INTERPRETING SPECTRAL ANALYSES IN TERMS OF TIME-DOMAIN MODELS

BY ROBERT F. ENGLE*

This paper derives relationships between frequency-domain and standard time-domain distributed-lag and autoregressive moving-average models. These relations are well known in the literature but are presented here in a pedogogic form in order to facilitate interpretation of spectral and cross-spectral analyses. In addition, the paper employs the conventions and discusses the estimation procedures used in the NBER's TROLL system.

1. INTRODUCTION

Although spectral analysis is a widely used tool in the statistical analysis of time series, the mathematical difficulties and the unfamiliarity of the concepts make it inaccessible to many economists. Increasing familiarity with time series models such as distributed lags and autoregressive-moving average processes (so called "Box–Jenkins" models), makes it now easy to describe and interpret spectral analysis without the use of complicated mathematics. Traditional time domain and more difficult frequency domain (spectral) analysis are just two ways of looking at the same phenomenon, but each has some advantages and the use of both can be an important aid in model building. Because frequency domain methods are more non-parametric, they are particularly useful in model specification.

The computational difficulties of performing spectral analysis have been substantially reduced through an innovative computer algorithm called the fast Fourier transform and the availability of the NBER TROLL computer system throughout the U.S. and many places abroad. The system is used by telephone and easily performs both time domain and frequency domain analysis. All the facilities to be described in this paper are available in TROLL.

The paper is intended to be a tutorial. It will develop in Sections 2 and 3 the correspondences between time and frequency domain analyses assuming only that the reader is familiar with time domain analysis. A short description of frequency domain estimation in Section 4 focuses upon confidence intervals and some adjustments which are available in TROLL. Finally, in Section 5, spectral analysis is used to provide a guide to the specification of time domain models with an example from economics.

There are many excellent reference works on spectral analysis which should be consulted for more details. Granger [6]† is perhaps the easiest to read, while Jenkins and Watts [10] is the most comprehensive. Fishman [7] focuses on some economic estimation problems and Dhrymes [3, 4] extends this direction with somewhat more mathematics. Hannan [8] gives a very rigorous treatment of the

† Parenthesized numerals refer to entries in the Reference section.

whole area. Relatively short and simple early expositions of the theory and practice are in Jenkins [9] and Parzen [13] with a good application to economics in Nerlove [12]. The book which is recommended as a companion to the TROLL system is Cooley, Lewis, and Welch [1] which is more application oriented and which describes, in Chapters 5 and 7, the basic concepts used in designing the system. Also see Cooley, Lewis, and Welch [2].

2. THE SPECTRUM

Many data series can be considered successive chance observations over time called stochastic processes. Possibly, each observation is independent of the preceding ones. However, for most applications, there is some suspected dependence between the observations. Both spectral analysis (frequency domain) and the more familiar time domain analysis are ways to characterize this dependence. High correlations between neighboring observations or seasonal components might be important forms of this dependence. Once the stochastic process is characterized, it may be possible to forecast its values, improve the efficiency of a regression where this is the disturbance, or make an inference about the economic model which produced such a variable.

Both frequency domain and time domain analyses begin with stochastic processes which are covariance stationary. This means that the covariance between an observation now and one a few periods later depends only on the time interval, not the dates themselves. Mathematically this can be expressed as

$$(1) \qquad \gamma(s) = E(x_{t+s} - \mu)(x_t - \mu)$$

where γ is the autocovariance function and μ is the mean. The important assumption is that neither depend upon t. While this assumption may seem strong, it is because of this condition that information from the past can be used to describe the present or future behavior.

Many economic time series appear to violate this assumption, particularly those with pronounced trends. It is generally possible, however, to create an approximately stationary series by taking first differences, or extracting a trend, thus leaving the series with a constant mean of zero. There may also be trends in variance which can often be removed by first taking logs of the series.

In the time domain the most common models are the autoregressive moving average models (ARMA). These may be purely autoregressive, purely moving average, or mixed.

A p-th order autoregressive and a q-th order moving average model are defined in equations (2) and (3), respectively, while (4) is an ARMA (p, q).

$$(2) \qquad x_t = a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p} + \varepsilon_t$$

$$(2') \qquad A(L)x_t = \varepsilon$$

$$(3) \qquad x_t = b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \ldots + b_q \varepsilon_{t-q} + \varepsilon_t$$

$$(3') \qquad x = B(L)\varepsilon$$

$$(4) \qquad x_t = a_1 x_{t-1} + \ldots + a_p x_{t-p} + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q} + \varepsilon_t;$$

$$(4') \qquad A(L)x = B(L)\varepsilon$$

In these equations ε is a series of independent, identically distributed random errors with ε_t independent of x_{t-i} for all i greater than zero; L is the lag operator and $A(L)$ and $B(L)$ are polynomials. These classifications are not unique since one type of process can, in general, be transformed into another. Nevertheless, they provide useful, simple models of time series which can be tested with data or used for analysis. Knowledge of $A(L)$ in (2'), $B(L)$ in (3') or $A(L)$ and $B(L)$ in (4') is equivalent to knowing the dynamics of the stochastic process of x.

The spectrum provides another way of characterizing time series. In this case we think of a series as being made up of a great number of sine and cosine waves of different frequencies which have just the right (random) amplitudes to make up the original series. Thus the list of how much of each frequency component was necessary is also a full description of the time series. The spectrum is a plot of the squared amplitude of each component against the frequency of that component. It is continuous and always greater than zero as long as there are no deterministic elements (that is, no exactly repeating components, or components which can be predicted exactly on the basis of the past). This is a very general way to describe a stochastic process.

The spectral density function is defined as the Fourier transform of the autocovariance function

$$(5) \quad f(\theta) = \gamma(0) + 2 \sum_{s=1}^{\infty} \gamma(s) \cos(2\pi\theta s) = \sum_{s=-\infty}^{\infty} \gamma(s) e^{-2\pi i \theta s} \qquad 0 \le \theta \le 1$$

where $i = \sqrt{-1}$, $e^{i\theta} = \cos(\theta) + i \sin(\theta)$, and the last equality follows from $\gamma(s) = \gamma(-s)$. There are several important features of this definition. First, although written in complex notation, the spectrum is real valued since all the imaginary sine terms cancel exactly. Second, since the cosine is symmetric $f(\theta) = f(1-\theta)$, and only the frequencies from 0 to 1/2 are needed to describe the spectrum. Third, integrating equation (5) from 0 to 1, shows that the area under the spectrum is equal to $\gamma(0)$, the variance. The spectrum is a decomposition of the variance into the components contributed by each frequency. A strict proof of the probabilistic basis of this interpretation is provided by the spectral representation theorem. Fourth, since θ is measured in cycles per period, it appears that there are no components from less than one cycle every two periods (the Nyquist frequency). The reason for this becomes clear upon reflection. When observing monthly data, weekly fluctuations will be indistinguishable from longer oscillations which have the same value at the moment the observation is taken. The weekly component will therefore be counted with these lower frequencies.

To clarify the interpretation of a spectrum and help with the notion of frequency components, consider the spectrum in Figure 1 which has been estimated from quarterly data.

From the definition of the spectrum in equation (5), the highest frequency oscillation which can be distinguished is 0.5 cycle per period. At this frequency, it takes two quarters to complete a cycle so there are two cycles per year. There is a peak at 0.25 cycle which corresponds to a four-quarter, or annual cycle. This is most likely a seasonal component. Similarly, the peak at 0.5 also indicates a seasonal component since it has an even number of cycles per year. The peak at
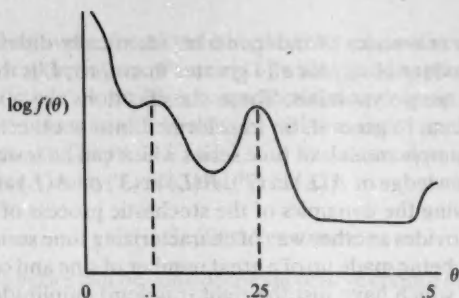
91

Figure 1 Typical spectrum

0.1 corresponds to a two and a half year oscillation. This might be a business cycle and economically interesting if it is significantly above its neighboring points. Generally, economic time series show behavior much like that of Figure 1.

The purpose of this paper is to emphasize the relationship between these concepts of frequency domain analysis and the more conventional time domain analysis. The first order serial correlation coefficient is easily calculated in the time domain and is generally large and positive for economic time series. This finding is easily observed in the frequency domain as well. Multiply the spectrum by $\cos(2\pi\theta)$ and integrate to obtain from equation (5) just $\gamma(1)$, the first order serial covariance. Roughly, this amounts to multiplying low frequencies by a positive number, high frequencies by a negative number, and adding. If the result is positive, there is positive first order serial correlation. Thus data series with generally downward sloping spectra have positive first order serial correlations while those with upward sloping spectra have negative serial correlations. Very important is the observation that spectra which are roughly symmetric about 0.25 will show no first order serial correlation.

A useful application of this analysis is found in interpretation of regression results. The assumption of no serial correlation in the disturbance is equivalent to the assumption that its spectrum is constant. The Durbin–Watson statistic gives a test against the possibility that there is first order serial correlation. This is a test against a general slope of the spectrum of the disturbance, whereas one would like to test against all forms of variation. In particular, notice that if the seasonality in Figure 1 were more severe, the spectrum might easily have no first order serial correlation but be far from constant. Durbin [5] formulated such a test based upon the spectrum of the residuals. In general, examination of the residual spectrum gives very useful information about the validity of the regression assumptions.

The link between time domain and frequency domain is completed by a derivation of the spectrum corresponding to the ARMA models of equations (2)–(4). The basic result is quite simple but will be established in the appendix.

Lemma 1: If x is a stochastic process generated by the model

$$A(L)x_t = B(L)\varepsilon_t$$

where ε_t is a series of independent identically distributed random variables with variance σ^2, and the polynomial $A(L)$ has all roots outside the circle, then the

92

spectrum of x is given by

(6)
$$f_x(\theta) = \sigma^2 |B(z)|^2 / |A(z)|^2, \qquad z = e^{(-2\pi i \theta)}$$

Notice that z is a complex function of θ.

Several examples should help to illustrate the usefulness of this result. First, notice that the spectrum of the very simple (white noise) process which has no time dependence, is just a constant. It has equal contributions from all frequencies.

Now consider the first order moving average process with parameter ρ, $x_t = \varepsilon_t + \rho \varepsilon_{t-1}$. From equation (6) the spectrum of x is

$$f_x(\theta) = |1 + \rho \varepsilon^{-2\pi i\theta}|^2 \sigma^2 = \{1 + \rho^2 + 2\rho \cos(2\pi\theta)\}\sigma^2$$

Evaluating this for θ in the range $(0, 1/2)$, gives a smooth spectrum which begins at $(1+\rho)^2$ and ends at $(1-\rho)^2$. If ρ is positive, this has the typical spectral shape which is common to most economic time series, and which implies a positive serial correlation coefficient, $\rho/(1+\rho^2)$. The first order autoregressive case is very similar but gives a somewhat steeper spectrum at low frequencies.

A very simple autoregressive model which captures the behavior of purely seasonal stochastic processes for monthly data is

$$x_t = \rho x_{t-12} + \varepsilon_t$$

From equation (6) the spectrum of this seasonal process is given by

$$f_x(\theta) = \sigma^2 / |1 - \rho \varepsilon^{-24\pi i\theta}|^2 = \sigma^2 / (1 + \rho^2 - 2\rho \cos(24\pi\theta))$$

which is plotted in Figure 2. There are peaks at all the harmonic frequencies: $\theta = 1/12, 2/12, 3/12, 4/12, 5/12, 6/12$, and all are equally important.
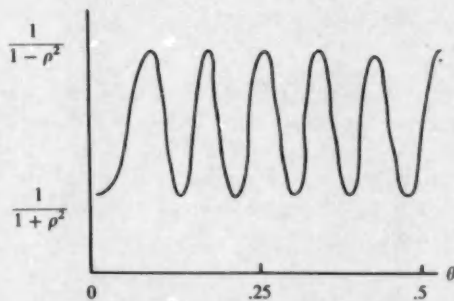

Figure 2   Spectrum of pure seasonal

3. THE CROSS SPECTRUM

The techniques used above can also be used to describe the relations between two jointly covariance-stationary time series. Both the individual behavior and the interrelations can be decomposed into basic sinusoidal elements.

The cross covariance function is a direct analogue of the autocovariance function. For two series with mean zero this is simply defined as:

(7)
$$\gamma_{xy}(s) = E(x_{t+s} y_t)$$

93

Again, notice that it does not depend on t. The cross spectrum is similarly defined as:

$$(8) \qquad f_{xy}(\theta) = \sum_{s=-\infty}^{\infty} \gamma_{xy}(s) \, e^{-2\pi i \theta s}$$

Because γ_{xy} is no longer symmetric the cross spectrum is not a real valued function of θ but rather a complex valued function.

Although the cross spectrum summarizes all the information in the series, it cannot be plotted directly. Instead, one examines statistics called "coherence squared," "gain," and "phase." These measures are defined here and given a rather extended interpretation below, connecting these concepts with the ideas of distributed lag regression models.

The coherence squared (COH) is like a correlation coefficient and is defined as:

$$(9) \qquad \text{COH}(\theta) = |f_{xy}(\theta)|^2 / f_x(\theta) f_y(\theta)$$

which is clearly between 0 and 1.

The gain (G) indicates how much the spectrum of x has been amplified to approximate that component of y.

$$(10) \qquad G_{xy}(\theta) = |f_{xy}(\theta)| / f_x(\theta)$$

This expression can clearly never be negative. However, if it is small, it indicates that at frequency θ, x has little effect on y.

The phase (PH) is a measure of the timing between the series. It is measured in the fraction of a cycle that y leads x.

$$(11) \qquad \text{PH}(\theta) = \frac{1}{2\pi} \arctan \left(\frac{-\text{Im} \, (f_{xy}(\theta))}{\text{Re} \, (f_{xy}(\theta))} \right)$$

where Im and Re are the imaginary and real parts of the cross spectrum.[*] There is a natural ambiguity about the phase since adding or subtracting 1 whole cycle from an angle will not change its tangent. The phase is known only up to adding or subtracting an integer and therefore even the lead-lag relation is not known for sure. The plot of the phase is designed to emphasize this fact. It is possible to combine the phase and the gain in a simple expression

$$(12) \qquad f_{xy}(\theta)/f_x(\theta) = G_{xy}(\theta) \, e^{-2\pi i \text{PH}(\theta)}$$

Two other potentially useful measures of the cross spectrum are its amplitude, which is merely its absolute value, and its time lag. The latter describes the phase in terms of the number of periods y leads x rather than the fraction of a cycle. Although this seems like a useful measure, the natural ambiguity of the phase also makes the time lag ambiguous and difficult to interpret. This may not be the case at low frequencies, where these difficulties are less likely to be important.

---

[*] The appropriate quadrant for PH is chosen on the basis of the signs of the real and imaginary components of the cross spectrum.

A natural and very general way for economists to think about the relations between two time series is in terms of a bivariate distributed lag model, such as

$$(13) \qquad y_t = \sum_{j=p}^{-p} w_j x_{t-j} + \varepsilon_t.$$

This is often rewritten in terms of the lag operator L as

$$(14) \qquad y_t = \sum_{j=p}^{-p} w_j L^j x_t + u_t = w(L)x_t + \varepsilon_t$$

where, for generality, leads as well as lags have been allowed and L^{-1} is interpreted as a lead operator. The same techniques required for equations (5) and (6), establish frequency domain interpretations of equation (14).

Lemma 2: If y is generated by a distributed lag model

$$y = w(L)x + \varepsilon$$

where x and ε are uncorrelated covariance stationary processes, then

$$(15) \qquad f_{xy}(\theta) = w(z^{-1})f_x(\theta),$$

and

$$(16) \qquad f_y(\theta) = |w(z)|^2 f_x(\theta) + f_\varepsilon(\theta)$$

where $z = e^{(-2\pi i\theta)}$.

Notice from (16) that the variance of y is broken into two parts: one which is the variation due to x modified by the lag distribution and the other due to the disturbance. Equation (15) shows that f_{xy}/f_x is an estimator of $w(z)$ which is just a function of the lag coefficients. Once $w(z)$ is known, all the lag coefficients can be found by merely taking the inverse Fourier transform. This is the basis of a very useful type of distributed lag estimation which is often called Hannan's inefficient method.*

Now consider running a regression of one component of y against the same frequency component of x. The regression coefficient would be the ratio of the covariance of x and y to the variance of x. In spectral notation this would be just $f_{xy}(\theta)/f_x(\theta)$. The R-squared of this regression is one minus the unexplained variance over the total variance. Substituting (15) into equation (16) demonstrates that the coherence squared is just the R-squared of this regression.

Similarly, from (12), the regression coefficient is the gain times $e^{-2\pi i PH(\theta)}$. The regression coefficient is just the gain if there is no time lag between the independent and dependent variables. If there is a time lag, the gain can be interpreted as the regression coefficient if the series were lagged just the right amount to eliminate any phase shift, and the phase is the angle by which they would have to be shifted (or the phase over the frequency is the time shift which is necessary). As a particular example, if the series are negatively related, the gain will still be positive but the phase will be 0.5.

---

* It is inefficient because it does not use the properties of the disturbance to construct an estimator with the smallest possible variance. It too is available in TROLL.

In summary, the coherence squared, the gain and the phase at a particular frequency can be interpreted in terms of a regression using only data at that frequency. The coherence squared is the R^2, the gain is the regression coefficient once any delay has been eliminated, and the phase is the angular shift needed to make this delay.

From Lemma 2 it is simple to plot the gain and phase corresponding to any particular time domain distributed lag model. A variety of these plots, often called Bode plots, are presented in Figure 3. In the balance of this section these figures will be analyzed for salient characteristics; and in Section 5 these are used to help specify a distributed lag model.
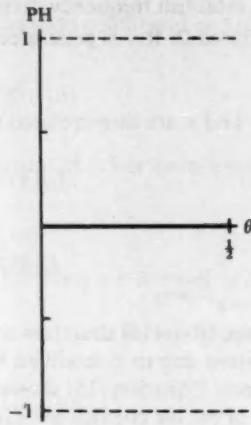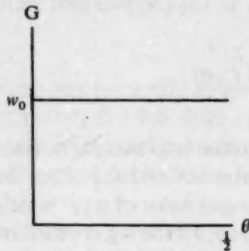
1. *Simple Static Model*

$$w(L) = w_0$$
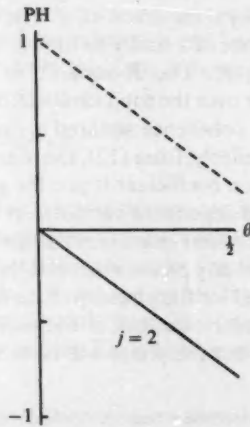
$$w(z^{-1}) = w_0$$

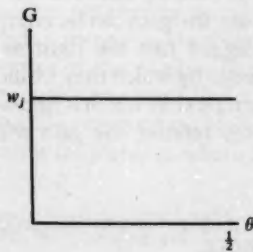$$G(\theta) = w_0$$

$$PH(\theta) = 0$$

2. *Simple Delay*

$$w(\theta) = w_j L^j$$

$$w(z^{-1}) = w_j e^{-2\pi i j \theta}$$
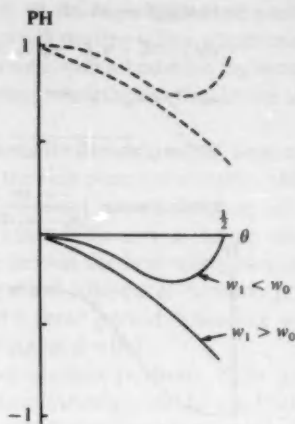
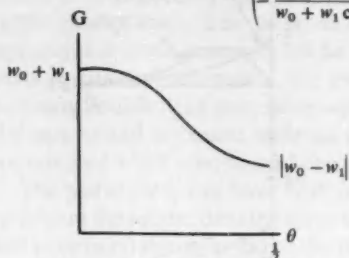$$G(\theta) = w_j$$

$$PH(\theta) = -j\theta$$

96

3. One Period Lag

$$w(L) = w_0 + w_1 L \quad w_0, w_1 > 0$$

$$w(z^{-1}) = w_0 + w_1 e^{2\pi i\theta}$$

$$G(\theta) = \sqrt{w_0^2 + w_1^2 + 2w_1 w_0 \cos 2\pi\theta}$$

$$PH(\theta) = \arctan\left(-\frac{w_1 \sin 2\pi\theta}{w_0 + w_1 \cos 2\pi\theta}\right)$$

4. Geometric Lag
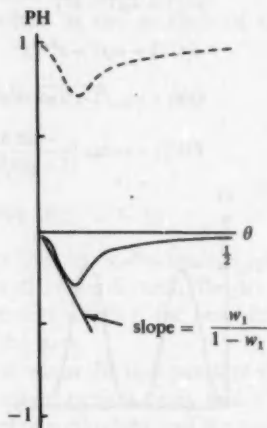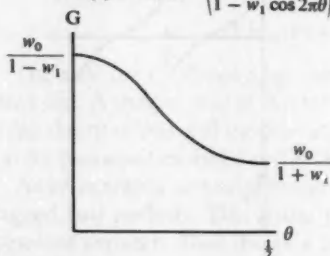
$$w(L) = \frac{w_0}{1 - w_1 L} = w_0(1 + w_1 L + w_1^2 L^2 + \ldots)$$

$$0 \le w_1 < 1$$

$$w(z^{-1}) = \frac{w_0}{1 - w_1 e^{2\pi i\theta}}$$

$$G(\theta) = w_0 \sqrt{1 + w_1^2 - 2w_1 \cos 2\pi\theta}$$

$$PH(\theta) = \arctan\left(\frac{-w_1 \sin 2\pi\theta}{1 - w_1 \cos 2\pi\theta}\right)$$

The simplest models are the static model described in (1) and the delay model shown in (2). Each has only one $w_j \neq 0$; thus

$$(17) \qquad\qquad w(z^{-1}) = w_j e^{2\pi i\theta j}.$$

The gain of these models is therefore constant and equal to w_j and the phase plot will be a straight line with slope $-j$. The (negative) slope of this line is just the delay of the dependent variable behind the independent variable. For any lag distribution, the mean lag is the slope of the phase at zero frequency (as long as the gain is non-zero). For the delay model, the slope is everywhere equal to the mean lag.
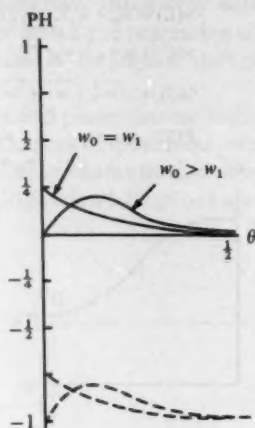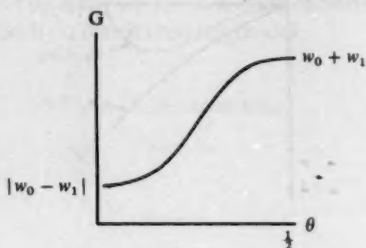
97

5. First Differences

$$w(L) = w_0 - w_1 L \quad w_0, w_1 > 0$$

$$w(z^{-1}) = w_0 - w_1 e^{2\pi i \theta}$$

$$G(\theta) = \sqrt{w_0^2 + w_1^2 - 2 w_0 w_1 \cos 2\pi\theta}$$

$$PH(\theta) = \arctan\left(\frac{-w_1 \sin 2\pi\theta}{w_0 - w_1 \cos 2\pi\theta}\right)$$

6. Four Period Differences

$$w(L) = w_0(1 - L^4)$$

$$w(z^{-1}) = w_0(1 - e^{8\pi i \theta})$$

$$G(\theta) = w_0 \sqrt{2 - 2\cos(8\pi\theta)}$$

$$PH(\theta) = \arctan\left(\frac{-\sin 8\pi\theta}{1 - \cos 8\pi\theta}\right)$$
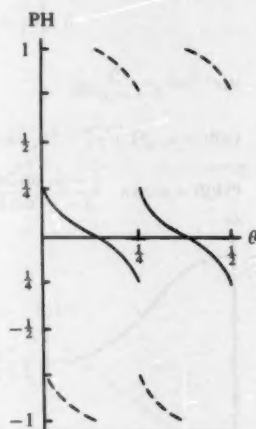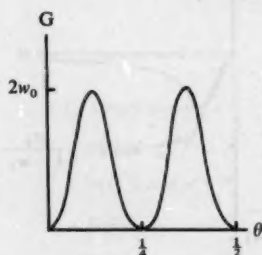
Models (3) and (4) include current and one lagged independent variable, and a lagged dependent variable, respectively. These very common types of lag distributions are used to model processes where the dependent variable only adjusts partially to a change in the independent variable in the current period. In model (3) the adjustment is completed in the second period, while in model (4) it continues forever but with geometrically declining effect.

In both of these cases, the gain is largest at low frequencies and then falls at higher frequencies. At zero frequency $z = 1$ and the gain is the sum of the lag coefficients or the long run propensity. Since z cannot exceed one, the peak of the

gain function will always be at zero frequency if all the lag weights are positive. The phase will be zero at zero frequency and then will decline with a slope equal to (minus) the mean lag. For more complicated models with all positive lag weights, the phase may change sign for higher frequencies; nevertheless, the initial slope will be the mean lag.

The final two models have a different character. In these cases, the dependent variable is influenced by the rate of change of the independent variable; the lag weights change sign. The peak in the gain function can now occur at any frequency, but it will generally not be at zero. In fact, if the sum of the lag weights is zero as with first differences, the gain will be zero at the beginning. Notice the interesting double peak associated with a four period difference. A twelve period difference would of course have six peaks and a three period difference would have one and a half (the second peak would occur at $\theta = 0.5$).

The phase plots can have both positive and negative portions. If the gain is non-zero at the origin, then the slope of the phase at the origin will be algebraically equal to (minus) the mean lag of the distribution. But, the mean lag is not a useful measure when the lag weights are of different signs. In particular, a negative mean lag does not imply a lead.

Both plots can be constructed for many other lag distributions either by hand or using the computer. A simple corollary of Lemma 2 will make it easy to combine these simple forms into more complicated lag distributions.

Corollary. If a lag distribution can be written as the product of two lag distributions such as

$$w(L) = u(L)v(L),$$

then

$$G_w(\theta) = G_u(\theta)G_v(\theta)$$

$$PH_w(\theta) = PH_u(\theta) + PH_v(\theta).$$

The rule for combining lag distributions is that the gains multiply and the phases add. A special case of this is familiar from the time domain: the product of two lag distributions will have a long run propensity which is the product of the separate propensities and a mean lag which is the sum.

As an example, consider the geometric form when the independent variable is lagged two periods. This is the product of a two period delay and a lagged dependent variable. Thus the gain is the product of a constant and the geometric gain, while the phase is the sum of the constant slope of the delay and the variation shown for the geometric.

4. SPECTRUM ESTIMATION

There are several distinct methods for estimating spectra and cross-spectra. The advantages and disadvantages of each have been extensively discussed. In particular, see Cooley, Lewis, and Welch [1] and Parzen [14]. Since the rediscovery of the fast Fourier transform, computational considerations suggest that periodogram averaging may be the most efficient method for spectrum estimation. In addition, it is conceptually simplest and leads to great versatility in the

estimation procedures. Finally, the usefulness of the periodogram in regression and various test procedures makes it sensible to compute this as a first step. See also Jones [11] and Tick [15].

The periodogram is defined as the square of the absolute value of the Fourier transform of the data series at each frequency, all divided by m, the number of observations. The formula for the periodogram is

$$(18) \qquad I_x(\theta_j) = \frac{1}{m} \left| \sum_{t=0}^{m-1} x_t e^{-2\pi i t \theta_j} \right|^2$$

where $\theta_j = j/m$ and $j = 0, 1, 2, \ldots, m$. This quantity is an estimator of the spectrum, but it is not a very good one. The expected value of the periodogram is

$$(19) \qquad E(I_x(\theta_j)) = \sum_{v=-m+1}^{m-1} \frac{(m-|v|)}{m} \gamma(v) e^{-2\pi i v \theta_j}.$$

For large values of m this estimator is an unbiased estimator of the spectrum, since $\gamma(v)$ is small for large v. Unfortunately, it is not a consistent estimator since the variance does not decrease as the sample approaches infinity. In fact, the periodogram at each frequency is approximately proportional to a chi squared random variable with two degrees of freedom, regardless of the number of observations. An intuitive explanation for this unusual circumstance is that as the sample becomes larger, more and more frequency points are estimated rather than obtaining better estimates of a fixed number of parameters. This explanation also suggests the solution. The average of a few neighboring points should give a better estimate of the spectrum in that neighborhood. Thus smoothing procedures must be used to obtain consistent spectrum estimators.

Two averaging or smoothing procedures, called "windows", are commonly used with periodogram averaging. A rectangular moving average gives the minimum variance for smoothing over a flat spectrum using only a certain number of points. However, when there are peaks in the spectrum, the rectangular window will lead to considerable bias and broadening of the peaks. An alternative window is a triangular window which gives the spectrum a much smoother appearance and is often better at describing the shape of peaks.

The width of the window is an important parameter in the estimation. The wider the window, the smaller is the variance of the resulting estimate; yet, the wider the window, the more serious may be the bias of smoothing over non-smooth portions of the spectrum. Two measures of width can be used to describe the windows, the bandwidth and the range. The bandwidth is the half-power width of the window. It is measured in frequency units, i.e., it is a fractional number of cycles per period. If, for example, the bandwidth is specified as 0.1, there will be five separate "bands" since the frequencies range from 0 to 0.5. For many purposes, spectral estimates separated by more than one bandwidth are considered to be independent.

The second measure is the range. This is merely the number of spectral points used in each moving average; it gives the separation between which two points are known to be completely independent. If the effective sample is 200 observations (implying 100 points in the spectrum) and the range is 20, there will be five

100

separate window widths in the estimation. A sensible value for the range is $\sqrt{m}$, where m is the number of observations.

Near the endpoints of the spectrum, the smoothing procedures must be modified. One choice is to decrease the range so that the window does not overlap the endpoints. Because the variance increases as the window becomes narrower, the variance increases markedly at very low or very high frequencies and one must be very cautious in interpreting low frequency peaks or troughs.

An alternative endpoint procedure is to keep the range constant but take advantage of the symmetry properties of the spectrum. The smoothing windows will "wrap" around the endpoints of the spectrum since the spectrum at $+\theta$ is just the same as at $-\theta$. This method will give smaller variances but bigger biases than the first method.

The spectral estimator resulting from smoothing the periodogram is approximately proportional to another chi squared random variable, this time with more degrees of freedom. The equivalent degrees of freedom are equal to

(20) $$\text{E.D.F.} = B \cdot m$$

where B is the bandwidth. This allows computation of a confidence interval for the spectrum. On the spectral plot, a 95 percent confidence interval can be constructed for each frequency separately.

Estimates of the cross spectrum are accomplished in exactly the same manner. The finite Fourier transform of one series is multiplied by the complex conjugate of the Fourier transform of the other to form the cross periodogram. The real and complex parts of this are then smoothed individually, just as for the periodogram. The sampling distributions for the various measures derived from the cross spectrum also depend only on the equivalent degrees of freedom of the estimate. With the coherence plot, the critical point for a 5 percent test of the hypothesis of zero coherence can be calculated. Approximate 50 percent confidence intervals for the gain can be plotted with the output. These depend on the sample coherence; where the coherence is small, the confidence interval is large.

When using a wide window, peaks tend to be spread out. For many series we know a priori where these peaks will be, either because the series is typical in having strong low frequencies or because it has important seasonality. In these cases "prewhitening" is often recommended. This amounts to dividing the raw periodogram by the expected or typical shape, smoothing this "prewhitened" periodogram which no longer has the large peaks, and then "recoloring" by multiplying by the typical spectral shape. A seasonal and non-seasonal version of the prewhitening filter might be important. Prewhitening can be done in connection with either spectrum or cross spectrum estimation.

A second characteristic which is likely to make the smoothing procedure badly biased in cross spectral estimation is misalignment of the series. When one series lags another by several periods, there is a peak in the cross covariance function which is not at zero. This leads to a regular oscillation in the amplitude of the cross periodogram. Smoothing this will obscure this particular bit of information as well as distorting other results. The recommended procedure is to first divide the cross periodogram by an aligning series, smooth the cross periodogram, and then remultiply it by the aligning series. To construct the alignment series one

first computes the inverse Fourier transform of the cross periodogram which is exactly the cross covariance function. This could have been computed from the data directly, but such a method is apparently inferior to the computation of Fourier and inverse Fourier transforms. Searching the cross covariance function for the maximum yields the information needed to construct the alignment series. If this procedure were applied to the estimate of the spectrum or a cross spectrum which was already aligned, the maximum covariance would be the zeroth estimate and thus the alignment series would be unity and would have no effect.

The Fourier transform algorithm used in these computations is the Cooley–Tukey fast Fourier trahsform. In its basic form it expects a series with 2^n elements and thus each series is padded out to this length with zeros (the mean). The number of spectral points estimated is therefore 2^{n-1} which are evenly spaced between the frequencies 0 and 1/2 cycle per basic time unit of the data. Frequently it is desirable to estimate the spectrum at particular points or not to pad with zeros. In this case, it is possible to pick an integer so that the series is padded to $N = 2^s r$. Choosing $r = 3$, for example, would insure factors of 12 which would be required in order to have exact seasonal points with monthly data. This variation can make substantial difference in the results when there is strong seasonality.

5. THE USE OF SPECTRA IN SPECIFYING MODELS

This paper has shown that the time domain and the frequency domain are just two different ways of looking at the same models. The estimation procedures are, however, dramatically different, and it is here that the two techniques can fruitfully be combined. In order to estimate an ARMA or a distributed lag model, using time domain methods, one must first specify (or "identify" as the statisticians say) the form of the model. The process of specification usually involves a series of trial forms and statistical tests coupled with a liberal amount of good judgment. The judgment is particularly important since the tests become of questionable validity when applied in sequence.

On the other hand, the spectral methods do not require the specification of the model. The estimation procedure is independent of the form of the model. This is clearly an advantage since the first difficult step can be avoided; however, it is also a disadvantage, since in general, a very large number of parameters must be estimated and will have relatively wide confidence intervals.

The situation is exactly comparable with the choice of parametric or non-parametric statistical methods. If the parameterization is correct, the parametric procedure is far more efficient than the non-parametric procedure. However, if the parameterization is incorrect, only the non-parametric method will give a valid result. The user of non-parametric statistics gives up some efficiency for insurance against a wrong parameterization.

This discussion suggests a two step procedure for the analysis of stochastic processes. Use the spectral methods to aid in the specification of time domain models, and then use standard time domain methods to estimate them, thereby avoiding the trial and error search for an appropriate model. This is an approach which has been used by engineers for many years but has not had wide acceptance

by economists. It will be illustrated by an economic model of housing expenditures in the U.S.

In Figure 4, quarterly housing expenditures and mortgage rates are plotted from 1950 through 1974. The two series are apparently non-stationary and growing exponentially. They appear to be negatively correlated as expected since higher interest rates generally discourage housing. With careful attention to the timing, it may be possible to see that declines in housing follow any interest rate increase with a lag. Questions of particular interest in this analysis might be (1) the timing of interest rate effects on housing and (2) whether the housing market overshoots so that the lag weights change sign for long lags. This is presumably the behavior of other types of investment as implied by the accelerator and neoclassical models of investment.

To calculate the cross spectrum of housing with interest rates, both series are expressed as the first difference of the logarithms, that is, the quarterly growth rates. The gain and the phase as well as the summary statistics from the one line TROLL command requesting this output are shown in Figure 5. The triangular smoothing window is used with the "wrap around" endpoint option and the bandwidth is set (by default) at 0.079. Notice that the rate is multiplied by minus one so that the phase will be centered at 0 rather than 0.5, since interest rates and housing starts are negatively related.

Figure 4   Housing data

Time bounds: 1946 1st to 1974 4th.   Data names: Rate Hous.

The task of choosing a time domain specification for a distributed lag model is now to compare the theoretical Bode plots in Figure 3 with the observed gain and phase in Figure 5. It is clear that none of the theoretical plots are close approximations of the estimated gain and phase, but this is to be expected. Some features are, however, recognizable. The gain begins quite small and then rises to a peak at a
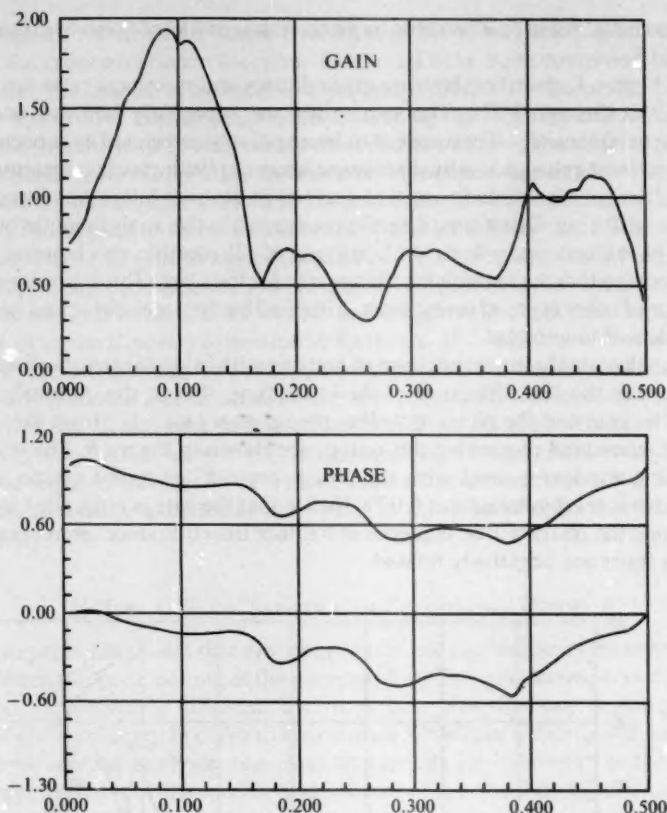
Summary statistics and options:
    Type of smoothing: Wrap-triangular. Range: 10.
    Type of prewhitening: None.
    FFT parameter: 1.
    Preprocessing: Demean.
    Number of observations in data series: 100
    Basis: 128.
    Bandwidth: 0.079.
    No alignment.
    Critical coherence squared (at 5 percent level): 0.636.
    $-$(Del(1 : Log(Rate))) variance: 0.001. Spectrum total: 0.001.
    Del(1 : Log(Hous)) variance: 0.003. Spectrum total: 0.003.

Figure 5   Gain and phase between housing and interest rates

period of about 10 quarters where the interest rate elasticity is almost 2. The second most important peak is at approximately 2 quarters per cycle. These look roughly like the four quarter first difference except that the second peak is at too high a frequency and is too small. The phase looks much like a one quarter delay for the first half of the spectrum but then deviates from this in the second half.

TABLE 1
REGRESSION COEFFICIENTS FOR HOUSING AND INTEREST RATES

	Spectral Specification	Alternative 1	Alternative 2	Alternative 3
R^2	.53681	.53566	.42562	.56004
$\bar{R}^2$	.52203	.51043	.40092	.52049
H_{-1}	.51 (6.74)	.48 (5.28)	.69 (6.71)	.51 (4.86)
H_{-2}	—	—	-.07 (.58)	.03 (.23)
H_{-3}	—	—	-.12 (1.18)	.16 (1.60)
R	.70 (5.71)	.17 (1.28)	.26 (1.94)	.08 (.60)
R_{-1}	—	.55 (3.75)	—	.62 (4.11)
R_{-2}	—	.22 (1.36)	—	.15 (.91)
R_{-3}	—	-.16 (1.10)	—	.05 (.32)
R_{-4}	-.24 (1.83)	—	—	.08 (.52)
Const.	.007 (1.71)	.0 (2.19)	.006 (1.25)	.01 (2.29)

* t-statistics in parentheses. All regressions have 98 observations.

To synthesize a close approximation to Figure 5, one must remember the rules for combining lag distributions. If two lag distributions are multiplied together, the gains are multiplied and the phases are added. Multiplying a geometric distribution times a four quarter first difference will increase the gain of the first peak relative to the second. If a three quarter difference is used, then the second peak will occur at 0.5 which is closer to the observed maximum, especially recalling that the confidence intervals at the endpoints are very large. Finally, the gain at low frequencies is not exactly zero, so probably the coefficients on the third differences should be allowed to differ. In summary, a model with one lagged dependent variable and a current and three period lagged independent variable appears capable of producing approximately the observed gain pattern.

If this model is used, the phase pattern would be the sum of a negative contribution, due to the lagged dependent variable, and a positive component, due to the three period lag. Altogether, this is unlikely to produce the appropriate decline in the phase with frequency which is observed. An additional delay will not alter the gain since it has gain of unity, but will add a linearly declining phase to the others. This line of reasoning leads to consideration of the time domain model:

$$(21) \qquad (1 - w_0 L)H = (w_1 + w_2 L^3)(L)R$$

where H and R are the first differences of the logs of housing and interest rates.

This model can be estimated directly with ordinary least squares.* The result is given in Table 1, along with the gain and phase plots which correspond to this set of parameters. The pictures are rather close to the calculated spectra and the time domain regression statistics are generally good with significant t-statistics. The change of sign in the regression coefficients which was anticipated for the lagged independent variable does occur and, as expected, the lagged coefficient is not as large as the current one.

To examine whether this identification procedure led to a particularly good model, both simpler and more complicated versions were estimated. Two naive ways to generate lag distributions are to successively increase the number of lagged independent or lagged dependent variables. Three models are shown in Table 1 with their associated gain and phase plots. One has the current and three quarters lag on interest rates with one lagged dependent variable, a second has three lagged dependent variables but only the current independent variable, and the third is the most general model with three lagged dependent variables and the current and four quarters of lagged independent variables. Notice that the first and second of these alternative models fail to reproduce the second peak of the gain function. The third model does show roughly the right shape, but it does so by using 9 parameters rather than the 4 required by the spectrally specified model. Furthermore, F tests for the restrictions implied by the spectral model are invariably less than one (even testing against alternative 1 when R_{-4} is included) and thus the restrictions are easily accepted.

In conclusion, the specification suggested by spectral analysis is a good specification which might have eventually been reached by time domain techniques after much trial and error. Of many models tried, it has almost the highest

---

* Least squares will be consistent only if the dynamic specification has eliminated serial correlation in the disturbance. A test of the hypothesis of white noise was accepted.

corrected R^2 and F tests supported the restrictions it imposed. The economic result is that there appears to be a one quarter delay before interest rates influence housing, thereafter there is a geometrically declining effect. Eventually (one year later) there is a change in the direction of housing response as the market has presumably caught up to the desired capital stock. This is easily seen in the time domain but is also apparent in the frequency domain since the gain is small for low frequencies.

This example lends credibility to the proposition that the use of both time and frequency domain techniques may enrich each and, in particular, that frequency domain methods may be very helpful in model specification.

<center>APPENDIX</center>

Lemma A.1: If x is a stochastic process generated by the model

$$A(L)x_t = B(L)\varepsilon_t$$

where ε_t is a series of independent identically distributed random variables with variance σ^2, and $A(L)$ has all roots outside the unit circle, then the spectrum of x is given by

$$f_x(\theta) = \sigma^2 |B(z)|^2 / |A(z)|^2$$

where $z = \varepsilon^{(-2\pi i \theta)}$.

Proof: Consider the moving average process

$$x_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j}$$

where the ε are all independent. Then

$$\gamma(s) = Ex_{t+s}x_t = E \sum_{j=0}^{q} b_j \varepsilon_{t+s-j} \sum_{k=0}^{q} b_k \varepsilon_{t-k}$$

where the expectation on the right only has non-zero values where $-k = s - j$ and $0 \le k \le q$. Therefore for $q \ge s \ge 0$

$$\gamma(s) = \sigma^2 \sum_{j=s}^{q} b_j b_{j-s}$$

and is 0 otherwise. The spectrum of x is defined using equation (5) and the symmetry of γ by

$$f(\theta) = \sum_{s=-\infty}^{\infty} \gamma(s)z^s = \sum_{s=1}^{q} \gamma(s)(z^s + z^{-s}) + \gamma(0)$$

$$= \sigma^2 \sum_{s=1}^{q} \sum_{j=s}^{q} b_j b_{j-s}(z^s + z^{-s}) + \sigma^2 \sum_{j=0}^{q} b_j^2$$

which can be written

$$f(\theta) = \sigma^2 \sum_{j=0}^{q} b_j z^j \sum_{k=0}^{q} b_k z^{-k} = \sigma^2 \left| \sum_{j=0}^{q} b_j z^j \right|^2$$

There is nothing in this proof which requires that q be finite. Since every stable ARMA process has a (possibly infinite dimensional) moving average representation, the result is true for any ARMA process.

Lemma A.2: If y is generated by a distributed lag model

$$y = w(L)x + \varepsilon$$

where x and ε are uncorrelated covariance stationary processes, then

$$f_{xy}(\theta) = w(z^{-1})f_x(\theta)$$

and

$$f_y(\theta) = |w(z)|^2 f_x(\theta) + f_\varepsilon(\theta)$$

where $z = \varepsilon^{(-2\pi i\theta)}$.

Proof: Without loss of generality take both x and y to have mean zero

$$\gamma_{xy}(s) = Ex_{t+s}y_t$$

$$= E\sum_j w_j x_{t-j}x_{t+s} + Ex_{t+s}\varepsilon_t$$

$$\gamma_{xy}(s) = \sum_j w_j \gamma(s+j)$$

$$f_{xy}(\theta) = \sum_{s=-\infty}^{\infty} \gamma_{xy}(s)z^s = \sum_j \sum_s w_j \gamma(s+j)z^{s+j}z^{-j}$$

$$= \sum_j w_j z^{-j} f_x(\theta)$$

$$f_{xy}(\theta) = w(z^{-1})f_x(\theta).$$

And

$$\gamma_y(s) = Ey_{t+s}y_t$$

$$= E(\sum w_j x_{t+s-j} + \varepsilon_{t+s})(\sum w_k x_{t-k} + \varepsilon_t)$$

$$\gamma_y(s) = \sum_j \sum_k w_j w_k \gamma_x(s-j+k) + \gamma_\varepsilon(s)$$

$$f_y(\theta) = \sum_s \gamma_y(s)z^s$$

$$= \sum_s \sum_j \sum_k w_j w_k \gamma_x(s-j+k)z^{s-j+k}z^{-k}z^j + \sum_s \gamma_\varepsilon(s)z^s$$

$$= \sum w_j z^j \sum w_k z^{-k} f_x(\theta) + f_\varepsilon(\theta)$$

$$f_y(\theta) = |w(z)|^2 f_x(\theta) + f_\varepsilon(\theta).$$

108

REFERENCES

[1] Cooley, J. W., P. A. W. Lewis, and P. D. Welch, *The Fast Fourier Transform Algorithm and its Applications*, IBM Research Paper RC-1743, 1967.

[2] —— "The Application of the Fast Fourier Transform Algorithm to the Estimation of Spectra and Cross Spectra," *Journal of Sound Vibrations*, 1970.

[3] Dhrymes, P. J., *Econometrics: Statistical Foundations and Applications*. New York: Harper and Row, 1970.

[4] —— *Distributed Lags; Problems of Estimation and Formulation*. San Francisco: Holden Day, 1971.

[5] Durbin, J., "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals," *Biometrica*, 56 (1969), pp. 1–14.

[6] Granger, C. W. J., in association with M. Hatanaka, *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press, 1964.

[7] Fishman, G. S., *Spectral Methods in Econometrics*. Cambridge: Harvard University Press, 1969.

[8] Hannan, E. J., *Multiple Time Series*. New York: J. Wiley, 1970.

[9] Jenkins, G. M., "General Considerations in the Analysis of Spectra," *Technometrics*, 3, 2 (1961), pp. 133–166.

[10] Jenkins, G. M. and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco: Holden Day, 1969.

[11] Jones, R. H., "A Reappraisal of the Periodogram in Spectral Analysis," *Technometrics*, 7 (1965), pp. 531–542.

[12] Nerlove, M., "Spectral Analysis of Seasonal Adjustment Procedures," *Econometrica*, 32 (July 1964), pp. 241–286.

[13] Parzen, E., "Mathematical Considerations in the Estimation of Spectra," *Technometrics*, 3, 2 (1961), pp. 167–190.

[14] —— "Multiple Time Series Modeling," in *Multivariate Analysis*. P. Krishnaia editor. New York: Academic Press, 1969.

[15] Tick, L. H., Letter to the Editor, *Technometrics*, 8 (1966), pp. 559–561.

109