CHAPTER 2

Probing the Underlying Structure in Dynamical Systems: An Introduction to Spectral Analysis

Michael McBurnett

This chapter examines one method of investigating the underlying dynamics of time-series data. It introduces the reader to spectral analysis, a tool for evaluating the frequency properties of a time series. This is distinguished from the analysis of the properties of time series in the time domain, which is the subject of most recent research in time-series analysis. Since this discussion serves as an introduction to the topic, I have eschewed mathematical rigor.

In this chapter, I focus on three basic types of time series—periodic time series, chaotic time series, and a random time series—and explore their frequency properties with numerous examples. I also introduce noise into some of the dynamic processes to illustrate how noise affects spectral analysis. In the final section of the chapter, I introduce a time series whose properties are unknown and analyze its frequency properties. This time series is constructed from survey data collected by the Center for Political Studies.

I also introduce some common problems that research has uncovered in the spectral analysis of time series, particularly as these problems relate to the analysis of chaotic time series. These problems involve the inability of spectral analysis to clearly discern cycles, even when they are known to be present. Sometimes this is a data problem: the time series is too short or the signal-to-noise ratio is too high. Other times the signal resembles noise even though it is not. Unfortunately, the latter case is the more critical problem and the more likely problem when dealing with chaotic processes.

The chapter is organized as follows. First, I provide the necessary mathematical introduction to the spectral density and spectral distribution functions. I show how to interpret these measures and explain some of the problems in their use. Then I demonstrate by example how to use spectral analysis and introduce two specific cases: pure noise and periodic. Here, I show how the presence of noise can interfere with the resolving power of the technique. Third, I introduce several known chaotic time series and examine their respective spectra. This section shows that no two time series have identical spectra and that there is no single characteristic spectrum for chaotic dynamics. Fourth, I introduce an empirical time series constructed from surveys conducted during the 1984 Democratic nomination race. These data are subjected to spectral analysis and compared to some of the spectra from the prior section. I determine that spectral analysis is not conclusive in determining the dynamics of this time series. I conclude with a discussion of the use of spectral analysis and suggest the use of some additional measures that can be used to quantify chaos if it is present in a time series.

Mathematical Background

Inspection of a time series may lead one to suppose that it contains a periodic oscillatory component with a known wavelength. This can be represented by

$$Y_t = R \cos (\omega t + \theta) + E_t \tag{1}$$

where ω is the frequency of the oscillation, R is known as the amplitude of the oscillation, θ is the phase, and E_t is a stationary random series. Sometimes the frequency, f, is expressed as $f = \omega/2\pi$, which is a measure of the number of cycles per unit time and is easier to interpret. I use this expression throughout in the interpretation of data. The period of a cycle, called the wavelength, is given by $1/f$ or $2\pi/\omega$. Figure 2.1 shows a graph of a time series with $f = 1/6$ and wavelength 6.

Equation 1 is extremely simple and, in practice, the variation in an observed time series may be caused by variation at several different frequencies. For example, presidential popularity may show variation at yearly, quarterly, monthly, and even weekly frequencies. This means that the series shows variation at high (weekly), medium (monthly or quarterly), and low (yearly) frequencies. Equation 1 can be generalized to account for the combination of variation in the observed series by

$$Y_t = \sum_{j=1}^{k} R_j \cos (\omega_j t + \theta_j) + E_t, \tag{2}$$

where R_j is the amplitude at frequency ω_j.

In the two equations shown thus far, it should be noted that neither is stationary if R, θ, $\{R_j\}$, and $\{\theta_j\}$ are constants since this condition implies that $E(Y_t)$ will shift over time. It is customary to assume that $\{R_j\}$ are uncorrelated random variables uniformly distributed on $(0, 2\pi)$, which are fixed for a particular value of the process (Chatfield 1992). This assumption allows the

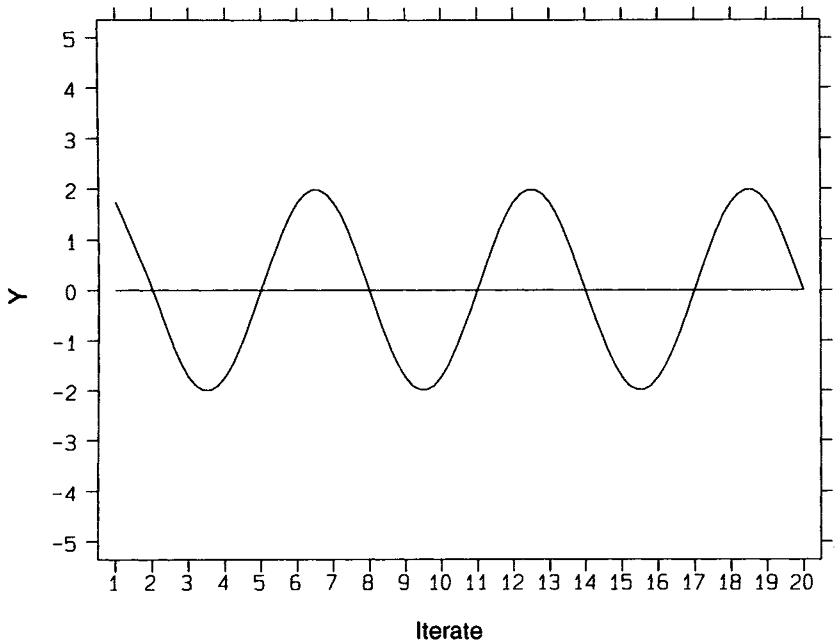

Fig. 2.1. Sample sinusoidal function with wave length 6 and frequency 1/6

treatment of time series suspected to contain more than one oscillating component as a stationary series.

Using the trigonometric identity $\cos(\omega t + \theta) = \cos \omega t \cdot \cos \theta - \sin \omega t \cdot \sin \theta$, equation 2 can be written as

$$Y_t = \sum_{j=1}^{k} (a_j \cos \omega_j t + b_j \sin \omega_{jt}) + E_t, \tag{3}$$

where $a_j = R_j \cos \theta_j$ and $b_j = -R_j \sin \theta_j$. Clearly, from equations 2 and 3 we see that only a finite number of frequencies are represented here (the index j counts only from 1 to k). Why are there not more, indeed, an infinite number of frequencies? Wiener (1949) showed that when $k \rightarrow \infty$, any discrete process measured at unit intervals can be represented as

$$Y_t = \int_0^{\pi} \cos \omega t \, du(\omega) + \int_0^{\pi} \sin \omega t \, dv(\omega) \tag{4}$$

L. Douglas Kiel and Euel Elliott, Editors
http://www.press.umich.edu/titleDetailDesc.do?id=14623
The University of Michigan Press, 1996

where $u(\omega)$ and $v(\omega)$ are uncorrelated continuous processes defined for all ω in the range $(0, \pi)$. This equation is called the spectral representation of the process. Intuitively it helps to think of Y_t as a linear combination of orthogonal oscillating terms.

One may also wonder why the upper limit of the integrals in equation 4 is π rather than $+\infty$. If the process were continuous, the limit would be $+\infty$. We are concerned with a *discrete* process measured at unit intervals.[1] Hence, there is no loss of generality in restricting ω to the range $(0, \pi)$, because

$$\cos(\omega + k\pi)t = \begin{cases} \cos \omega t & k, t \text{ integers where } k \text{ is even.} \\ \cos(\pi - \omega)t & k, t \text{ integers where } k \text{ is odd.} \end{cases}$$

Variation at frequencies higher than π cannot be distinguished from variation in a corresponding frequency in the interval $(0, \pi)$. The frequency $\omega = \pi$ is called the Nyquist frequency. If the change in the unit interval over which measurements of the time series are taken is given by Δt, then the Nyquist frequency is given by $\pi/\Delta t$. The introduction of the spectral representation of the time series serves to demonstrate that all frequencies in the interval $(0, \pi)$ may contribute to the variation in the observed series.

The Spectral Density and Spectral Distribution Functions

The power spectral distribution function, denoted $F(\omega)$, is the empirical measure that describes how the oscillations in a time series, if they are present, are distributed with respect to the frequency. This measure is the one most frequently consulted when examining a time series for a periodic component.

Suppose we have a stochastic stationary process. Then the spectral distribution function is a continuous, monotone increasing, bounded function over the interval $(0, \pi)$. The function may be differentiated with respect to ω in $(0, \pi)$. The derivative is denoted $f(\omega) = dF(\omega)/d\omega$. This is the (power) spectral density function, which is simply referred to as the spectrum.

When the derivative exists, and we have a stationary stochastic process with an autocovariance function $\gamma(k)$, we can express the relationship between the autocovariance function and the spectral distribution function as

$$\gamma(k) = \int_0^\pi \cos \omega k \, f(\omega) \, d\omega. \tag{5}$$

Set $k = 0$ and we have

$$\gamma(0) = \sigma_y^2 = \int_0^{\pi} f(\omega)\, d\omega = F(\pi). \tag{6}$$

The physical meaning of the spectrum is that $f(\omega)\, d\omega$ represents the contribution to the variance of components with frequencies in the range $(\omega, \omega + d\omega)$. When the spectrum is drawn, the total area under the curve given by equation 6 is equal to the variance of the process. A peak in the spectrum indicates an important contribution to variance in the appropriate interval. In practice one searches for peaks in the spectrum when it is drawn, as this is an indication that periodicity is present in the series.

The integral in equation 6 has another use in analysis. We are actually dealing with a finite number of frequencies in the interval $[0, \pi]$. Summing the contribution to the variance of each of these frequencies produces a characteristic shape. For a simple sinusoidal function with a single wavelength, such as that given by equation 1, one observes a single peak that translates into a step function. This is seen by using equation 1 to generate some data, say 2,500 iterates, and computing its spectrum. Then if one computes the running sum of the contribution to the variance of the spectral density function and plots that sum against the cycles per iteration, the result will have a single step. More generally, for a series with two incommensurate (nonoverlapping or orthogonal) cycles, the result will have two steps separated by the distance corresponding to their respective frequencies, each of which provides an important contribution to the variance of the process. Continuing, a process that exhibits n cycles will show n steps and each step will contribute only a small amount to the total variance of the process. Figure 2.2 shows how this phenomenon translates into problems when the number of cycles present is large.

This figure shows five selected spectral density functions displayed as described by equation 6 above. The equations used here are: a simple periodic function (equation 1), an equation which contains a four-cycle (equation 7 following), the Lorenz equations (equation 8 following), a chaotic version of the logistic equation (equation 7 with its driving parameter set to 3.99), and a random time series with mean zero.

The variance of the spectrum is given by the sum over all the frequencies present between 0 and π. Since all computer implementations of spectral analysis report discrete frequencies, the maximum number of which is given by one-half the length of the time series, the procedure for computing equation 6 is to sum the spectral density function over all cycles. All the examples used in figure 2.2 use time series with length 2,500. The vertical axis in this figure is arbitrary; all the spectral densities have been normalized to match the scale for the random time series to allow comparison of their respective shapes.

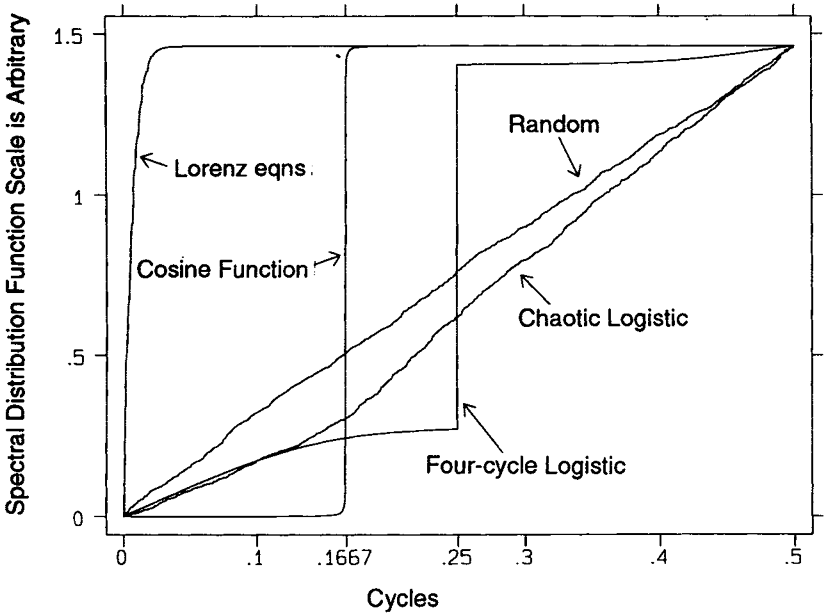

Fig. 2.2. Comparison of spectral distribution functions for five selected dynamic equations

The distinctions are quite dramatic. The spectral density for equation 1 is shown as a step function, which means that all the variance in the spectrum is accounted for by a single peak. This is also the case for the logistic equation with the driving parameter at 3.5, which produces a four-cycle. Its spectral distribution function has a single peak at 0.25 cycles per iteration and a corresponding step in this figure at 0.25 cycles. The Lorenz equations produce a very different spectral density, which is mainly confined to the left region of this figure. This figure shows that the Lorenz equations are dominated by a number of short (high) frequencies spread over a relatively narrow range, which accounts for all the variance. The two remaining equations are the chaotic version of the logistic equation and a random time series. These functions have spectral density functions that are indistinguishable from one another. Both show that all frequencies contribute about the same amount to the total variance in the time series. The chaotic logistic equation has frequencies of all cycles present in the series and these cycles are uniformly distributed across the entire spectrum. This makes it is easy to see why the random series and this equation look similar in this graph: each frequency contributes about the same amount to the total variance. This is a serious impediment to the sole use of this technique for attempting to identify chaotic dynamics

L. Douglas Kiel and Euel Elliott, Editors
http://www.press.umich.edu/titleDetailDesc.do?id=14623
The University of Michigan Press, 1996

(Glass and Mackey 1988). It shows that deterministic chaos may not be distinguished from noise in some instances by spectral analysis. Despite this difficulty we can often learn about the dynamics of a process from the frequency properties of a time series. Below I introduce examples of the spectral distribution of a variety of time series.

Investigation of Selected Time Series

The empirical investigation of the spectrum of a process involves the examination of a graph of the spectral density function over the range $(0, \pi)$. One first considers the theory underlying the process, and then examines the time series itself. If one suspects that a periodic component is present, then the series is analyzed. Suppose we have a random process represented by a time series governed by a random process with mean 0 confined to the range $(-3.25, +3.25)$. Such a process can be generated with any random number generator. For a purely stochastic process, we should suspect that no frequency will contribute more to the variance of the spectral distribution function than any other. In other words, the spectrum will be essentially flat over its entire range, though some peaks may be present. Such a spectrum is shown in figure 2.3.

It is common to rescale the vertical axis by taking the logarithm of the spectral density estimates and to scale the x-axis by converting the frequency to cycles per unit time. The latter is accomplished by multiplying the frequency by $d/2\pi$, where d is the number of observations per unit time ($d = 1$ for discrete series). This follows from the fact that variation at frequencies higher than π cannot be distinguished from a corresponding frequency in the interval $(0, \pi)$.

Examination of figure 2.3 shows that the random time series ($N = 1{,}000$) is featureless. Put another way, we could say that this spectrum has the same power at all cycles; it could be easily described by a line of slope zero. Having the same power at all frequencies means that each component of the spectrum contributes about the same amount to the total variance of the function. This spectrum is the least interesting case but it provides a backdrop against which to check other, more interesting time series.

Let us reconsider equation 1. The function $Y_t = R \cos(\omega t + \theta) + E_t$, with $R = 2$, $\omega = (\pi/3)$, $\theta = (\pi/6)$, and $E_t = 0$ completes one wavelength for each six iterations. This is confirmed by inspection of figure 2.1. Recall that the Nyquist frequency is a method of scaling the interval over which the spectrum is to be estimated. Since the function given by equation 1 completes one cycle each six iterations, we should anticipate a peak at 1/6 of the Nyquist frequency. In translating the frequency into cycles per unit time, we normalize the ordinate to unity, so the expectation is that the peak is present at .16667.

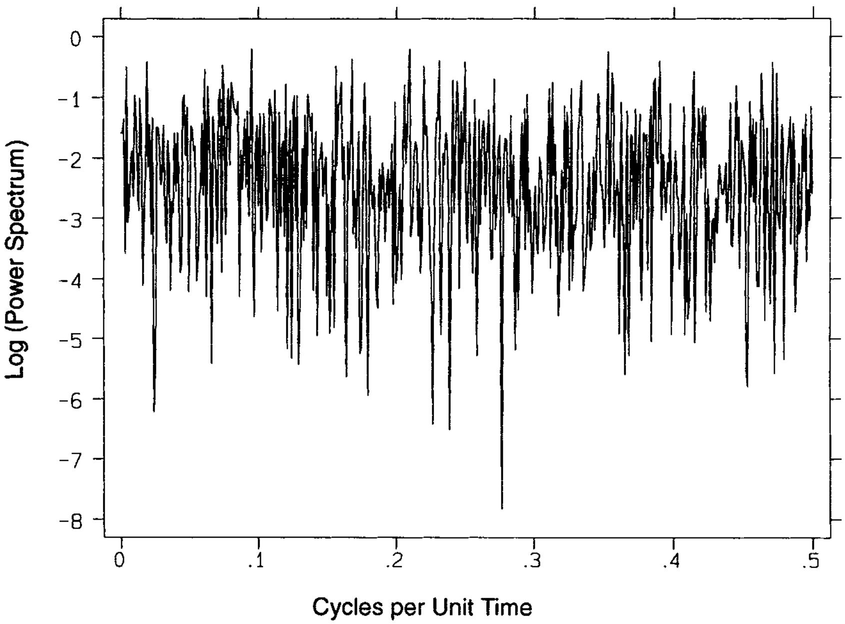

Fig. 2.3. Featureless spectrum for a random process. (*N* = 1,000.)

Figure 2.4 shows the spectrum for equation 1, and the spike is present precisely where it is expected. This spectrum shows that the variance in equation 1 is accounted for by a single frequency that corresponds to the precise wavelength of the time series. This example demonstrates convincingly the resolving power of spectral analysis with periodic data.[2]

These two examples use a purely random process and a known periodic process to illustrate what to look for in the spectrum of a time series. In the next section, I introduce three specific time series and their respective spectra. First I show how the spectrum responds to the presence of noise in the time series. Second, because our interest lies in the examination and characterization of (potentially) chaotic time series I demonstrate the analysis of a selection of time series with known chaotic properties both with and without noise present. Third, I analyze the spectrum of a time series derived from survey data that may have some chaotic properties.

Spectra of Selected Time Series

In order to illustrate the spectra of chaotic series, it is useful first to demonstrate what happens when noise is present in a periodic time series. Social

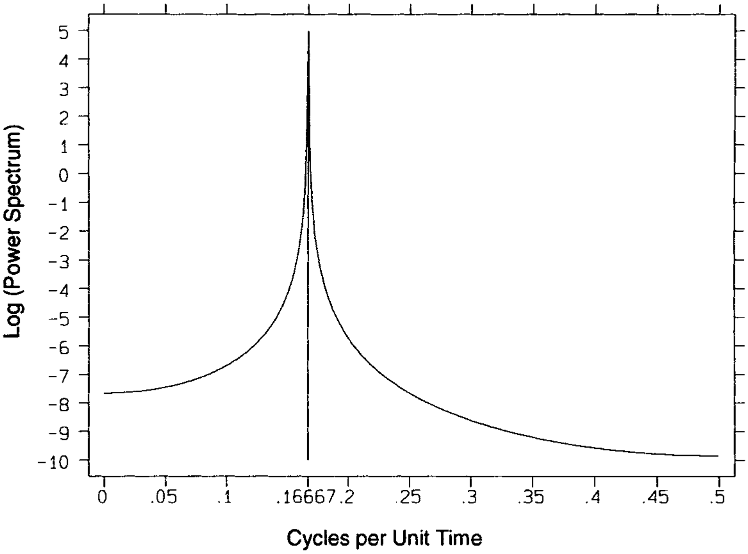

Fig. 2.4. Spectrum for a periodic function

scientists are well aware that noise is present in all measurements of social processes. The error in measurements can be due to random or systematic measurement error, the failure to account for a relevant variable, or unknown other means. In statistics, measurement errors can be problematic (Achen 1986; Greene 1978; Kenny and McBurnett 1992). In deterministic dynamics, the presence of error can have a catastrophic effect (Peitgen et al., 1992). As a matter of fact, all models of dynamical systems will contain some error. One source of error is the computer. For many models, the only source of error is that due to computer rounding error. In physical experiments, error can creep into the measurements from imprecision in the instruments used to take measurements. Social scientists are well aware that many measurements we make are imprecise. Inferences from surveys rely on sampling theory and the sample error increases as the sample size decreases. What we need to know is how stochastic error affects the resolving power of the spectral distribution function. I first examine two variants of a well-known equation to illustrate the resolving power of the spectral density function in a periodic function, the periodic function with an additive stochastic component, and a chaotic variant of the same equation. I then turn to an examination of two additional well-known chaotic time series.

L. Douglas Kiel and Euel Elliott, Editors
http://www.press.umich.edu/titleDetailDesc.do?id=14623
The University of Michigan Press, 1996

A Periodic Function

The logistic equation

$$Y_t = 3.5 \cdot Y_{t-1}(1 - Y_{t-1}) \qquad (7)$$

with $Y = 0.5$ produces a time series that contains a four-cycle. This series contains 500 elements. The four-cycle is a stable attractor. The series converges to the four-cycle by the third iteration. The points are .38281, .82694, .50088, and .87499. The series contains no noise except that due to computer rounding errors.

Figure 2.5 shows the spectrum for the logistic equation above. A four-cycle translates into .25 cycles per unit time, and the graph shows a very sharp peak at precisely that value. The remainder of the graph is flat. What happens when error is present?

The same time series produced by equation 7 is used again, but 20 percent error has been added to each observation. The error is normal, with variance equal to .07. This series also contains 500 data points. In this case,

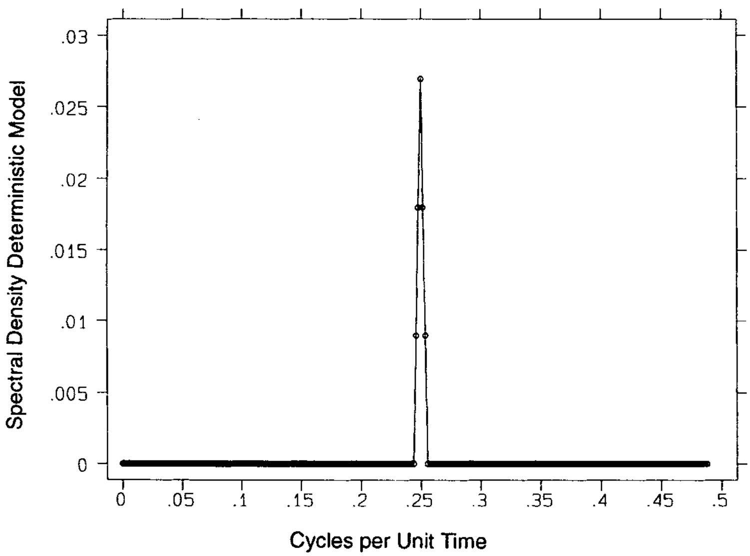

Fig. 2.5. Spectral density of the logistic equation with parameter = 3.5. (The time series contains a four-cycle in the unit interval.)

the spectrum is recovered precisely, with the sharp peak at .25 cycles per iteration, as before. The relatively large noise component manifests itself as background noise surrounding the sharp peak. Figure 2.6 shows that noise does not necessarily compromise the resolving power of spectral analysis. What does the spectrum of a series with multiple cycles look like?

If the driving parameter in equation 7 is changed to 3.6, the logistic equation produces a chaotic time signal. This means that the series contains 1, 2, 3, 4, . . . , ∞-cycles. The spectrum for this series should be quite different from the previous series, but the length of this series is not sufficient to allow cycles whose period exceeds 500 to be determined from the data. Figure 2.7 shows the spectrum for this series and it differs dramatically from the previous two. Here a sharp peak is seen again at .25 cycles, but this peak is straddled by two additional peaks, which are straddled by two additional peaks, and these too are bounded by another two peaks, after which the resolution ceases to allow a clear identification of the presence of additional peaks. This chaotic signal contains multiple cycles as well as broad background noise and the analysis of its frequency spectrum allows one to identify the presence of some of them. Notice that the spectrum shows that the time series appears to be

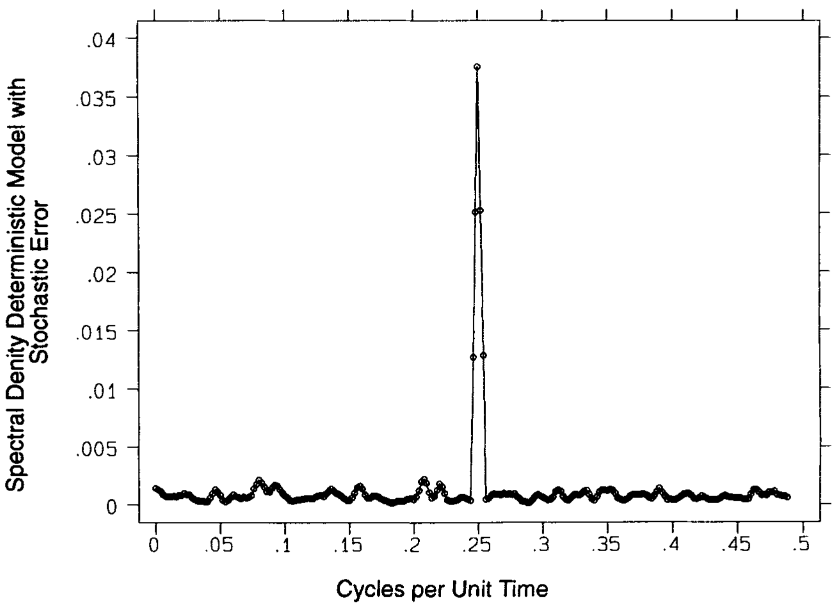

Fig. 2.6. Spectral density of the logistic equation with parameter = 3.5. (The time series contains a four-cycle in the unit interval.)

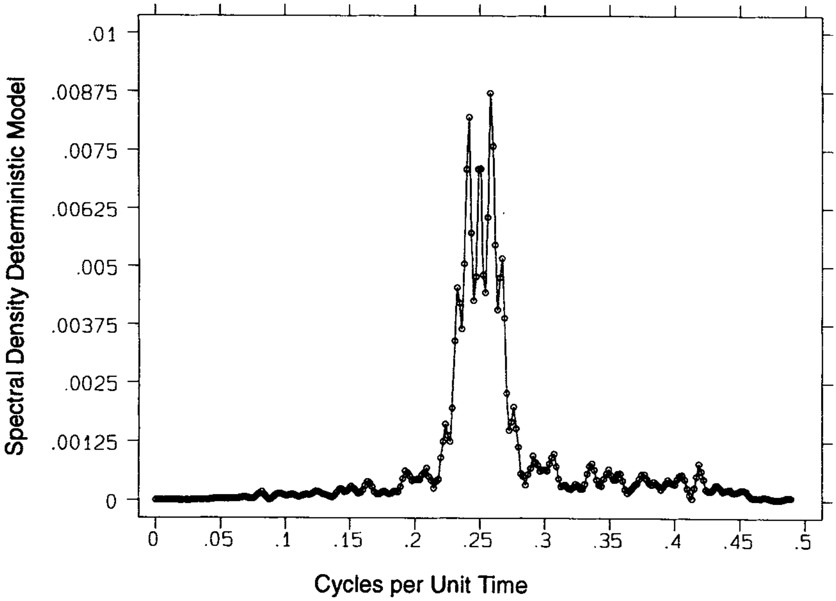

Fig. 2.7. Spectral density of the logistic equation with parameter = 3.6. (The time series contains multiple cycles.)

periodic, while it is known that this series is chaotic. We must recognize that spectral analysis does not allow the identification of any cycles that take longer than the length of the time series to complete.

Indeed, one prefers a time series of sufficient duration so that long cycles have ample time to complete several oscillations. With social science data we have to recognize that this requirement is almost never met; hence, we run a constant risk of failing to characterize correctly the periodicities present in the data. Note also that the use of spectral analysis to develop a model specification containing periodic components may also fail to characterize correctly the "true" model because the data fail to span the full range of oscillations. If one is prepared to assume that the length of the available time series is sufficient to explore the range of oscillation present in the true time series, then the examination of the spectral distribution function is appropriate.

The Lorenz and Rössler Equations and Their Spectra

The most well-known system of equations in chaos theory is the Lorenz equations. These equations comprise an early deterministic model of the weather formulated by Edward Lorenz (1963). The system of equations (in dot notation) is

$$\begin{aligned} \dot{x}_1 &= 10(-x_1 + x_2) \\ \dot{x}_2 &= 28x_1 - x_2 - x_1x_3 \\ \dot{x}_3 &= -(8/3)x_3 + x_1x_2. \end{aligned} \tag{8}$$

When integrated numerically they appear to possess extremely complicated dynamics (Hale and Koçak 1991).

The spectrum for the Lorenz equations has been studied extensively. Here I reproduce the work of Farmer et al. (1980). Figure 2.8 shows the spectrum for the Lorenz equations for the parameter values in equation 8 above. The equations were integrated numerically with the Runge-Kutta integrator in Phaser for fifteen iterations with a step size of 0.001. The result is a discrete time series of length 15,000. The variable $x_{1(t-T)}$ is used to investigate the spectrum.

This spectrum is essentially featureless, with high power at low frequen-

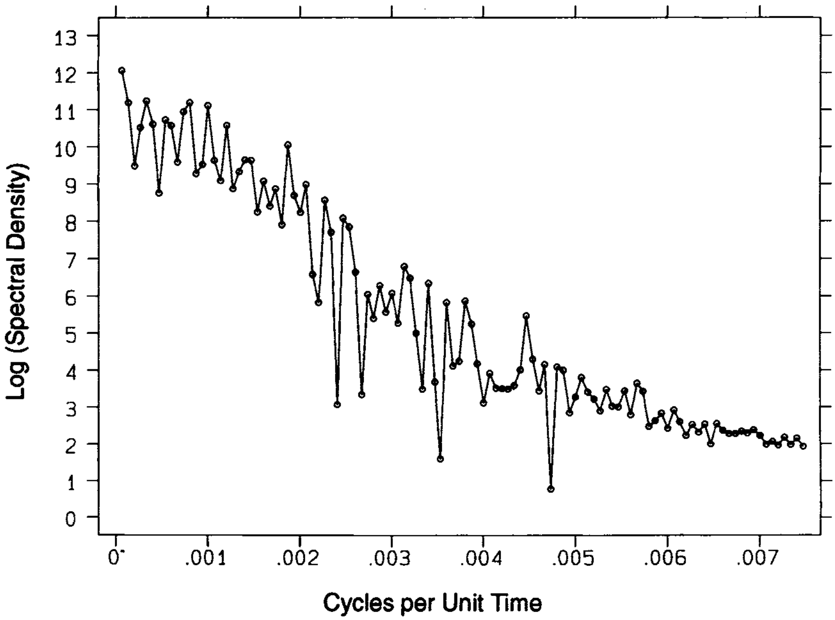

Fig. 2.8. Essentially featureless spectrum for the Lorenz equations (highest power at low frequencies with rapid decay)

cies, and it decays rapidly. The peaks that are present are not well defined and it appears that no periodicities are easily discernible. This result confirms results reported by Farmer et al. (1980) and is one characterization of chaotic time series.

The Rössler equations are another system of dynamic equations known to display chaotic dynamics. The Rössler equations describe the spread of disease and have been used effectively to model measles and whooping cough epidemics in children (Rössler 1976; Schaffer et al. 1986; Schaffer 1987; Schaffer et al. 1990). The system of equations is again written in dot notation.

$$\begin{aligned} \dot{x}_1 &= -(x_2 + x_3) \\ \dot{x}_2 &= x_1 + .2x_2 \\ \dot{x}_3 &= .2 + x_1x_3 - 5.7x_3 \end{aligned} \tag{9}$$

The parameters given in this system of equations describe what is called "the simple Rössler attractor" (Farmer et al. 1980). The equations were integrated numerically with the Runge-Kutta integrator in Phaser for 1,000 iterations with a step size of 0.1. The result is a discrete time series of length 10,000.

The spectrum for this system of equations (using x_{1t}), as shown in figure 2.9, displays features dramatically different from the Lorenz spectrum. The Rössler equations contain many identifiable peaks corresponding to cycles of particular length, with broad background noise evident between the peaks.

The logistic equation, the Lorenz equations, and the Rössler equations all display different spectra. Where periodicities are known to exist a priori, spectral analysis clearly identifies them as with the logistic equations. The Lorenz and Rössler systems are known to contain aperiodic cycles. This is evident in the spectrum for the Lorenz system, which displays high power at low frequencies with rapid decay in the spectrum. In contrast, the Rössler system has broad background noise and clearly identifiable peaks corresponding to periodicity at particular frequencies.

What we observe in these three examples, with and without noise, is the full range of possibilities in the spectral distribution functions of chaotic data. The spectrum for the logistic equation exhibits period doubling, which is a well-known route to chaos. The Lorenz equations, in contrast, show no identifiable structure other than high power at low frequency with a decay in power as the frequency increases. This pattern resembles the spectrum for many economic time series. (See Granger and Newbold 1986, 66 for a similar spectrum derived from the Federal Reserve Board index of industrial production.) Finally, the spectrum for the Rössler equations shows some identifiable peaks at a variety of frequencies that appear above broad background noise. Why do we observe so much variability in these spectra, each of which is

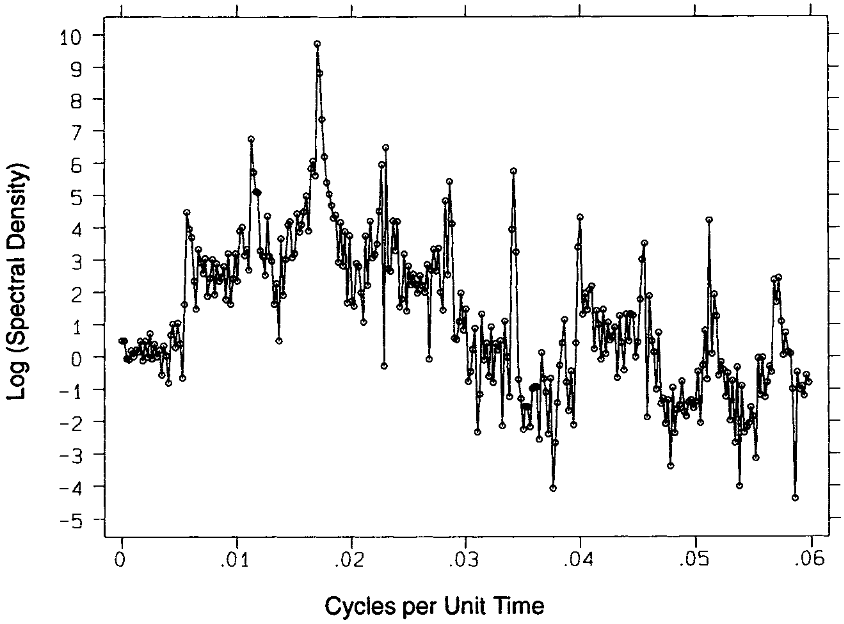

Fig. 2.9. Spectrum for the Rössler equations. (This spectrum has distinct features.)

known to be chaotic? The problem lies in the difficulty in distinguishing between signal and noise in chaotic series. As Glass and Mackey describe the problem:

> [b]road-band power spectra, perhaps with superimposed peaks, are often associated with chaotic dynamics. Unfortunately "noise" is also associated with broad-band spectra, and consequently the presence of a broadband spectrum is not adequate to establish chaos as opposed to noise (1988, 48).

The fact is that the spectrum provides us with only one piece of evidence for chaotic dynamics and even that may be suspect. The examples used here were taken from time series whose dynamics are known to be chaotic. If, for example, the driving parameter in the logistic equation were to be shifted downward by only 0.1 (to 3.5), the spectrum would show the single peak identified in figure 2.4, rather than multiple peaks surrounded by background noise.

Spectrum for the Public Opinion Series

In 1984, the Center for Political Studies conducted a telephone poll of the national electorate, using a probability sample over a very long period. This survey is known as the Rolling Cross-Section. Although it has some drawbacks, such as a fairly large sampling error (Allsop and Weisberg 1988), it is the only national sample of its type. This analysis uses that portion of the survey containing those respondents interviewed after January 11, 1984, and before June 12, 1984 (153 days). Those respondents were interviewed during the period in which all primary and caucus activity took place.

The survey asked respondents to rate the chances of the various Democratic candidates winning their party's nomination. As Bartels points out, most respondents failed to rate candidates' chances as if they had probability theory in mind (1988, app. A). Though Bartels used this survey for a quite different purpose, I make use of his coding scheme, which measures each individual's perception of each candidate's chance of winning the Democratic nomination (1988, 322–23).

The individual responses were averaged according to the date of the

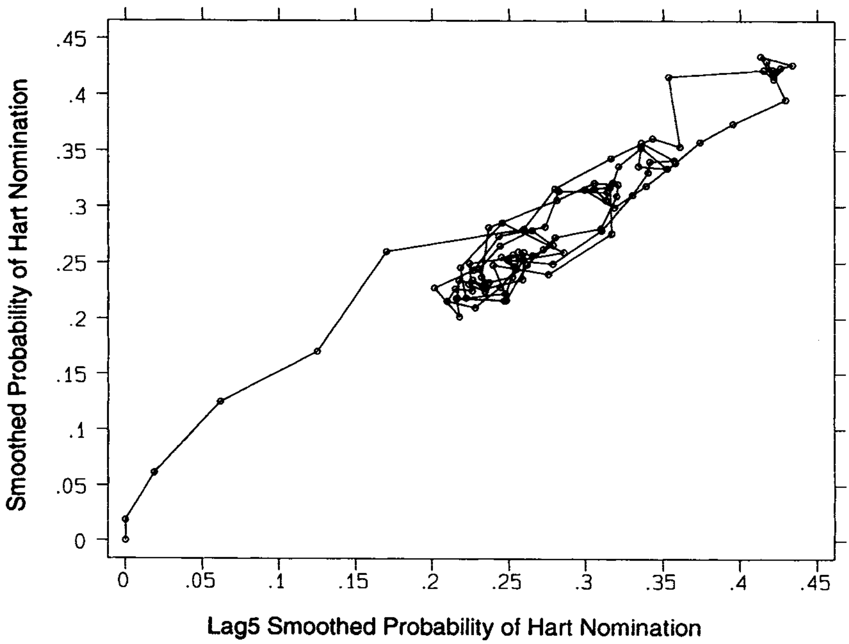

Fig. 2.10. Phase portrait of probability of Hart nomination time series

L. Douglas Kiel and Euel Elliott, Editors
http://www.press.umich.edu/titleDetailDesc.do?id=14623
The University of Michigan Press, 1996

responses. This resulted in a measure of the perceived mean daily probability of each candidate's chance of winning the Democratic nomination. In order to see if any interesting dynamical patterns were evident, I smoothed the data using a five-day moving average. Smoothing a time series that contains noise may allow a cyclical or otherwise interesting pattern to be identified visually, if one is present. Smoothing reduces, but does not eliminate, the level of noise in the time series.

Figure 2.10 shows the phase portrait for Gary Hart's chances of winning the 1984 Democratic nomination. For a review of how phase portraits are generated, see chapters 1 and 8 in this volume. The dynamical time path indicates that the probability Hart will win the nomination approaches the value .25. This matches Bartels's observation of Hart's chances of winning the nomination (1989, 251). Use of a moving average does not alter the gross structure of a time series, it merely serves to accentuate what is already present. The approach appears to be quite smooth, given the known level of noise in the data. The time path progresses upward from the origin, then appears to approach a limiting value. People's perceptions of Hart's chance of

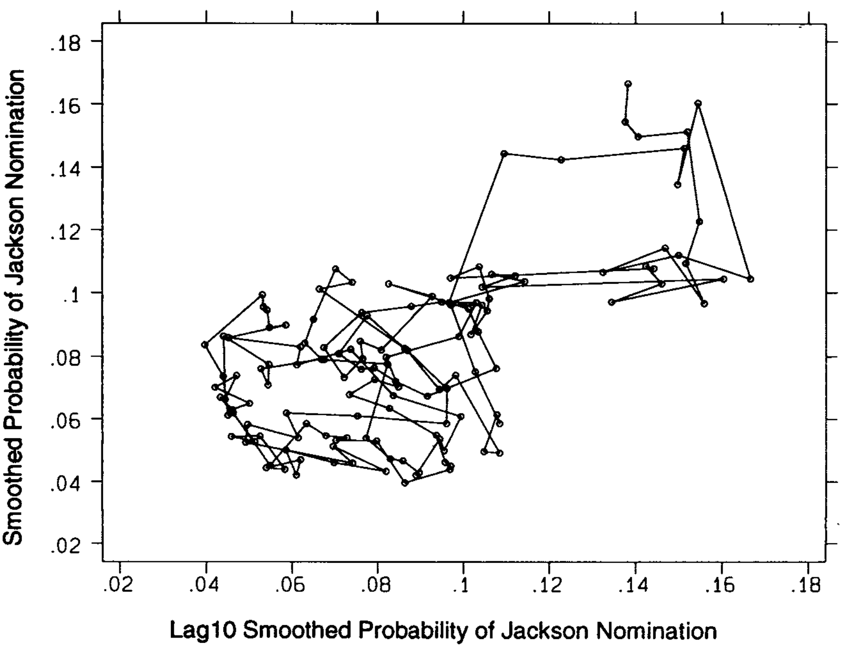

Fig. 2.11. Phase portrait of probability of Jackson nomination time series

winning the nomination appear to cycle about a particular value. Notice that Hart is always thought to have a chance of winning the nomination. He is not eliminated from the competition; he is considered to be a contender throughout the campaign. This pattern in public opinion matches the pattern in the electoral results closely. The expectation is that these data may contain a cycle with long periodicity, as do the electoral results.

Figure 2.11 shows the dynamics of the probability of Jesse Jackson winning the nomination. The dynamics here are anything but smooth and orderly. The graph contains a shape reminiscent of a figure eight. The dynamical behavior cycles in one direction and then reverses, and the two lobes revolve about separate and distinct values. Though the data are obviously noisy, the phase portrait shows that the dynamical pattern is bounded, that is, the dynamics remain near the attractor, and that this does not appear to be an ordinary cycle.

Figure 2.12 shows the phase portrait of the probability that Walter Mondale will win the 1984 Democratic nomination. Though it is difficult to see in the figure, the dynamics indicate that the perception of Mondale's chances of winning increases rapidly at first, cycling about one value (near 0.50), and

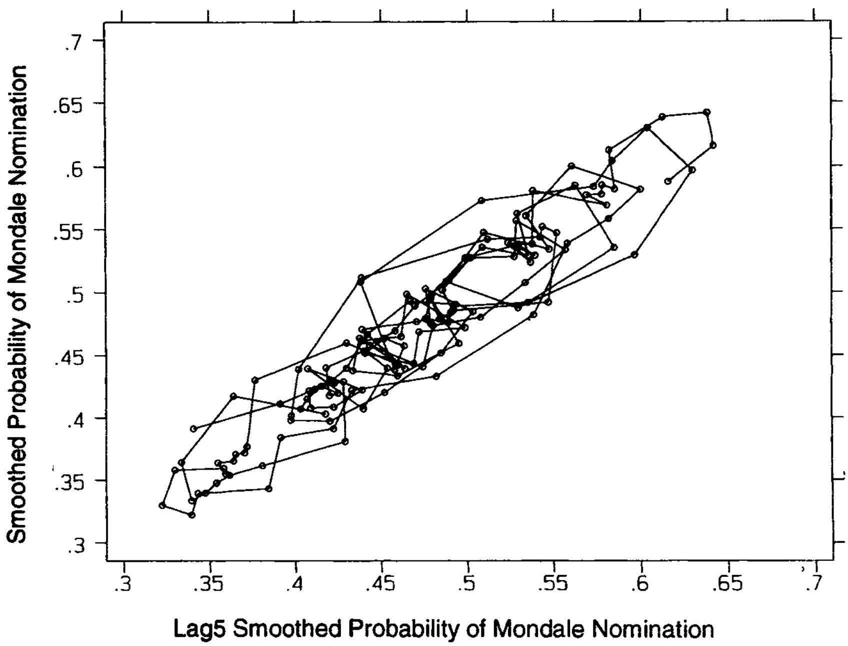

Fig. 2.12. Phase portrait of probability of Mondale nomination time series

then drops to a lower value (near 0.35), about which the series continues to cycle. Toward the end of the race the mean perception of his chances rises beyond the first level (near 0.60) and continues to cycle about that value. This graph shows extremely complex behavior but it is hidden by the noise present in the data.

Taken together, these figures reveal what may be a strange attractor resembling the Lorenz attractor from climate dynamics (Lorenz 1963; Schuster 1989). Strange attractors are associated with chaotic dynamical systems (Schuster 1989). The identification of this type of structure underlying the dynamics of public opinion is important because of the implications this has for forecasting the outcome of nomination races as well as the study of change in public opinion over time generally (Huckfeldt 1990). For example, if the time series here can be shown to be chaotic, then we will gain a deeper understanding of why polls are so variable while election outcomes are relatively easy to forecast (Gelman and King 1993).

The spectral distribution function for the Mondale time series has some features in common with the chaotic logistic spectrum. This spectrum is displayed in figure 2.13. It appears that all frequencies contribute about equally to the spectral density. The spectrum is dominated by broadband

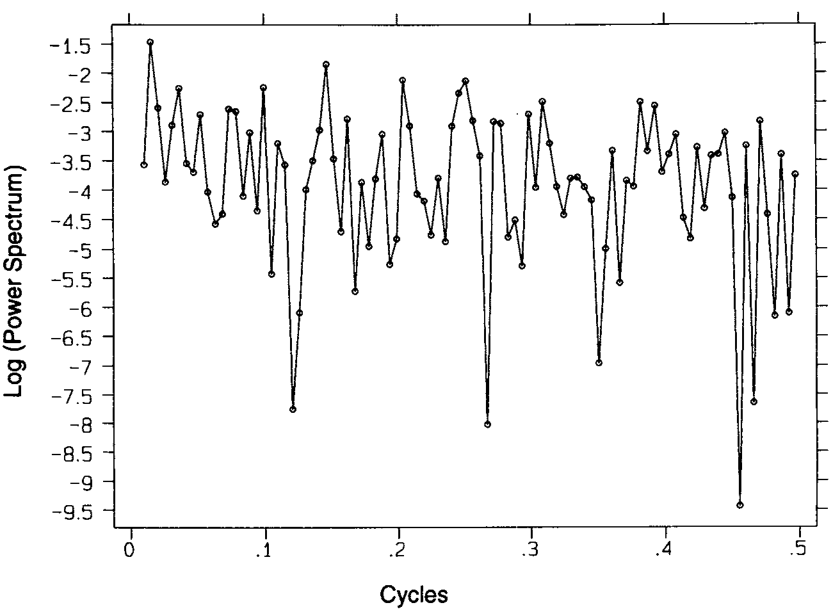

Fig. 2.13. Power spectrum for the Mondale time series (spectrum is essentially featureless)

noise. As Moon (1987) points out, spectral analysis may aid in identifying chaos in an experimental time series, but the technique is not conclusive if the system has "many hidden degrees of freedom of which the observer is unaware" (Moon 1987, 45).

The Lorenz, Rössler, and chaotic logistic equations have decidedly different spectra. The analysis of the Mondale series is based on the same principles, and the result reveals a spectrum that favors the chaotic logistic spectrum. Evidently there is no fundamental frequency for the Mondale series, which would appear as a sharp peak clearly extending above the background. Admittedly, the level of noise in this time series is high and that is probably one source of the problem in identification of cycles, if cycles are present. Some chaotic time series contain cycles, but in many cases (Lorenz equations, Rössler equations) these cycles are aperiodic and this fact manifests itself as a broadening of the spectrum. Spectral analysis of the Mondale series does not eliminate a deterministic component in the dynamics.

Conclusions

Spectral analysis serves to illuminate hidden periodicities in data, when they are present. Many chaotic series such as the Lorenz system are aperiodic while some, such as the logistic map, contain many, indeed an infinite number of, cycles. The presence of clear periodic behavior in a time series is easily discerned. This was demonstrated with the Rössler equations. The Lorenz equations contain aperiodic cycles and the spectrum fails to sift these out of the series, despite its great length. The logistic map, with driving parameter at 3.6, contains infinitely many cycles, of all lengths (Li and Yorke 1975). Spectral analysis cannot discern any cycle whose length exceeds the length of the data set (500 elements in the logistic equation). This may be the problem for the time series constructed from the CPS survey data. If the series has a cycle (or many) whose length exceeds the length of the observed time series, we cannot discover its presence. It seems intuitively reasonable to assume that the aggregated survey data contain no cycles longer than the entire nomination period. However, chaotic time series are known to have aperiodic cycles of many lengths, so it seems reasonable to assume that if they are present in the candidate time series they should have been observed, though the large stochastic component here is problematic.

This analysis has highlighted one problem present in many frequency analyses of chaotic time series: the results are ambiguous. The ambiguity arises on one hand from the difficulty in distinguishing between the stochastic component and the periodic signal when the signal-to-noise ratio is high, while on the other hand chaotic signals closely resemble pure noise. In some cases spectral analysis can discern between the two components and in others it cannot. While it seems that the Mondale (and the Jackson) series contains identifiable peaks corresponding to several frequencies, the series also contains considerable noise. Noise introduces problems in the resolution of periodicity, especially when the signal-to-noise ratio is high. Glass and Mackey (1988) and Moon (1989) clearly state that spectral analysis is only one tool to be used in the analysis of time series that are thought to be chaotic. Here we have an example of why no particular technique is to be relied upon exclusively in the analysis of suspected chaotic time series. A set of tests need to be applied to the time series. Spectral analysis is a good place to start, but measures of the dimension of the series (the correlation integral) as well as of the rate of divergence of nearby points on the attractor (the Lyapunov exponent) are required for empirical identification of chaotic motion.

NOTES

I would like to thank Thad Brown and Jim Kuklinski for their cogent critiques of earlier versions of this paper, and Gottfried Mayer-Kress for suggestions regarding the spectra of a variety of time series.

1. All time series used here are discrete because they are produced by numerical integration or recursion. Numerical integrators rely on small step sizes, which are a discrete approximation to an integral solution.

2. A variety of smoothing (or weighting) techniques are commonly used with spectral analysis to accentuate periodicity (see Chatfield 1991; Wei 1990 for examples). None of these techniques are used here because smoothing a chaotic time series destroys the basic property of interest! None of the basic research in the natural sciences uses smoothing in spectral analysis because the interest lies not in modeling the process but rather in examining the extant time series for periodicity in its natural state. Hence smoothing is not used here, nor should it be used in general.