the special case in which the largest common divisor is the constant interval N between samples.

II. Application to Economic Time Series

In this section we illustrate the method developed by applying it to the economic time series for commercial bank cash assets. Data on commercial bank cash assets were collected for the period January, 1948 to May, 1979 for a total of 377 observations.[4] The uneven spacing for these data arise because they are collected for the last Wednesday of the month except for June and December when they are for the last day of the month.[5] We let the sampling interval τ be one day. Then defining $\Delta m_j = m_j - m_{j-1}$, we find that Δm_j takes on the integer values 26-30 and 34-37 for these data. Thus, it is obvious that the largest common divisor of $m_1,\ldots,m_{377}$ is unity, and the Nyquist frequency is π meaning that the shortest cycle which can be observed is that with a length of two days.

In order to examine the periodicities in the bank cash asset data, we first extracted the mean and time trend from the data using a linear regression with cash assets regressed on a constant, time, time squared and time cubed. Time was measured in calendar days beginning with January 1, 1948.

The regressions reported in Table 1 test for the presence of weekly cycles in the bank cash asset data. In equations (1) and (2) we find that the coefficients of both $\cos(2\pi t/7)$ and $\sin(2\pi t/7)$ are significantly different from zero at the 0.01 level of significance. In equation (3) the first harmonic of $2\pi/7$ is added. We find that only the coefficient of $\sin(4\pi t/7)$ is significantly different from zero at even the 0.05 level. However, the large standard errors of the $\cos(2\pi t/7)$, $\sin(2\pi t/7)$ and $\cos(4\pi t/7)$ terms are due to the high multicollinearlity between the four sine and cosine terms at the frequencies $2\pi/7$ and $4\pi/7$.[6] The F-statistic for the test of the

TABLE 1

Regression Results - Bank Cash Assets

Equation	(1)	(2)	(3)	(4)	(5)
Constant	7.400** (0.7897)	3.888** (1.591)	7.300** (0.7263)	3.895** (1.462)	-0.0077 (0.2737)
$\cos 2\pi t/7$	-6.266** (1.052)	-0.6095 (2.428)	-5.889** (0.9687)	-0.4237 (2.232)	
$\sin 2\pi t/7$	5.717** (0.6089)	-1.039 (2.675)	5.917** (0.5606)	-0.6012 (2.459)	
$\cos 4\pi t/7$		0.9909 (1.150)		0.8332 (1.057)	
$\sin 4\pi t/7$		5.311** (2.067)		5.138** (1.901)	
$\cos 4\pi t/11508$			-2.755** (0.3309)	-2.747** (0.3286)	-2.685** (0.3877)
$\sin 4\pi t/11508$			0.3372 (0.3294)	0.3048 (0.3272)	0.4338 (0.3866)
Σe^2	0965.70	8902.42	7623.6	7473.0	10564.8
S.E.E.	4.923	4.892	4.527	4.494	5.315
F (degrees of freedom)	59.60*** (2,374)	31.89*** (4,374)	52.84*** (4,372)	36.98*** (6,370)	24.61*** (2,374)
R^2	0.24	0.26	0.36	0.37	0.12
D.W.	1.38	1.36	1.64	1.61	1.77

Standard errors in parenthesis

* significant at 0.05 level (two-tailed)

** significant at 0.01 level (two-tailed)

*** significant at 0.001 level (two-tailed) null hypothesis that the coefficients of these four terms are zero against the alternative that at least one is nonzero is 31.89, which indicates that the null hypothesis can be rejected at better than the 0.001 level of significance. A regression was also run including the second harmonic of $2\pi/7$ as well. The high multicollinearity between the six sine and cosine terms in this case caused all individual standard errors to be extremely large. However, the F-statistic for this regression was 20.79, which once again indicates that the null hypothesis of no weekly cycle can be rejected at the 0.01 level. Equations (3) and (4) perform the same tests including a long term cycle $4\pi/11508$ which did not appear to have been eliminated by our detrending procedures. The tests on the $2\pi t/7$ and $4\pi t/7$ terms in these regressions are similar to those in (1) and (2). Thus, the regression evidence indicates that weekly cycles are present in bank cash asset holdings.[7]

In order to check on the weekly cycle shown in the regression results, we also took the periodogram of the detrended bank cash asset series to determine whether or not the periodogram would exhibit a spike at the frequency of the weekly cycle or any of its harmonics. Specifically, we took the periodogram of the series

$$x_t = \begin{cases} b_t & \text{if } t = t_j, \quad j=1,\ldots,377 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $t=1,\ldots,11508$ and b_t denotes the detrended level of bank cash assets at time t.[8] Thus, x_t is a (calendar) daily series on bank cash assets with a zero assigned to those days for which we did not have an observation.

Since 11508 is evenly divisible by 7, the Fourier frequency $2\pi(1644)/11508 = 2\pi/7$ will have a period of exactly seven days.

The periodogram of the detrended bank cash series is presented in Figure 1. Examination of the figure as well as the raw data used to generate it shows that there are relative peaks in the periodogram at the frequency $2\pi(1644)/11508$ as well as at its first and second harmonics.[9] Thus, the evidence from the periodogram supports the previous conclusion that weekly cycles do exist in bank cash asset holdings.

III. Leakage

The leakage problem in Fourier analysis occurs when the presence of a particular harmonic component in a data series causes the values of the Fourier coefficients and hence the values of the periodogram to be nonzero at other frequencies. Our motivation for examining the leakage problem is the presence of peaks in the periodogram of the bank cash asset data at what might be considered unusual frequencies. For example, peaks in this periodogram occur at frequencies $2\pi(1134)/11508$, $2\pi(2400)/11508$, $2\pi(2910)/11508$, and $2\pi(3666)/11508$ to list a few. These frequencies correspond to periods of 10.148, 4.795, 3.955, and 3.139 days, respectively. We wish to determine if these peaks could be caused by the leakage from some other harmonic in the bank cash asset series.

There are two standard cases of the leakage problem in the case of evenly spaced data. The first is when there exists a harmonic in the data series which is not at a Fourier frequency. The second is the case in which the data record is extended by zeroes before the Fourier transform is calculated. In each of these cases the leakage has the form of a Dirichlet function. We now examine the question of whether leakage will occur in the case of unevenly spaced data and, if so, what will be its nature.

To answer the question of whether or not leakage will occur with unevenly spaced data, consider the data series

$$y_t = \begin{cases} \cos(\omega_m t) & t = t_j \\ 0 & \text{otherwise}, \end{cases} \tag{9}$$

where the t_j are those defined for the bank asset series and ω_m is any of the Fourier frequencies used in the analysis of the bank cash asset data in the previous section. Then we know that the Fourier coefficients for Fourier expansion (1) of this data series will be

$$a_k = [\sum_j \cos(\omega_m t_j)\cos(\omega_k t_j)]\frac{2}{n} \qquad k \neq n/2$$

and

$$b_k = [\sum_j \cos(\omega_m t_j)\sin(\omega_k t_j)]\frac{2}{n}, \qquad k \neq 0,\ n/2$$

where $\omega_k = 2\pi k/11508$ rad/day. It is obvious by inspection that a_k and b_k will not in general zero when $k \neq m$. Thus, it is obvious that there will be leakage at frequencies other than ω_m. Further, it is obvious that this leakage will not in general have the form of a Dirichlet function, but instead for the data series y_t the leakage will be large at those frequencies ω_k at which the vector $\cos(\omega_m t_j)$ is highly correlated with either the vector $\cos(\omega_k t_j)$ or the vector $\sin(\omega_k t_j)$. To put it another way, the leakage for the data series y_t will occur at those frequencies at which the vectors $\cos(\omega_m t_j)$ and $\cos(\omega_k t_j)$ or the vectors $\cos(\omega_m t_j)$ and $\sin(\omega_k t_j)$ have a high degree of multicollinearity.

In order to examine the leakage associated with the series y_t, its periodogram is presented as Figure 2 when $\omega_m = 2\pi/7$. A comparison of this periodogram and that for bank cash assets indicates that the peaks at unusual frequencies in the periodogram of the bank cash asset series occurs at leakage peaks from a cosine with a period of 7 days.

Of course, two additional points must be made with regard to the leakage problem. The first is that since there is a peak in the periodogram of $\cos(2\pi t/7)$ at the frequency $2\pi(2910)/11508$ means that a high degree of

Figure 1

Periodogram, Bank Cash Assets

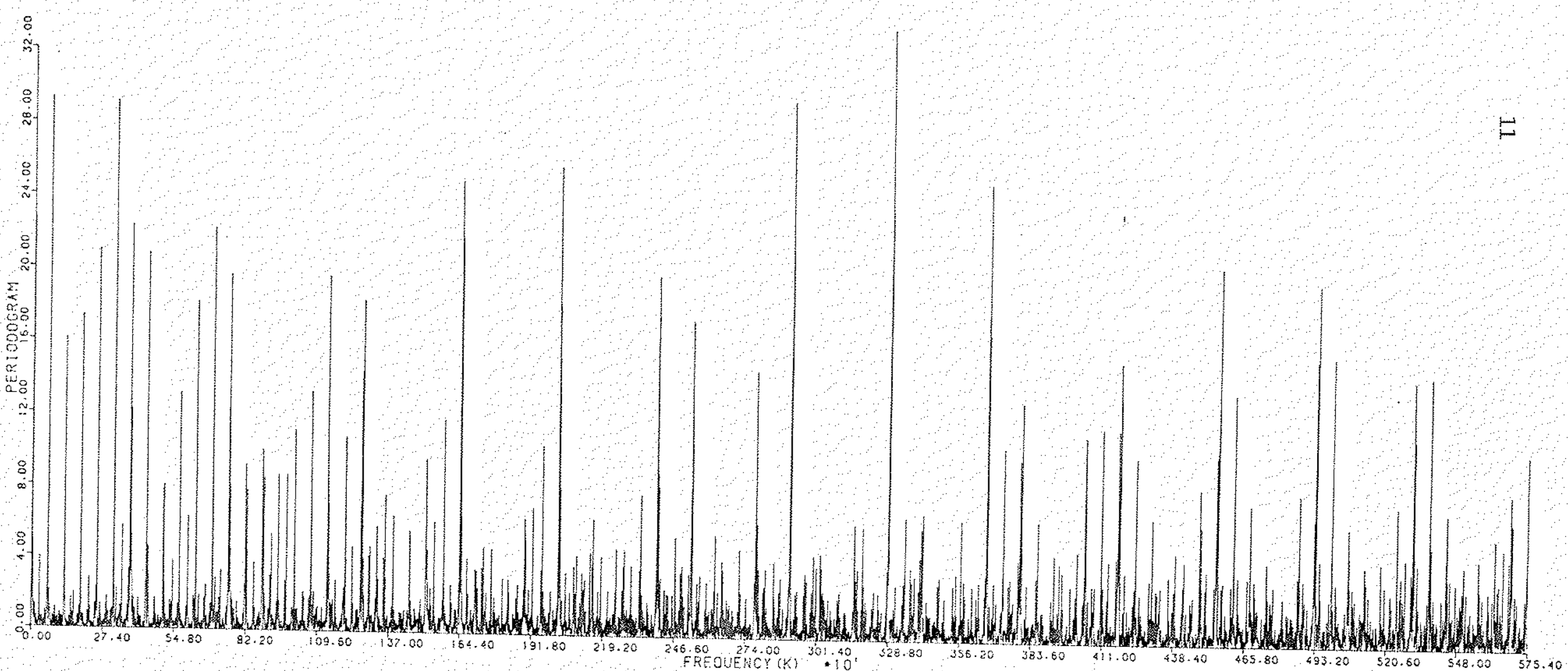
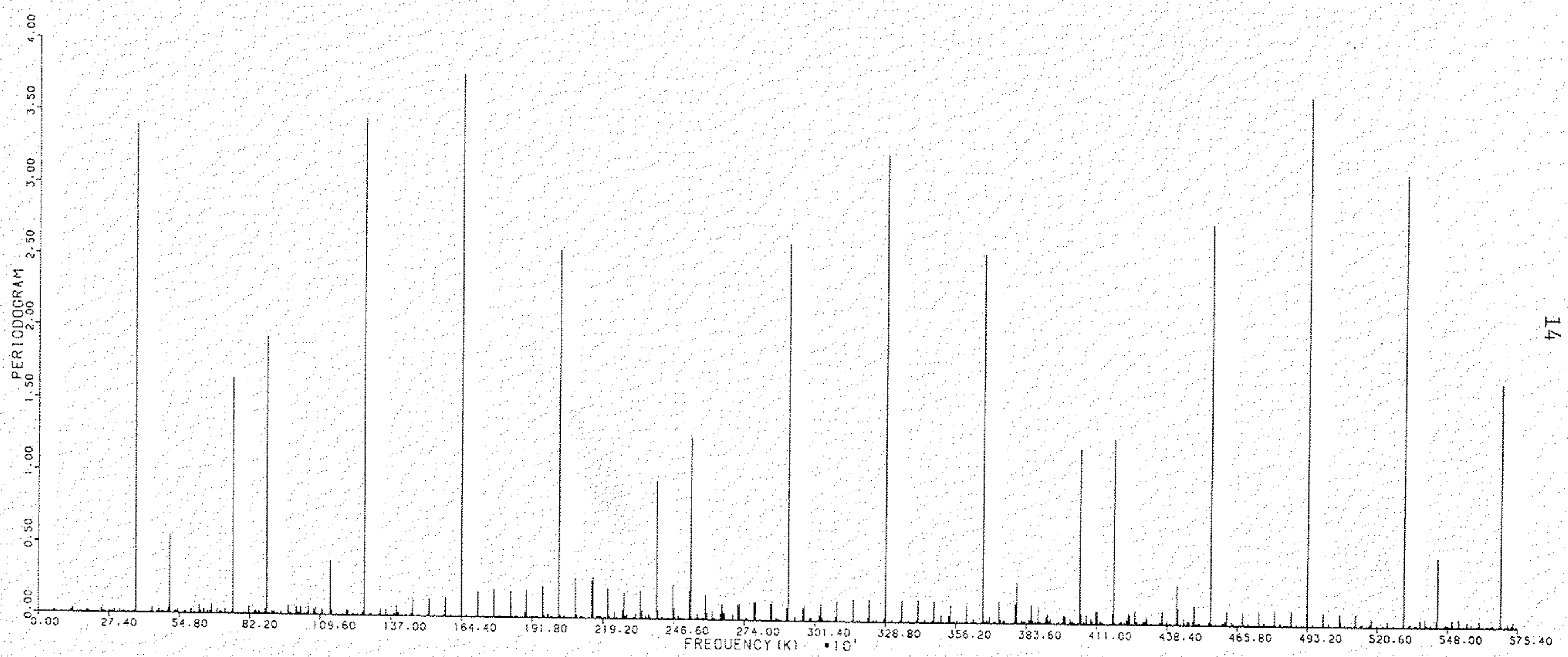


Figure 2

Periodogram, $\mathrm{Cos}(2\pi t_j/7)$



multicollinearity exists between the vector $\cos(2\pi t_j/7)$ and either the vector $\cos(5820\pi t_j/11508)$ or the vector $\sin(5820\pi t_j/11508)$. Thus, if we were to replace the $\cos(2\pi t/7)$ and $\sin(2\pi t/7)$ terms in regression equations (3) and (4) with the terms $\cos(5820\pi t/11508)$ and $\sin(5820\pi t/11508)$ we would not expect a significant loss of explanatory power to occur. Thus, we could not determine whether, in fact, the cycle in the data were a 7 day cycle or an 11508/2910 = 3.955 day cycle. However, this is where knowledge of the institutions is important. There are good reasons to expect a 7 calendar day cycle in bank cash asset holdings since reserve requirements must be met on Wednesdays and many individuals are paid on Fridays and want to cash pay checks at this time. No such reasons exist for a 3.955 day cycle.

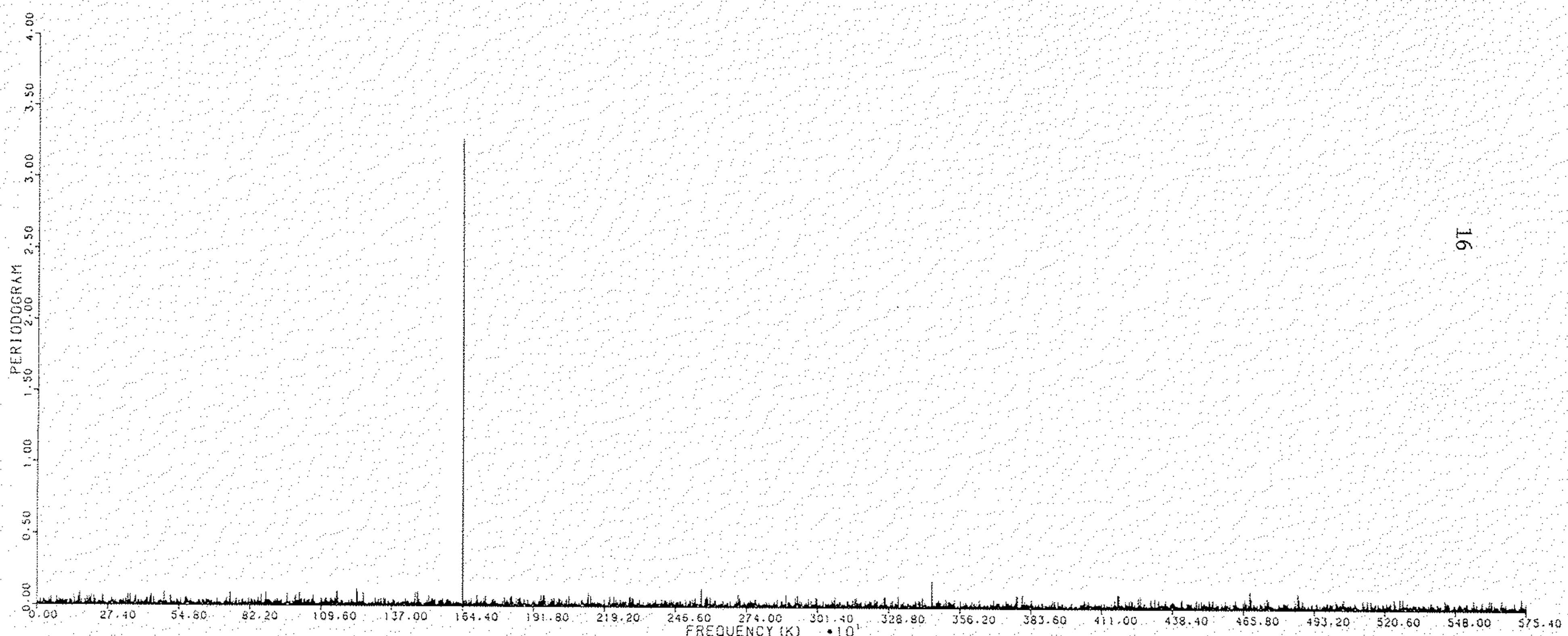
The second major point is that the frequencies at which the leakage would occur will vary with the exact sequence of the t_j's. For another data series which was sampled at different times, the leakage pattern would be entirely different. Thus, an investigator working with unevenly spaced data should construct series like y_t above only using the t_j for his sample and then run periodograms of this series in order to determine the leakage pattern associated with his particular uneven spacing. The leakage pattern will also vary as the frequency ω_m is changed in the y_t series.

To check on this latter point, we also ran the periodogram of a y_t (see (9)) series with $\omega_m = 2\pi/7$, but with the t_j randomly sampled. The periodogram for this series is presented as Figure 3 and indicates that the leakage found in Figure 2 has disappeared. This confirms our conjecture that the leakage pattern observed in Figure 2 occurs because of the sampling pattern associated with the bank cash asset data, and that it would disappear if these data had been randomly sampled unevenly spaced observations instead.

Figure 3

Periodogram, $\text{Cos}(2\pi t_j/7)$

Random Sample



Footnotes

1. In terms of real functions, the independent variables are $\cos\omega_k t_j$ and $\sin\omega_k t_j (k = 1,\ldots,M)$.

2. This follows from the result that

$$\sum_{j=0}^{n-1} \exp(i2\pi \frac{kj}{n}) = \frac{1 - \exp(i2\pi k)}{1 - \exp(i2\pi \frac{k}{n})}$$

$$= 0 \quad \text{for any integer } k, \ |k| < n.$$

3. Most noise processes have considerable low frequency energy (variance) due to nonstationarity of the mean. This low frequency variation will swamp many of the high frequency aliases.

4. Included in bank cash assets are "Currency and Coin," "Reserves with Federal Reserve Banks," "Reserves with Banks," and "Cash Items in Process of Collection." The sources for these data are Banking and Monetary Statistics, 1941-70, Annual Statistical Digest, and the Federal Reserve Bulletin.

5. This is true except for the years 1957-1960 when the last Wednesday in June was used. We are indebted to Mary Harrington of the Board of Governors of the Federal Reserve System for this information.

6. We can illustrate the multicollinearlity problem using $\cos(2\pi t/7)$ and $\cos(4\pi t/7)$ terms. Because of the nature of the cosine function, the values of $\cos(2\pi t_j/7)$ will be equal whenever Δm_j equals 26, 28, 33 or 35. This is true in all but 22 of the 377 observations on bank cash asset holdings. The same is true for $\cos(4\pi t_j/7)$. Hence, $\cos(4\pi t_j/7)$ does not contain much information which is not already contained in $\cos(2\pi t_j/7)$ and hence we have a multi-collinearlity problem between these two series.

7. The question may arise as to whether our regression results may be biased due to the omission of some other relevant frequencies in our estimations. Of course, if we were working with evenly spaced observations, no such bias would exist since the time series of the sines and cosines at the Fourier frequencies are orthogonal. This is not necessarily true for the case of unevenly spaced observations. However, to the extent that the time series of the sines and cosines at the Fourier frequencies are nearly orthogonal, the bias due to the omission of relevant variables will be small.

8. This notation is different from that used by Parzen, E. (1963), "On spectral analysis with missing observations and amplitude modulation." Sankya, A. 25, 383-392. Also see Clinger, W. and Van Ness, J. (1976), "On unequally spaced time points in time series." Ann. Statist. 4, 736-745. The length of the record was made 11508, so that the Fourier decomposition would include frequencies corresponding to 2 day, 3 day, and 7 day cycles.

9. The other striking feature of the periodogram of this series is the peak at the frequency $2\pi(63)/11508$ and at several of its harmonics. This frequency corresponds to a period of 182.67 days, so that the data indicate that there is a half yearly cycle in bank cash asset holdings. The strong first harmonic of this frequency also indicates the presence of a cycle of 91.33 days which would approximate a quarterly cycle. Regression similar to those reported as equations (3), (4) and (5) in Table 1 were run including $\cos 2\pi t(63)/11508$, $\sin 2\pi t(63)/11508$ and their first harmonics. These results confirmed the presence of cycles at these frequencies and left the conclusions regarding the presence of the weekly cycle unchanged.