Chapter 17

TIME SERIES AND SPECTRAL METHODS IN ECONOMETRICS

C. W. J. GRANGER and MARK W. WATSON

Contents

1.	Introduction	980
2.	Methodology of time series analysis	980
3.	Theory of forecasting	993
4.	Multiple time series and econometric models	1002
5.	Differencing and integrated models	1006
6.	Seasonal adjustment	1009
7.	Applications	1016
8.	Conclusion	1019
References		1019

Handbook of Econometrics, Volume II, Edited by Z. Griliches and M.D. Intriligator

1. Introduction

A discrete time series is here defined as a vector x_t of observations made at regularly spaced time points $t=1,2,\ldots,n$. These series arise in many fields, including oceanography, meterology, medicine, geophysics, as well as in economics, finance and management. There have been many methods of analysis proposed for such data and the methods are usually applicable to series from any field. For many years economists and particularly econometricans behaved as though either they did not realize that much of their data was in the form of time series or they did not view this fact as being important. Thus, there existed two alternative strategies or approaches to the analysis of economic data (excluding cross-sectional data from this discussion), which can be called the time series and the classical econometric approaches. The time series approach was based on experience from many fields, but that of the econometrician was viewed as applicable only to economic data, which displayed a great deal of simultaneous or contemporaneous interrelationships. Some influences from the time series domain penetrated that of the classical econometrician, such as how to deal with trends and seasonal components, Durbin–Watson statistics and first-order serial correlation, but there was little influence in the other direction. In the last ten years, this state of affairs has changed dramatically, with time series ideas becoming more mainstream and the procedures developed by econometricians being considered more carefully by the time series analysts. The building of large-scale models, worries about efficient estimation, the growing popularity of rational expectations theory and the consequent interest in optimum forecasts and the discussion of causality testing have greatly helped in bringing the two approaches together, with obvious benefits to both sides.

In Section 2 the methodology of time series is discussed and Section 3 focuses on the theory of forecasting. Section 4 emphasizes the links between the classical econometric and time series approaches while Section 5 briefly discusses the question of differencing of data, as an illustration of the alternative approaches taken in the past. Section 6 considers seasonal adjustment of data and Section 7 discusses some applications of time series methods to economic data.

2. Methodology of time series analysis

A discrete time series consists of a sequence of observations x_t taken at equi-spaced time intervals, examples being annual automobile production, monthly unemployment, weekly readings on the prime interest rate and daily (closing) stock market prices. x_t may be a vector. Underlying these observations will be a theoretical stochastic process X_t which can, of course, be fully characterized by a (possibly

countable-infinitely dimensioned) distribution function. The initial and basic objective of time series analysis is to use the observed series x_t to help characterize or describe the unobserved theoretical sequence of random variables X_t. The similarity between this and the ideas of sample and population in classical statistics is obvious. However, the involvement of time in our sequences and the fact, or assumed fact, that time flows in a single direction does add a special structure to time-series data and it is imperative that this extra structure be fully utilized. When standing at time t, it is important to ask how will the next value of the series be generated. The general answer is to consider the conditional distribution of x_{t+1} given x_{t-j}, $j \geq 0$, and then to say that x_{t+1} will be drawn from this distribution. However, a rather different kind of generating function is usually envisaged in which the x_{t+1} is given by:

$$x_{t+1} = (\text{function of } \tilde{x}_t) + e_{t+1}, \tag{2.1}$$

where

$$\tilde{x}_t = (x_t, x_{t-1}, \ldots)$$

and the parameters of the distribution of e_{t+1} other than the mean, can depend on x_{t-j}, $j \geq 0$. It is usually overly ambitious to consider the whole distribution of e_{t+1} and, at most, the variance is considered unless e_{t+1}, or a simple transformation of it, is assumed to be normally distributed. An obviously important class of models occurs when the function in (2.1) is linear, so that:

$$x_{t+1} = \sum_{j=0}^{\infty} \alpha_{j,t} x_{t-j} + e_{t+1}. \tag{2.2}$$

For linear models, an appropriate set of characterizing statistics are the first and second moments of the process, that is the mean:

$$\mathrm{E}[X_t] = \mu_t;$$

the variance:

$$\mathrm{E}\left[(X_t - \mu_t)^2\right] = \sigma_t^2;$$

and the covariances:

$$\mathrm{E}[(X_t - \mu_t)(X_{t-s} - \mu_{t-s})] = \lambda_{t,s},$$

assuming that these quantities exist.

Given a finite amount of data and a single realization, which is the usual case in practice with economic data, it is fairly clear that one cannot estimate these quantities without imposing some further structure. A case which provides a good base situation is when the process is stationary. A process is said to be second-order stationary if the mean and variance, μ and σ^2, do not vary with time and the covariances, λ_s, depend only on the time interval between X_t and X_{t-s} rather than on time itself. A general definition of stationarity has that any group of x's, and the same group shifted by a finite time interval, have identical joint distributions. In terms of the generating function (2.1), x_t will be stationary if the form and parameters of the function do not vary through time. For the linear form (2.2) a sufficient set of conditions are that the parameters of the distribution of ε_t are time invariant and the parameters $\alpha_{j,t}$ are both time invariant and are such that the difference equation:

$$X_{t+1} = \sum_{j=0}^{\infty} \alpha_j X_{t-j},$$

is stable. An assumption of stationarity is not made because it is believed to be realistic, but because a number of important results derive from the assumption and these results can then be studied as the stationarity assumption is relaxed in useful ways.

If x_t is a univariate, stochastic process, its linear properties can be studied from knowledge of its mean, which is henceforth assumed known and to be zero, variance σ^2 and the autocovariances λ_s, or equivalently the autocorrelations $\rho_s = \lambda_s/\sigma^2$. Given a single realization x_t, $t = 1, \ldots, n$, consistent estimates of these quantities are easily found provided that the process is ergodic, which essentially means that as n increases the amount of useful information about the process continually increases. (An example of a non-ergodic process is $X_t = a\cos(bt)$ where a is a random variable with finite mean.) Although these quantities, particularly the autocorrelations, do characterize the linear properties of the process, they are not always easy to interpret or to use, if, for example, one is interested in forecasting. For many purposes there is greater interest in the generating process, or at least approximations to it. Ideally, one should be able to look at the correlogram, which is the plot of ρ_s against s, decide which is the appropriate model, estimate this model and then use it. To do this, one naturally first requires a list, or menu of possible and interesting models. There is actually no shortage of time series models, but in the stationary case just a few models are of particular importance.

The most fundamental process, called white noise, consists of an uncorrelated sequence with zero mean, that is ε_t such that $\mathrm{E}[\varepsilon_t] = 0$, $\mathrm{var}(\varepsilon_t) < \infty$ and $\mathrm{corr}(\varepsilon_t, \varepsilon_{t-s}) = 0$, all $s \neq 0$. The process can be called pure white noise if ε_t and

ε_{t-s} are independent for $s \neq 0$. Clearly a pure white-noise process cannot be forecast from its own past, and a white noise cannot be forecast linearly, in each case the optimal forecast is the mean of the process. If one's objective when performing an analysis is to find a univariate model that produces optimum linear forecasts, it is clear that this objective has been reached if a linear transformation of x_t can be found that reduces the series to white noise, and this is why the white noise process is so basic. It can be shown that any univariate stationary process can, in theory at least, be reduced uniquely to some white-noise series by linear transformation. If non-linear or multivariate processes are considered there may not be a unique transformation.

A class of generating processes, or models, that are currently very popular are the mixed autoregressive moving averages given by:

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=0}^{q} b_j \varepsilon_{t-j}, \qquad b_0 = 1,$$

where ε_t is white noise. In terms of the extremely useful backward shift operator, B, where

$$B^k x_t = x_{t-k},$$

these ARMA (p, q) models can be expressed as:

$$a(B)x_t = b(B)\varepsilon_t,$$

where

$$a(B) = 1 - \sum_{j=1}^{p} a_j B^j$$

and

$$b(B) = \sum_{j=0}^{q} b_j B^j, \qquad b_0 = 1.$$

If $q = 0$, one has an autoregressive, AR(p), model and if $p = 0$ the model is a moving average, denoted MA(q). The ε_t's are, of course not directly observable, but a model is said to be invertible if the original ε_t can be re-constructed from the observed x_t. Given a long enough series for x_t, the models are invertible if the roots of the equation $b(z) = 0$ all lie outside the unit circle.

Consider now the AR(1) model:

$$x_t = \alpha x_{t-1} + \varepsilon_t.$$

This simple difference equation has the solution:

$$x_t = \sum_{j=0}^{t+n} \alpha^j \varepsilon_{t-j},$$

if the process started up at time $t = -n$. If ε_t has zero mean and variance σ^2, then clearly the variance of x_t is:

$$\text{var}(x_t) = \left[\frac{1-\alpha^{2[n+t+1]}}{1-\alpha^2}\right]\sigma^2,$$

and x_t has mean zero. If now the starting up time is moved into the distant past, the variance of x_t tends to $1/(1-\alpha^2)$ if $|\alpha| < 1$, but increases exponentially and explodes if $|\alpha| > 1$. A borderline case, known as a random walk when $\alpha = 1$, has $\text{var}\, x_t = (t+n+1)\sigma^2$. It is clear that if $|\alpha| \geq 1$, x_t will have infinite variance. More generally, if all of the roots of $a(z) = 0$ lie outside the unit circle and the process started in the distant past, the series will be stationary, if any roots lie inside the unit circle the series will be explosive. If d roots lie on the unit circle and all others outside one has an integrated process. Suppose that x_t is generated by

$$(1-B)^d a(B) x_t = b(B)\varepsilon_t,$$

where $a(B)$ is a polynomial of order p with all roots outside the unit circle and $b(B)$ is a polynomial of order q, then x_t is said to be an integrated autoregressive–moving average series, denoted $x_t \sim \text{ARIMA}(p, d, q)$ by Box and Jenkins (1976) who introduced and successfully marketed these models. It should be noted that the result of differencing x_t d times is a series $y_t = (1-B)^d x_t$, which is ARMA(p, q) and stationary. Although, when $d > 0$ and x_t is not stationary, then these models are only a rather simple subset of the class of all non-stationary series. There has been a rather unfortunate confusion in the literature recently about distinguishing between integrated and general non-stationary processes. These terms have, incorrectly, been used as synonyms.

One reason for the popularity of the ARMA models derives from Wold's theorem, which states that if x_t is a stationary series it can be represented as the sum of two components, x_{1t} and x_{2t}, where x_{1t} is deterministic (i.e. $x_{1,t+k}$, $k > 0$, can be forecast without any error by a linear combination of $x_{1,t-j}$, $j > 0$) and x_{2t} has an MA(q) representation where q may be infinite. As an infinite series can frequently be well approximated by a rational function, the MA(∞) process may be adequately approximated by an ARMA(p, q) process with finite p and q. The ARIMA(p, d, q) models give the analyst a class of linear time series processes that are general enough to provide a good approximation to the true model, but

are still sufficiently uncomplicated so that they can be analyzed. How this is done is discussed later in this section.

Many other models have been considered. The most venerable considers a series as being the sum of a number of distinct components called trend, long waves, business cycles of various periods, seasonal and a comparatively unimportant and undistinguished residual. Many economic series have a tendency to steadily grow, with only occasional lapses, and so may be considered to contain a trend in mean. Originally such trends were usually represented by some simple function of time, but currently it is more common to try to pick up these trends by using integrated models with non-zero means after differencing. Neither technique seems to be completely successful in fully describing real trends, and a "causal" procedure, which attempts to explain the trend by movements in some other series—such as population or price—may prove to be better. The position that economic data contains deterministic, strictly periodic cycles is not currently a popular one, with the exception of the seasonal which is discussed in Section 5. The ARIMA models can adequately represent the observed long swings or business cycles observed in real economics, although, naturally, these components can be better explained in a multivariate context.

The decomposition of economic time series into unobserved components (e.g. permanent and transitory, or, "trend" and seasonal components) can be accomplished by signal extraction methods. These methods are discussed in detail in Nerlove, Grether and Carvalho (1979). In Section 6 we show how the Kalman filter can be used for this purpose.

A certain amount of consideration has been given to both non-stationary and non-linear models in recent years, but completely practical procedures are not usually available and the importance of such models has yet to be convincingly demonstrated in economics. The non-stationary models considered include the ARIMA models with time-varying parameters, the time variation being either deterministic, following a simple AR(1) process or being driven by some other observed series. Kalman filter techniques seem to be a natural approach with such models and a useful test for time-varying autoregressive parameters has been constructed by Watson and Engle (1980).

Estimation and prediction in models with time varying autoregressive parameters generated by an independent autoregressive process is a straightforward application of the techniques discussed by Chow in Chapter 20 of this Handbook. Stochastically varying moving average coefficients are more difficult to handle. Any stochastic variation in the coefficients yields a model which is not invertible as it is impossible to completely unscramble the shocks to the coefficients from the disturbance. In the moving average model this introduces a non-linear relationship between the unobservables, the disturbances and the coefficients. The Kalman filter cannot be used directly. It is possible to linearize the model and use an extended Kalman filter as Chow does in Chapter 20 for the simultaneous

equation model. The properties of the coefficient estimates and forecasts derived from this method are not yet established.

Useful classes of non-linear models are more difficult to construct, but a class with some potential is discussed in Granger and Andersen (1978). These are the bilinear models, an example being:

$$x_t = \alpha x_{t-1} + \beta x_{t-2}\varepsilon_{t-1} + \varepsilon_t.$$

When $\alpha = 0$, this particular model has the interesting property that the autocorrelations ρ_s all vanish for $s \neq 0$, and so appears, in this sense, to be similar to white noise. Thus, in this case x_t cannot be forecast linearly from its own past, but it can usually be very well forecast from its own past non-linearly. Conditions for stationarity and invertibility are known for some bilinear models, but it is not yet known if they can be used to model the types of non-linearity that can be expected to occur in real economic data.

Priestly (1980) introduces a state-dependent model which in its general form encompasses the bilinear model and several other non-linear models. The restricted and conceivably practical form of the model is a mix of the bilinear and stochastically time varying coefficient models.

Engle (1982) has proposed a model which he calls autoregressive conditional heteroscedastic (ARCH) in which the disturbances, ε_t, have a variance which is unconditionally constant, but conditional on past data may change, so that:

$$\mathrm{E}\left[\varepsilon_{t+1}^2\right] = \sigma^2,$$

but

$$\mathrm{E}\left[\varepsilon_{t+1}^2 | x_t, x_{t-1}, \ldots, x_1\right] = h(x_t, x_{t-1}, \ldots, x_1) = h_{t+1}.$$

As will be shown in the next section, ε_{t+1} is just the one step ahead forecast error x_{t+1}. The ARCH model postulates that x_{t+1} will sometimes be relatively easy to forecast from x_t, i.e. $h_{t+1} < \sigma^2$, while at other times it may be relatively difficult. This seems an attractive model for economic data.

One of the basic tools of the time series analyst is the correlogram, which is the plot of the (estimated) autocorrelations ρ_s against the lag s. In theory, the shape of this plot can help discriminate between competing linear models. It is usual practice in time series analysis to initially try to identify from summaries of the data one or just a few models that might have generated the data. This initial guess at model specification is now called the identification stage and decisions are usually made just from evidence from the data rather than from some preconceived ideas, or theories, about the form of the true underlying generating process. As an example, if a process is ARMA (p, q) with $p > 0$, then $\rho_s = \theta^s$ for s

large, with $|\theta| < 1$, but if $p = 0$, $\rho_s = 0$ for $s \geq q+1$ so that the shape of the correlogram can, theoretically, help one decide if $p > 0$ and, if not, to choose the value of q. A second diagram, which is proposed by Box and Jenkins to help with identification is the partial correlogram, being the plot of $a_{s,s}$ against s, where $a_{k,k}$ is the estimated coefficient of x_{t-k} when an kth order AR model is fitted. If $q > 0$, this diagram also declines as θ^s for s large, but if $q = 0$, then $a_{s,s} = 0$ for $s \geq p+1$. Thus, the pair of diagrams, the correlogram and the partial correlogram, can, hopefully, greatly help in deciding which models are appropriate. In this process, Box and Jenkins suggest that the number of parameters used, $p + q$, should be kept to a minimum—which they call the principal of parsimony—so that estimation properties remain satisfactory. The value of this suggestion has not been fully tested.

The Box and Jenkins procedure for identifying the orders p and q of the ARMA(p, q) model is rather complicated and is not easily conducted, even by those experienced in the technique. This is particularly true for the mixed model, when neither p nor q vanishes. Even for the pure AR or MA models difficulties are often encountered and identification is expensive because it necessitates decision making by a specially trained statistician. A variety of other identification procedures have been suggested to overcome these difficulties. The best known of these is the Akaike information criteria (AIC) in which if, for example, an AR(k) model is considered using a data set of size N resulting in an estimated residual variance $\hat{\sigma}_k^2$, then one defines

$$AIC(k) = \log \hat{\sigma}_k^2 + 2k/N.$$

By choosing k so that this quantity is minimized, an order for the AR model is selected. Hannan and Quinn (1979) have shown that this criteria provides upward-biased estimates of the order of the model, and that minimization of the criterion:

$$\phi_k = \log \hat{\sigma}_k^2 + N^{-1} 2kc \log\log N, \qquad c > 1,$$

provides better, and strongly consistent estimates of this order.

Although c is arbitrary, a value $c = 1$ appears to work well according to evidence of a simulation. So for instance, if $N = 100$ an AR(4) model would be prefered to an AR(5) model if the increase in $\hat{\sigma}^2$ is less than 2% using AIC and less than 3% using ϕ. These procedures can be generalized to deal also with mixed ARMA(p, q) models. (A critical discussion on the use of information criteria in model selection can be found in Chapter 5 of the Handbook.) Another partly automated method has been proposed by Gray, Kelly and McIntire (1978) which is particularly useful with the mixed model. Although the method lacks intuitive appeal, examples of its use indicate that it has promise. As these, and other,

automated methods become generally available, the original Box–Jenkins procedures will probably be used only as secondary checks on models derived. There is also a possibility that these methods can be used in the multiple series case, but presently they are inclined to result in very non-parsimonious models.

The identification stage of time series modeling is preceded by making an estimate of d, in the ARIMA(p, d, q) model. If $d > 0$, the correlogram declines very slowly—and theoretically not at all—so the original series is differenced sufficiently often so that such a very smooth correlogram does not occur. In practice, it is fairly rare for a value of d other than zero or one to be found with economic data. The importance and relevance of differencing will be discussed further in Section 5. Once these initial estimates of p, d and q have been obtained in the identification stage of analysis, the various parameters in the model are estimated and finally various diagnostic checks applied to the model to see if it adequately represents the data.

Estimation is generally carried out using maximum likelihood or approximate maximum likelihood methods. If we assume the ε's are normally distributed with mean zero and variance (conditional on past data) σ_ε^2, the likelihood function is proportional to:

$$\left(\sigma_\varepsilon^2\right)^{-T/2} f(\beta)\exp\left[-S(\beta, X_T)/-2\sigma_\varepsilon^2\right],$$

where β contains the parameters in $a(B)$ and $b(B)$ and now $X_T = (x_1, x_2, \ldots, x_T)'$. Analytic expressions for $f(\beta)$ and $S(\beta, X_T)$ can be found in Newbold (1974).

One of three methods, all with the same asymptotic properties, is generally used to estimate the parameters. The first is the exact maximum likelihood method, and Ansley (1979) proposes a useful transformation of the data when this method is used. The second method, sometimes called exact least squares, neglects the term $f(\beta)$, which does not depend on the data, and minimizes $S(\beta, X_T)$. The method is called exact least squares since $S(\beta, X_T)$ can be written as:

$$\sum_{t=-\infty}^{T} = \tilde{\varepsilon}_t^2,$$

where $\tilde{\varepsilon}_t = \mathrm{E}[\varepsilon_t | X_T, \beta]$. Box and Jenkins (1976) suggest approximating this by "back-forecasting" (a finite number of) the pre-sample values of ε. The third and simplest approach, called conditional least squares, is the same as exact least squares except pre-sample values of the disturbances are set equal to their unconditional expected values.

Monte Carlo evidence [see Newbold and Ansley (1979)] suggests that the exact maximum likelihood method is generally superior to the least squares methods. Conditional least squares performs particularly poorly when the roots of the MA polynomial, $b(z)$, are near the unit circle.

Once the model has been estimated diagnostic checks are carried out to test the adequacy of the model. Most of the procedures in one way or another test the residuals for lack of serial correlation. Since diagnostic tests are carried out after estimation Lagrange Multiplier tests are usually the simplest to carry out (see Chapter 12 of this Handbook). For the exact form of several of the tests used the reader is referred to Hosking (1980). Higher moments of the residuals should also be checked for lack of serial correlation as these tests may detect non-linearities or ARCH behavior.

The use of ARIMA models and the three stages of analysis, identification, estimation and diagnostic testing are due to Box and Jenkins (1976), and these models have proved to be relatively very successful in forecasting compared to other univariate, linear, time-invariant models, and also often when compared to more general models. The models have been extended to allow for seasonal effects, which will be discussed in Section 6.

A very different type of analysis is known as spectral analysis of time series. This is based on the pair of theorems [see, for instance, Anderson (1971, sections 7.3 and 7.4)] that the autocorrelation sequence ρ_s of a discrete-time stationary series, x_t has a Fourier transformation representation:

$$\rho_s = \int_{-\pi}^{\pi} e^{i\omega s} dS(\omega),$$

where $S(\omega)$ has the properties of a distribution function, and the spectral representation for x_t:

$$x_t = \int_{-\pi}^{\pi} e^{it\omega} dz(\omega),$$

where

$$\begin{aligned} E[dz(\omega)\overline{dz}(\lambda)] &= 0, && \omega \neq \lambda, \\ &= \sigma^2 dS(\omega), && \omega = \lambda, \end{aligned}$$

where $\sigma^2 = \text{var}(x_t)$. When x_t contains no purely cyclical components $dS(\omega)$ can be replaced by $s(\omega)d\omega$, where $s(\omega)$ is known as the spectral function and is given by:

$$s(\omega) = \frac{1}{2\pi} \sum_{\text{all } s} (\rho_s e^{-is\omega}).$$

The spectral representation for x_t can be interpreted as saying that x_t is the sum of an uncountably infinite number of random components, each associated with a

particular frequency, and with each pair of components being uncorrelated. The variance of the component with frequencies in the range $(\omega, \omega + d\omega)$ is $\sigma^2 s(\omega)d\omega$ and the sum (actually integral) of all these variances is σ^2, the variance of the original series. This property can obviously be used to measure the relative importance of the frequency components. Small, or low, frequencies correspond to long periods, as frequency $= 2\pi$ (period)$^{-1}$, and thus to long swings or cycles in the economy if x_t is a macro-variable. High frequencies, near π, correspond to short oscillations in the series. In one sense, spectral analysis or frequency-domain analysis gives no more information than the more conventional time-domain analysis described earlier, as there is a unique one-to-one relationship between the set of autocorrelations ρ_s, $s = 1, 2, \ldots$, and the spectral function $s(\omega)$. However, the two techniques do allow different types of interpretation to be achieved and for each there are situations where they are clearly superior. Thus, for example, if one is interested in detecting cycles or near cycles in one's data, spectral analysis is obviously appropriate.

If x_t is a stationary series and a second series is formed from it by a linear transformation of the form:

$$y_t = \sum_{j=0}^{m} g_j x_{t-j},$$

then their respective spectral representations are easily seen to be:

$$y_t = \int_{-\pi}^{\pi} e^{it\omega} g(\omega)\, dz(\omega),$$

if

$$x_t = \int_{-\pi}^{\pi} e^{it\omega}\, dz(\omega),$$

where

$$g(\omega) = \sum_{j=0}^{m} g_j z^j, \qquad z = e^{-i\omega}.$$

By considering the autocovariance sequence of y, it follows immediately that the spectrum of y_t is $g(\omega)\bar{g}(\omega)s_x(\omega)$ where $s_x(\omega)$ is the spectrum of x_t and $\bar{g}$ is the complex conjugate of g. y_t is known as a (one-sided) filter of x_t and the effect on a series of the application of a filter is easily determined in the frequency domain.

A zero-mean, white-noise series ε_t with variance of σ_ε^2 has spectrum $s_\varepsilon(\omega) = \sigma_\varepsilon^2/(2\pi)$, so that the spectrum of a white noise is flat, meaning that all frequency components are present and contribute equal proportions to the total variance. Considering a series x_t generated by an ARMA(p, q) process as a filtered version

of ε_t, that is:

$$a_p(B)x_t = b_q(B)\varepsilon_t,$$

or

$$x_t = a_p^{-1}(B)b_q(B)\varepsilon_t,$$

it follows that the spectrum of x_t is:

$$\frac{b_q(\omega)\bar{b}_q(\omega)\sigma_\varepsilon^2}{a_p(\omega)\bar{a}_p(\omega)^{2\pi}}.$$

Some applications of spectral analysis in econometrics will be discussed in Section 7. Potentially, the more important applications do not involve just single series, but occur when two or more series are being considered. A pair of series, x_t, y_t, that are individually stationary are (second-order) jointly stationary, if all cross correlations $\rho_s^{xy} = \text{corr}(x_t y_{t-s})$ are time invariant. In terms of their spectral representations, it is necessary that:

$$\begin{aligned} \mathrm{E}\left[\mathrm{d}z_x(\omega)\overline{\mathrm{d}z_y}(\lambda)\right] &= 0, & \omega \neq \lambda, \\ &= \mathrm{cr}(\omega)\mathrm{d}\omega, & \omega = \lambda, \end{aligned}$$

where x_t and y_t have spectral representations:

$$x_t = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\omega}\mathrm{d}z_x(\omega)$$

and

$$y_t = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\omega}\mathrm{d}z_y(\omega).$$

$\mathrm{cr}(\omega)$ is known as the cross spectrum and is, in general, a complex valued quantity. Interpretation is easier in terms of three derived functions, the phase $\phi(\omega)$, the coherence $C(\omega)$, and the gain $R_{xy}(\omega)$ given by:

$$\phi(\omega) = \tan^{-1}\left[\frac{\text{imaginary part of } \mathrm{cr}(\omega)}{\text{real part of } \mathrm{cr}(\omega)}\right],$$

$$C(\omega) = \frac{|\mathrm{cr}(\omega)|^2}{s_x(\omega)s_y(\omega)},$$

$$R_{xy}(\omega) = \frac{|\mathrm{cr}(\omega)|}{s_y(\omega)}.$$

When the two series are related in a simple fashion:

$$x_t = ay_{t-k} + v_t,$$

where v_t is a stationary series uncorrelated with y_{t-s}, all s, the phase diagram takes the form:

$$\phi(\omega) = k\omega.$$

This is true whether k is an integer or not, so a plot of the estimate of $\phi(\omega)$ against ω will give an estimate of the lag k in this simple model. Models relating x_t and y_t involving more complicated structures do not lead to such easily interpreted phase diagrams, this being particularly true for two-way causal relationships. The coherence function measures the square of the correlation between corresponding frequency components of the two series and is always important. For instance, it might be found that two series are highly interrelated at low frequencies ("in the long rum") but not at high frequencies ("in the short run") and this could have interesting econometric implications. The gain can be interpreted as the regression coefficient of the ω-frequency component of x on the corresponding component of y.

The extension of spectral techniques to analyze more than two series is much less well developed, although partial cross spectra can be easily determined but have been little used.

Spectral estimation has generated a considerable literature and only the rudiments will be discussed here. Since the spectral density function is given by:

$$s(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho_j e^{-ij\omega}.$$

A natural estimator is its sample counterpart:

$$\hat{s}(\omega) = \frac{1}{2\pi} \sum_{j=-T+1}^{T+1} \hat{\rho}_j e^{-ij\omega}.$$

This estimator has the desirable property of being asymptotically unbiased but also has the undesirable properties of being inconsistent and producing a rather "choppy" graph when plotted against frequency even when $s(\omega)$ is smooth. This last property follows from the fact that $\hat{s}(\omega_1)$ and $\hat{s}(\omega_2)$ will be asymptotically uncorrelated for $\omega_1 \neq \omega_2$.

To alleviate these problems $\hat{s}(\omega)$ is usually smoothed to produce an estimator $\hat{s}_k(\omega)$ given by:

$$\hat{s}_k(\omega) = \int_{-\pi}^{\pi} k(\lambda)\hat{s}(\omega - \lambda)\,d\lambda.$$

The weighting function $k(\lambda)$ is called the spectral window. It is symmetric about

ω and most of its mass is concentrated around this frequency. Specific forms for spectral windows are given in the references below.

Since $\hat{s}_k(\omega)$ is a weighted averaged of $\hat{s}(\lambda)$ for λ near ω large changes in the spectrum near ω cause a large bias in $\hat{s}_k(\omega)$. These spillover effects are called leakage, and will be less of a problem the flatter the spectrum. To avoid leakage series are often "prewhitened" prior to spectral estimation and the spectrum is then "recolored". A series is prewhitened by applying a filter to the series to produce another series which is more nearly white noise, i.e. has a flatter spectrum than the original series. So, for example, x_t might be filtered to produce a new series y_t as:

$$y_t = \phi(B)x_t.$$

The filter $\phi(B)$ may be chosen from a low order autoregression or an ARMA model. Once the spectrum of y_t has been estimated, the spectrum of x_t can be recovered by recoloring, that is:

$$\hat{s}_x(\omega) = |\phi(\omega)|^{-2}\hat{s}_y(\omega).$$

The details of spectral estimation and the properties of the estimators can be found in the books by Anderson (1971), Fishman (1969), and Koopmans (1974). There are many computer packages for carrying out spectral and cross-spectral estimation. For the length of time series generally encountered in economics computation costs are trivial.

If in the spectral representation,

$$x_t = \int_{-\pi}^{\pi} e^{it\omega} dz(\omega),$$

the random amplitudes $dz(\omega)$ are not orthogonal, so that

$$E[dz(\omega)\overline{dz}(\lambda)] = d^2F(\omega, \lambda),$$

which is not necessarily zero when $\omega \neq \lambda$, a very general class of non-stationary processes result, known as harmonizable processes. They have recently been discussed and applied to economic data by Joyeux (1979).

3. Theory of forecasting[1]

In applied economics as well as many other sciences much of the work on time series analysis has been motivated by the desire to generate reliable forecasts of future events. Many theoretical models in economics now assume that agents in

[1]This section relies heavily on Granger and Newbold (1977).

the economy optimally or "rationally" forecast future events and take actions based on these forecasts. This section will be devoted to discussing certain aspects of forecasting methodology and forecast evaluation.

Let X_t be a discrete time stochastic process, and suppose that we are at time n (n = now) and seek a forecast of X_{n+h} (h = hence). Anything that can be said about X_{n+h} at time n will obviously be based on some information set available at time n, which will be denoted by I_n. As an example, a univariate forecast might use the information set:

$$I_n' = (x_t, -\infty < t \leq n; \text{ model}),$$

where by "model" we mean the process generating the data. Any information set containing the past and present of the variable being forecast will be called a proper information set.

Everything that can be inferred about X_{n+h} given the information set I_n is contained in the conditional distribution of X_{n+h} given I_n. Typically it is too ambitious a task to completely characterize the entire distribution, and the forecaster must settle for a confidence band for X_{n+h}, or a single value, called a point forecast.

To derive an optimal point forecast a criterion is needed, and one can be introduced using the concept of a cost function. Agents engage in forecasting presumably because knowledge about the future aids them in deciding which actions to take today. An accurate forecast will lead to an appropriate action and an inaccurate forecast to an inappropriate action. An investor, for example, will forecast the future price of an asset to decide whether to purchase the asset today or to sell the asset "short". An accurate forecast implies a profit for the investor and an inaccurate forecast implies a loss. A cost function measures the loss associated with a forecast error. If we define the forecast of X_{n+h} based on information set I_n as $f^x_{n,h}(I_n)$, then the forecast error will be:

$$e^x_{n,h}(I_n) = X_{n+h} - f^x_{n,h}(I_n). \tag{3.1}$$

The cost associated with this error can be denoted as $c(e^x_{n,h}(I_n))$. (For notational convenience we will often suppress the subscripts, superscripts, and information set when they are easily inferred from the context.) A natural criterion for judging a forecast is the expected cost of the forecast error.

The most commonly used cost function is the quadratic:

$$C(e) = ae^2,$$

where a is some positive constant. This cost function is certainly not appropriate in all situations—it is symmetric for example. However, it proves to be the most tractable since standard least squares results can be applied. Many results

obtained from the quadratic cost function carry over to other cost functions with only minor modification. For a discussion of more general cost functions the reader is referred to Granger (1969) and Granger and Newbold (1977).

Standard theory shows that the forecast which minimizes the expected squared forecast error is:

$$f_{n,h} = \mathrm{E}(X_{n+h}/I_n).$$

Calculating the expected value of the conditional distribution may be difficult or impossible in many cases, since as mentioned earlier the distribution may be unknown. Attention has therefore focused on forecasts which minimize the mean square forecast error and which are linear in the data contained in I_n. Except for a brief mention of non-linear forecasts at the end of this section, we will concern ourselves only with linear forecasts.

We will first derive the optimal linear forecast of X_{n+h} for the quadratic cost function using the information set I_n' introduced above. We will assume that X_t is covariance stationary and strictly non-deterministic. The deterministic component of the series can, by definition, be forecast without error from I_n so there is no loss in generality in the last assumption. For integrated processes, X_t is the appropriate differenced version of the original series. Since the infinite past of X_t is never available the information set I_n' is rather artificial. In many cases, however, the backward memory of the X_t process [see Granger and Newbold (1977)] is such that the forecasts from I_n' and

$$I_n'' = (x_t, t = 0,1,\ldots,n; \text{ model}).$$

differ little or not at all.

The optimal forecast for the quadratic cost function is just the minimum mean square error forecast. The linear minimum mean square error forecasts from the information set I_n' will be of the form:

$$f_{n,h} = \sum_{i=0}^{\infty} c_i x_{n-i} = c(B)x_n,$$

where $c(B)$ minimizes:

$$\mathrm{E}\left[(x_{n+h} - c(B)x_n)^2\right]$$

From Wold's theorem x_t has a moving average representation:

$$x_t = b(B)\varepsilon_t, \tag{3.2}$$

where ε_t is white noise. If we define:

$$w(B) = b(B)c(B),$$

and we assume that $b(B)$ is invertible, the problem reduces to finding $w(B)$ which minimizes:

$$\mathrm{E}\left[\left(x_{n+h} - w(B)\varepsilon_n\right)^2\right].$$

It is then straightforward to show [Granger and Newbold (1977, p. 121)] that the equations which characterize $w(B)$ are:

$$w_i = b_{i+h}, \qquad i = 0,1,\ldots.$$

A compact way of writing this is:

$$w(B) = \left[\frac{b(B)}{B^h}\right]_+,$$

where "+" means ignore all negative powers of B. The linear mean square error forecast can then be written as:

$$f_{n,h} = \left[\frac{b(B)}{B^h}\right]_+ \varepsilon_n,$$

or:

$$f_{n,h} = \left[\frac{b(B)}{B^h}\right]_+ \frac{1}{b(B)} x_n. \tag{3.3}$$

Substituting (3.2) and (3.3) into (3.1) shows that the forecast error will be:

$$e_{n,h} = \sum_{i=0}^{h-1} b_i \varepsilon_{n+h-i},$$

so that the h step forecast errors are generated by a moving average process of order $h-1$. The one step ahead forecast error is just ε_{n+1} which is white noise. Furthermore, x_{n+h} can be decomposed as:

$$x_{n+h} = f_{n,h} + e_{n,h},$$

where $f_{n,h}$ and $e_{n,h}$ are uncorrelated. The variance of the forecast will therefore be bounded above by the variance of the series.

The formulae given above for the optimal univariate forecast may look rather imposing, but simple recursions can easily be derived. Note, for instance, that:

$$\begin{aligned} f_{n,h} &= f_{n-1,h+1} + b_n \varepsilon_n \\ &= f_{n-1,h+1} + b_n (X_n - f_{n-1,1}), \end{aligned}$$

so that forecasts of X_{n+h} can easily be updated as more data becomes available. A very simple method is also available for ARMA models. Suppose that x_t is ARMA(p, q) so that:

$$x_{n+h} = a_1 x_{n+h-1} + \cdots + a_p x_{n+h-p} + \varepsilon_{n+h} - b_1 \varepsilon_{n+h-1} - \cdots - b_q \varepsilon_{n+h-q}.$$

$f_{n,h}$ can be formed by replacing the terms on the right-hand side of the equation by their known or optimal forecast values. The optimal forecast for ε_{n+k} is, of course, zero for $k \geq 0$.

While univariate forecasting methods have proved to be quite useful (and popular) the dynamic interaction of economic time series suggests that there may be substantial gains from using wider information sets. Consider the forecast of x_{n+h} from the information set:

$$I_n''' = \{(x_t, y_t'), -\infty < t \leq n; \text{model}\}.$$

where y is a vector of other variables. If we assume that (X_t, Y_t') is a covariance stationary process, then an extension of Wold's theorem allows us to write:

$$z_t \equiv \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a_{11}(B) & a_{12}(B) \\ a_{21}(B) & a_{22}(B) \end{bmatrix} \xi_t,$$

where ξ is vector of white noise with contemporaneous matrix Σ, so that $A(0) = I$. The linear mean square error forecast will be of the form:

$$f_{n,h}^{x}(I_n''') = Q(B) z_n,$$

where $Q(B)$ minimizes:

$$\mathrm{E}[X_{n+h} - Q(B) z_n]^2.$$

If the matrix polynomial, $A(B)$ is invertible it can be shown that:

$$Q(B) = \left[\frac{a_1(B)}{B^h}\right]_+ [A(B)]^{-1},$$

where $a_1(B)$ is the first row of $A(B)$.

Once again, the forecast errors, $e_{n,h}(I_n''')$ will follow a moving average process of order $h-1$. Furthermore, it must be the case that:

$$\operatorname{var}\left(e_{n,h}(I_n')\right) \geq \operatorname{var}\left(e_{n,h}(I_n''')\right),$$

since adding more variables to the information set cannot increase the forecast error variance.

These optimal forecasting results have been used to derive variance bounds implied by a certain class of rational expectations models. [The discussion below is based on Singleton (1981); see also Shiller (1981) and LeRoy and Porter (1981).] The models under consideration postulate a relationship of the form:

$$P_n = \sum_{i=0}^{k} \delta_i f_{n,i}^{x}(I_n'''), \tag{3.4}$$

where the forecasts are linear minimum mean square error. In some models P_n could represent a long-term interest rate and X_n a short-term rate, while in others P_n represents an asset price and X_n is the value of services produced by the asset over the time interval.

If we define

$$P_n^* = \sum_{i=0}^{k} \delta_i x_{n+i}$$

and

$$\tilde{P}_n = \sum_{i=0}^{k} \delta_i f_{n,i}^{x}(I_n'),$$

where $f_{n,i}^{x}(I_n')$ is the linear mean square error forecast, then:

$$P_n^* = P_n + \eta_n = \tilde{P}_n + \nu_n,$$

where

$$\eta_n = \sum_{i=1}^{k} \delta_i e_{n,i}(I_n''')$$

and

$$\nu_n = \sum_{i=1}^{k} \delta_i e_{n,i}(I_n').$$

Since P_n and $\tilde{P}_n$ are linear combinations of optimal forecasts:

$$\mathrm{E}[P_n\eta_n] = \mathrm{E}[\tilde{P}_n\nu_n] = 0,$$

which implies:

$$\sigma_{P^*}^2 = \sigma_P^2 + \sigma_\eta^2 = \sigma_{\tilde{P}}^2 + \sigma_\nu^2.$$

Furthermore, since I_n' is a subset of I_n''':

$$\sigma_\eta^2 \leq \sigma_\nu^2,$$

which leaves us with the inequality

$$\sigma_{P^*}^2 > \sigma_P^2 \geq \sigma_{\tilde{P}}^2.$$

The variances $\sigma_{P^*}^2$ and $\sigma_{\tilde{P}}^2$ are then the bounds for the variance of the observed series. If σ_P^2 falls outside of these bounds the model (3.4) must be rejected. The first two variances can be calculated from the available data in a straightforward manner. Singleton proposes a method for estimating the last variance, derives the asymptotic distribution of these estimators and proposes a test based on this asymptotic distribution.

The discussion thus far has dealt only with optimal forecasts. It is often the case that a researcher has at his disposal forecasts from disparate information sets, none of which may be optimal. These forecasts could be ranked according to mean square error and the best one chosen, but there may be gains from using a combination of the forecasts. This was first noted by Bates and Granger (1969) and independently by Nelson (1972) and has been applied in a number of research papers [see, for example, Theil and Feibig (1980)].

To fix notation, consider one step ahead forecasts of x_{N+1}, denoted $f^1, f^2, \ldots, f^m$, with corresponding errors $e^1, e^2, \ldots, e^m$. Since bias in a forecast is easily remedied we will assume that all of the forecasts are unbiased. An optimal linear combined forecast is:

$$f^c = \sum_{i=1}^{m} a_i f^i,$$

where the a_i's are chosen to minimize:

$$\mathrm{E}(x_{n+1} - f^c)^2.$$

If the mean of X is not zero the resulting combined forecast will be unbiased only if:

$$\sum_{i=1}^{m} a_i = 1.$$

The papers by Bates and Granger and Nelson derive the weights subject to this constraints. This is just a constrained least squares problem.

Granger and Ramanathan (1981) point out that the constraint will generally be binding and so a lower mean square root error combined forecast is available. As an example suppose that x_t is generated by:

$$x_t = y_{t-1} + z_{t-1} + \eta_t,$$

where y_t, z_t, and η_t are independent white noise. If I_n^1 contains only past and present y and I_n^2 contains only past and present z, the optimal forecasts are:

$$f^1 = y_{n-1},$$

$$f^2 = z_{n-1}$$

and

$$f^c = f^1 + f^2.$$

The combined forecast has a mean square error equal to σ_η^2. Imposing the constraint yields:

$$\tilde{f}^c = a_1 f_1 + a_2 f_2,$$

where

$$a_1 = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_z^2}; \qquad a_2 = \frac{\sigma_z^2}{\sigma_y^2 + \sigma_z^2},$$

and the mean square error of $\tilde{f}^c$ is:

$$\sigma_\eta^2 + \frac{\sigma_y^2 \sigma_z^2}{\sigma_y^2 + \sigma_z^2} \geq \sigma_\eta^2.$$

When the weights are unconstrained the combined forecast will generally be biased. This is easily remedied. One merely expands the list of available forecasts to include the mean of X. There is no need to impose the constraint as it will be

satisfied by the unconstrained least squares solution, for the same reson that if a constant is included in an OLS regression the residuals will sum to zero.

Evaluation of forecast performance is by no means a clear-cut procedure. The discussion of optimal forecasts does however suggest some properties which are easily checked. The optimal linear forecast of X_{n+h} based on the information set I_n is the projection of X_{n+h} on the data I_n. This implies that the forecast error, $e_{n,h}$, is orthogonal to any linear combination of variables in the information set. Forecast errors can then be regressed on linear combinations of data in the information set and the estimated coefficients can be tested to see if they are significantly different from zero. Care must be taken in carrying out these tests. We showed earlier that the optimal h-step forecast errors from a proper information set followed a moving average process of order $h-1$, and therefore even under the null the residuals in this regression will not be white noise for h larger than 1. One step ahead forecast errors from proper information sets should be white noise and this is an easy property to check. The variance bounds derived above also suggest a weak test. The variance of the forecast should be less than the variance of the series being forecast.

When more than one forecast of the same quantity is available additional tests can be constructed. Forecasts can be ranked on a mean square error criterion and the best chosen. More demanding tests can also be constructed. If f is the optimal forecast from an information set I_n^f, and g is a forecast from an additional information set I_n^g, which is a subset of I_n^f, then the forecast error from f will be uncorrelated with g. A regression of the forecast error, e_f, on g should yield a coefficient which is not significantly different from zero. Equivalently, if the optimal combined forecast using f and g is formed the weights on f and g should not be significantly different from one and zero, respectively. Tests similar to these have been constructed to evaluate the forecasting performance of macro models and are briefly discussed in Section 7. A thorough discussion of these tests and others is contained in Granger and Newbold (1977, ch. 8).

We have largely been concerned in this section with linear forecasts; however, even for covariance stationary processes considerable gains can occur from considering nonlinear forecasts. Consider for example a special case of the bilinear model introduced in Section 2:

$$x_t = \beta \varepsilon_{t-1} x_{t-2} + \varepsilon_t,$$

where ε_t is white noise. The process will be covariance stationary if $\beta^2 \sigma_\varepsilon^2 < 1$ [Granger and Andersen (1978, p. 40)]. Since the lagged autocovariances are all zero, it follows that the optimal univariate linear one step ahead forecast of X_{n+1} is zero. The forecast mean square error is then:

$$\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-\beta^2\sigma_\varepsilon^2}.$$

The optimal non-linear one step ahead forecast is $\beta\varepsilon_{t-1}x_{t-2}$ which will have an expected mean square forecast error σ_ε^2.

Identification of complicated bilinear models is a difficult procedure, but the book by Granger and Andersen suggests methods which seem practical for simple models. Their procedure is to examine the autocorrelations of the squares of the residuals from linear time series models. Many nonlinear models have linear approximations with serially correlated squared residuals. If the squared residuals appear to be serially correlated it is not clear which non-linear models should be considered as alternatives. A further discussion of non-linear forecasting and forecasting non-linear transformations of the data can be found in Granger and Newbold (1976, 1977) and in Priestley (1980).

4. Multiple time series and econometric models

Econometric models (for time series data) and multiple time series models both attempt to describe or at least approximate the dynamic relationship between the variables under consideration. As mentioned in the first section the approaches taken in building these two types of models have historically been quite different. To facilitate the comparison of these approaches it is useful to introduce a variety of multiple time series representations.

Let $\boldsymbol{Z}_t$ be an $N \times 1$ vector stationary time series. Then an extension of Wold's theorem [Hannan (1970)] allows us to write:

$$\boldsymbol{Z}_t = \boldsymbol{c}(B)\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{c}(B)$ is an $N \times N$ matrix of (possibly infinite degree) polynomials in the backward shift operator and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector white noise, that is:

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt}),$$

with

$$\mathrm{E}[\boldsymbol{\varepsilon}_t] = 0$$

and

$$\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s'] = \delta_{st}\Sigma,$$

where δ is the Kronecker delta.

As was the case with the univariate model, it may be true that $\boldsymbol{c}(B)$ can be represented, or at least well approximated, by the rational function:

$$\boldsymbol{c}(B) = \boldsymbol{a}^{-1}(B)\boldsymbol{b}(B), \tag{4.1}$$

where both $\boldsymbol{a}(B)$ and $\boldsymbol{b}(B)$ are $N \times N$ matrices of finite order polynomials in B. We will assume that these matrices are of full rank, so that their inverses exist. When (4.1) is satisfied, $\boldsymbol{Z}_t$ is said to follow a vector ARMA or VARMA process of order (P, Q). P and Q are now $N \times N$ matrices with p_{ij} equal to the order of the polynomial $a_{ij}(B)$ and q_{ij} equal to the order of the polynomial $b_{ij}(B)$. The generating process for $\boldsymbol{Z}_t$ can then be written as:

$$\boldsymbol{a}(B)\boldsymbol{Z}_t = \boldsymbol{b}(B)\boldsymbol{\varepsilon}_t. \tag{4.2}$$

The AR side of (4.2) states that each component of $\boldsymbol{Z}_t$ is at least partially explained by its own past and the present and past of the other components. The whole model then states that when the lag operator $\boldsymbol{a}(B)$ is applied to $\boldsymbol{Z}_t$, the resulting vector time series is such that its autocovariances and cross covariances can be represented by the multivariate moving average model $\boldsymbol{b}(B)\boldsymbol{\varepsilon}_t$. It should be noted that the variables which are observed are the components of $\boldsymbol{Z}_t$ and that the disturbances, $\boldsymbol{\varepsilon}_t$, are at best estimated from the model, provided that the moving average part is invertible. Invertibility is satisfied in the multivariate model if $\boldsymbol{b}^{-1}(B)$ exists.

The representation (4.2) is by no means unique and normalizations must be imposed if the parameters are to be identified in the econometric sense. One source of under-identification comes from the contemporaneous relationship or causality of the data. The elements $\boldsymbol{Z}_t$ will be contemporaneously related if any of the off-diagonal elements of $\boldsymbol{a}(0)$, $\boldsymbol{b}(0)$, or Σ are non-zero. Clearly, there will be no way to tell these apart given only data on $\boldsymbol{Z}$. A common normalization sets $\boldsymbol{a}(0) = \boldsymbol{b}(0) = I$ and leaves Σ unrestricted. Others are, of course, possible. Sims (1980) for example uses the recursive form of the model for his vector autoregressions in which $a(0)$ is lower triangular and Σ is diagonal. This is a useful form for forecasting and for the vector AR model implies that the parameters can be efficiently estimated by ordinary least squares. Sufficient conditions for parameter identification in VARMA models are given in Hannan (1969).

As $\boldsymbol{a}(B)$ is assumed of full rank, (4.2) may also be written as:

$$\boldsymbol{Z}_t = \boldsymbol{a}^{-1}(B)\boldsymbol{b}(B)\boldsymbol{\varepsilon}_t.$$

If $\boldsymbol{a}^*(B)$ is the adjoint matrix associated with $\boldsymbol{a}(B)$ and $|\boldsymbol{a}(B)|$ is the determinant of this matrix. This results in the equivalent model:

$$|\boldsymbol{a}(B)|\boldsymbol{Z}_t = \boldsymbol{a}^*(B)\boldsymbol{b}(B)\boldsymbol{\varepsilon}_t,$$

and the jth equation of this system is:

$$|\boldsymbol{a}(B)|\boldsymbol{Z}_{jt} = \alpha_j(B)\boldsymbol{\varepsilon}_t, \qquad j = 1, \ldots, N,$$

where $\alpha_j(B)$ is the jth row of $a^*(B)b(B)$. If no cancellation of factors of the form $(1-\beta B)$ from both sides of these equations occurs, it follows that all the single series ARMA (p,q) models for the components of Z_t will have identical AR parts, and further that p and q will be very large if the number of components is large. As neither of these features is actually observed, this suggests that considerable cancellations do occur or that the present single series modeling techniques tend to choose models that are too simple. Zellner and Palm (1976) and Protheo and Wallis (1976) have suggested that the common AR property can be utilized to indicate relevant constraints on the form of the matrix $a(B)$ in the full model (4.2), but the technique has been applied only to small systems so far. A possible limitation to this technique can be seen by noting that the Z_{jt}'s could all be univariate white noises, but still be related through a model of the form (4.2), although this model will be constrained so that $|a(B)|$ and the moving average process implied by $\alpha_j(B)\varepsilon_t$ are equal for all j. Such constraints are not easily used in practice.

Time series identification, that is the choice of p and q, for VARMA models is a difficult task and completely satisfactory methods are not yet available. Tiao et al. (1979) suggest a method similar to univariate methods of Box and Jenkins which is practical for AR or MA models. Mixed models are substantially more difficult. A procedure for bivariate models is proposed in Granger and Newbold (1977). A computer package, Tiao et al. (1979), is available for estimating small scale (up to five series) VARMA models.

A model more familiar to traditional econometricians is achieved by using the partition:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix},$$

where the lag operators have not been shown for notational convenience. If it is now assumed that $a_{21} \equiv 0$, $b_{12} \equiv 0$, and $b_{21} \equiv 0$, one obtains the two sets of equations:

$$a_{11}(B)y_t + a_{12}(B)x_t = b_{11}(B)\varepsilon_{1t} \tag{4.3}$$

and

$$a_{22}(B)x_t = b_{22}(B)\varepsilon_{2t}. \tag{4.4}$$

If, furthermore, there are no contemporaneous correlations between the components of the white-noise vector ε_{1t} and the white noise vector ε_{2t}, the Z_t is decomposed into x_t and y_t, where the components of x_t are called exogenous. The question of how exogeneity should be defined and tested is discussed in Chapter 18, on causality, in this Handbook. Alternative definitions of exogeneity can be found in Engle, Hendry, and Richard (1981). The correct division of

variables into these two classes is clearly important for forecasting, as well as other purposes. Equations (4.3) and (4.4) provide the link between times series and econometric models. Equation (4.3) can be viewed as the structural form of a dynamic simultaneous equation model, while (4.4) describes the evolution of the exogenous variables. Traditionally, the existence of the subsystem (4.4) is not considered, as the exogenous variables are said to be "generated outside of the system." In the time series literature, systems such as (4.3) are now being called ARMAX systems, for autoregressive-moving average with exogenous variables.

Although the structural form (4.3) is of fundamental importance, some other derived models are also of interest. Denote $\boldsymbol{a}_{110}(B) = \boldsymbol{a}_{11}(B) - \boldsymbol{a}_{11}(0)$, then (4.3) may be written either as:

$$\boldsymbol{y}_t = -\boldsymbol{a}_{11}^{-1}(0)\boldsymbol{a}_{110}(B)\boldsymbol{y}_t + \boldsymbol{a}_{11}^{-1}(0)\boldsymbol{a}_{12}(B)\boldsymbol{x}_t + \boldsymbol{a}_{11}^{-1}(0)\boldsymbol{b}_{11}(B)\boldsymbol{\varepsilon}_{1t},$$

which is known as the reduced form, or as:

$$\boldsymbol{y}_t = -\boldsymbol{a}_{11}^{-1}(B)\boldsymbol{a}_{12}(B)\boldsymbol{x}_t + \boldsymbol{a}_{11}^{-1}(B)\boldsymbol{b}_{11}(B)\boldsymbol{\eta}_{1t}, \tag{4.5}$$

which has been called the final form, a multidimensional rational-distributed lag model, or of a unidirectional transfer-function form. In the reduced form, endogenous variables are explained by "predetermined variables"—that is, exogenous and lagged endogenous variables—whereas in the final form y_t appears to be explained by just the exogenous variables. If parameter values are known, or have been estimated, both the reduced form and the final form can be used to produce forecasts. The reduced form used the information set $I_n^{(1)}$: $[x_{n-j}, y_{n-j}, j \geq 0]$, plus forecasts of exogenous variables and the final form appears to use just $I_n^{(2)}$: $[x_{n-j}, j \geq 0]$, plus exogenous variable forecasts. However, as is easily seen from (4.5), the use of $I_n^{(2)}$ will generally produce forecasts with errors that are not white noise. These forecasts can then be improved by modeling the residuals, but to do this earlier values of the residuals are required and to know this earlier values of y_t are needed, so that effectively one ends up using $I_n^{(1)}$. As situations are rare in which past values of exogenous variables are available, but not the past values of endogenous variables, the proper information set $I_n^{(1)}$ is the appropriate one in most cases.

Traditionally, econometricians have viewed their task as specifying and estimating the model (4.3) while ignoring (4.4). The time series analyst, on the other hand, would identify and estimate both (4.3) and (4.4). To the econometrician, the parameters of (4.3) were thought to be the most important as these presumably contained the sought after information about the working of the economy. These parameters could then be subjected to hypothesis tests, etc. Time series analysts, being primarily interested in forecasting and not economic theory, required both (4.3) and (4.4) for their purpose. Lucas (1976) showed that the parameters of (4.3)

were in general not the parameters of economic interest. He persuasively argued that the important economic parameters could not be deduced without knowledge of the process generating the exogenous variables. The main point of Lucas is that the parameters of (4.3) are not structural at all. They will in general be functions of underlying structural parameters and the parameters of (4.4). The Lucas critique has spawned a new class of econometric models in which the time series properties of the exogenous variables play a crucial role. Examples can be found in Wallis (1980) and Sargent (1981).

Other clear differences between the time series and classical econometric approaches are the size of the information sets used and the intensity with which they are analyzed. Time series models often involve just a few series, but a wide variety of different lag structures are considered. Classically, econometric models involved very large numbers of series, a model of 400 equations now being classified as moderate in size, but are sparse in that most variables do not enter most equations. To the time series analyst's eyes, econometric models involve remarkably few lags. It has been said that when a time series analyst is unhappy with his model, he adds further lagged terms, but an unhappy econometrician is inclined to add further equations. One reason why econometricians rely heavily on an economic theory is that they have so many variables, but usually with rather small amounts of data, so that it would be impossible to consider a wide enough variety of models to be able to get anywhere near the true model. The use of the theory severely limits the number of alternative model specifications that need to be considered. Thus, the theory effectively greatly expands the available data set, but the difficulty is that if an incorrect theory is imposed an incorrect model specification results.

A further use of time series analysis in econometric model building is based on the precept that one man's errors may be another man's data. Thus, the residuals from an econometric model can be analyzed using time series methods to check for model mis-specification. Calling the procedure TSAR, for time series analysis of residuals, Ashley and Granger (1979) looked at the residuals from the St. Louis Federal Reserve Bank Model. Some of the individual residual series were found not to be white-noise and so could be forecast from their own past, and some residuals could be forecast from other residuals, suggesting missing variables, model mis-specification and inefficient estimation. The classification of some variables as exogenous was also found to be questionable.

5. Differencing and integrated models

An example of differences in attitudes between time series analysts and the classical econometricians concerns the question of whether the levels or changes of economic variables should be modeled. If one has a properly specified model in

levels, then there will correspond an equally properly specified model in changes. Forecasting from either will lead to identical results, for example, by noting that next level equals next change plus current level. However, if it is possible that the model is mis-specified, which is certainly a sensible viewpoint to take, there can be advantages in using differenced data rather than levels. The occurrence of spurious relationships between independent variables has been known for a long time and was documented again, using theory and simulation, by Granger and Newbold (1974). There it was shown, for example, that if x_t and y_t were each ARIMA (p, d, q), with $d = 1$, but independent, then regressions of the form:

$$x_t = \alpha + \beta y_{t-k} + \varepsilon_t,$$

when estimated by ordinary least squares would frequently show apparently significant β and R^2 values. The problem can be seen by considering the null hypothesis, $\beta = 0$, which implies $\varepsilon_t = x_t - \alpha$. This shows that ε_t is serially correlated under the null so that standard t-tests based on ordinary least squares are not appropriate. Estimation methods which assume ε_t is AR(1) improve matters, but do not totally remove the problem, as spurious relationships can still occur. Clearly, if a sufficiently general model is allowed for the errors, the problem is less likely to occur, but if the dependent variable x_t has infinite variance, as occurs when $d = 1$, but the model for ε_t only allows finite variance, then spurious relationships are often found. If all series involved are differenced, the residual need not be white noise, so that ordinary least squares is not efficient, but now at least the change series and the residual all have finite variance. Plosser and Schwert (1977, 1978) have shown that, in a sense, over-differencing is less dangerous than under-differencing and have provided illustrations using real data of spurious relationships and the effects of differencing. Using differenced data is not, of course, a general panacea and, as Plosser and Schwert state "the real issue is not differencing, but an appropriate appreciation of the role of the error term in regression". As some econometricians were traditionally rather casual about the error specification, to the eyes of a time series analyst, until recently the possibility that apparently significant relationships were spurious or weaker than they appear remained.

Despite these results, some econometricians have been reluctant to build models other than in levels or have rejected the idea of differencing all variables. Partly this is because they feel more comfortable in specifying models in levels from their understanding of economic theory and also because differencing may not always seem appropriate, particularly when non-linear terms are present or if a change in one variable is to be explained by the difference between the levels of two other variables. Another reason for this reluctance is that econometricians have become used to extremely high R^2, or corrected R^2, values when explaining levels, but R^2 often falls to modest, or even embarrassingly low values, when

changes in a variable are explained. Partly this is due to the removal of spurious relationships, but is largely due to the fact that a very smooth, high-momentum variable, such as many levels, are very well explained from past values of this variable, but this is no longer true with the highly variable change series. An extreme case is stock market prices, the levels following a random walk and the changes being white noise, or very nearly so. Econometricians have also been worried that differencing may greatly reduce or even largely remove the very important low-frequency component, corresponding to the long-swings and the business cycle. This can certainly occur if one over-differences, but should not be a problem if the correct amount of differencing occurs to reduce the series to an ARMA generated sequence. Differencing may also exacerbate errors in variables problems, but the presence of errors in variables can often be tested, and these tests can be carried out on the differences as well as the levels. There has also been some debate about the usefulness of differencing by time series analysts. It has been pointed out that if a series has a mean, then this mean cannot be reconstructed from the differenced series, but this would not be so if the difference operator $(1-B)$ is replaced by $(1-\alpha B)$ with α near, but less than, one. The obvious response is that an ARIMA series need not possess a mean.

A way of generalizing this discussion in a potentially useful fashion follows by noting that differencing a series d times means that the spectrum of the series is multiplied by:

$$|1-z|^{2d},$$

where

$$z = \varepsilon^{i\omega}.$$

If a series x_t has a spectrum of the form:

$$|1-z|^{-2d} f(\omega),$$

where $f(\omega)$ is the spectrum of a stationary ARMA series, it will be said to be integrated of order d, and denoted $x_t \sim I(d)$. Note that x_t needs to be differenced d times to become stationary ARMA. As just defined, d need not be an integer and one can talk of fractional differencing a series if a filter of the form $a(B) = (1-B)^d$ is applied to it. It has been shown that integrated series, with non-integer d, arise from the aggregation of dynamic microvariables and from large dynamic systems [see Granger (1980a)].

When $d \geq \frac{1}{2}$, x_t will have infinite variance and if $d < \frac{1}{2}$, the series has finite variance. An integrated series with $d \geq \frac{1}{2}$ will be inclined to be identified by standard Box–Jenkins techniques as requiring differencing. Note that if also

$d < 1$, the differencing will produce a series whose spectrum is zero at zero frequency. Thus, the time series analysts will, in a sense, be correct in requiring differencing to remove infinite variance, but the econometricians' worries about losing their critical low-frequency components are well founded. The proper procedure is, of course, to fractionally difference, provided that the correct value of d is known. The best way to estimate d has yet to be determined, as has the importance and actual occurrence of integrated series with non-integer d.

Possible use of fractional integrated models, if they occur in practice, is in long-run forecasting. It can easily be shown that if the MA(∞) model corresponding to $x_t \sim I(d)$ is considered, then the coefficients will decline in the form:

$$b_j \sim Aj^{d-1},$$

whereas a stationary ARMA(p, q) model, with infinite p and q, will have coefficients declining at least exponentially, i.e.

$$b_j \sim A\theta^j, \qquad |\theta| < 1.$$

This "long-memory" property can be utilized to improve long-run forecasts in a simple fashion, once d is known or has been reliably estimated.

6. Seasonal adjustment

Many important economic series show a consistent tendency to be relatively high in one part of the year and low in another part, examples being unemployment, retail sales, exports, and money supply. It is fairly uncontroversial to say that a series contains seasonal variation if its spectrum shows peaks, that is extra power, at the seasonal frequencies, which are:

$$2\pi\frac{j}{12}, \qquad j = 1, \ldots, 6,$$

for monthly series. For some series, the seasonal component is an important one, in that the seasonal frequencies contribute a major part of the total variance. For reasons that are not always clearly stated, many econometricians feel that if the seasonal component is reduced, or removed, analysis of the remaining components becomes easier. Presumably, the seasonal part is considered to be economically unimportant or easily understood, but that leaving it in the series confuses the analysis of the more important low-frequency business cycle components. By "seasonal adjustment" is meant any procedure that is designed to remove, or reduce, the seasonal component. The problem of how best to design seasonal

adjustment procedures is a very old one and it has generated a considerable literature. Although much progress has been made the problem can hardly be classified as solved. Two excellent recent references are the extensive collection of papers and discussions edited by Zellner (1979) and the survey by Pierce (1980).

Much of the discussion of seasonal adjustment begins with the additive decomposition of an observed series y_t into two unobserved components:

$$y_t = n_t + s_t,$$

where s_t is strongly seasonal—so that its spectrum is virtually nothing except peaks at the seasonal frequencies and n_t is non-seasonal. For this model, "seasonal adjustment" is any procedure which yields an estimate of the non-seasonal component. If this estimate is based on an information set which contains only the past, present, and possibly future values of y_t, the method is called auto-adjustment. A procedure based on a wider information set, called causal adjustment, will be discussed at the end of this section. Most of the literature on seasonal adjustment concerns auto-adjustment procedures and these are by far the most widely used methods. Consequently, much of our discussion will be devoted to these methods.

Early methods of seasonal adjustment relied on the additive decomposition above, and assumed that s_t followed a periodic deterministic process, an example for monthly data being:

$$s_t = \sum_{i=1}^{12} \alpha_i D_{ti},$$

where the D_{ti}'s are a set of monthly dummy variables or sine and cosine terms. The non-seasonal component was assumed to be composed of a "trend" and "irregular" component. These components were approximated by a polynomial in t and white noise. The seasonal component in this model can be estimated using standard regression techniques. Subtracting this estimate from the observed series yields an estimate of the non-seasonal component. This method and its statistical properties are discussed in Jorgenson (1964, 1967).

The causes of seasonal fluctuations, e.g. weather, and the inspection of estimated spectra for economic time series suggest that the deterministic model for s_t is a poor one. A popular approach is to assume that each component is stochastic and generated by an ARMA model. (The possible need to difference the series can be handled, but introduces further complications that will not be discussed here. More details can be found in the references given above.) Thus, we can write:

$$a_n(B)n_t = b_n(B)\eta_t$$

and

$$a_s(B)s_t = b_s(B)\varepsilon_t,$$

where η_t and ε_t are independent white noise:

$$a_n(B) = 1 - a_1^n B - a_2^n B^2 - ,\ldots, - a_{p_n}^n B^{p_n},$$
$$b_n(B) = 1 - b_1^n B - b_2^n B^2 - ,\ldots, - b_{q_n}^n B^{q_n},$$

and $a_s(B)$ and $b_s(B)$ are similarly defined. The polynomials are such that s_t is strongly seasonal, so that

$$s_s(\omega) = \frac{b_s(z)\bar{b}_s(z)}{a_s(z)\bar{a}_s(z)} \frac{\sigma_\varepsilon^2}{2\pi}, \qquad z = e^{-i\omega},$$

has most of its power concentrated around the seasonal frequency and n_t is non-seasonal. The implied model for y_t is:

$$a(B)y_t = b(B)e_t,$$

where $a(B) = a_s(B)a_n(B)$ if $a_s(B)$ and $a_n(B)$ have no common roots, and $b(B)e_t$ is a moving average having the same autocovariances as $a_n(B)b_s(B)\varepsilon_t + a_s(B)b_n(B)\eta_t$.

Since only the sum of n_t and s_t is observed it is impossible to deduce the values of the components if both σ_ε^2 and σ_η^2 are non-zero. We will denote the seasonal adjustment error at time t by:

$$\sigma_t = n_t - \hat{n}_t = \hat{s}_t - s_t,$$

where $\hat{s}_t$ and $\hat{n}_t$ are the estimated values of the components. The linear estimate of n_t which minimizes the mean square seasonal adjustment error is the projection of n_t on the available data (conditional expected value if y_t is normal). If an entire realization of y_t is available the optimal linear estimate of the seasonally adjusted series is then:

$$\hat{n}_t = P(n_t | y_k, -\infty < k < \infty) = V(B)y_t,$$

where P is the projection operator and [Weiner (1949), Whittle (1963), Grether and Nerlove (1970)]:

$$V(z) = \frac{\text{spectrum of } n_t}{\text{spectrum of } y_t},$$

where $z = e^{-i\omega}$.

Several properties of the optimal linear estimate follow immediately. First, $\hat{n}_t$ is obtained from a time invariant linear filter applied to y_t, so that the coherence between $\hat{n}_t$ and y_t is one. Second, the filter is symmetric, $v_j = v_{-j}$, implying that

the phase between y_t and $\hat{n}_t$ is zero. Finally, the spectrum of $\hat{n}_t$ is

$$s_{\hat{n}}(\omega) = |V(z)|^2 s_y(\omega)$$
$$= \frac{s_n(\omega)}{s_y(\omega)/s_n(\omega)},$$

and since

$$s_y(\omega) = s_n(\omega) + s_s(\omega),$$
$$s_{\hat{n}}(\omega) = \frac{s_n(\omega)}{1 + s_s(\omega)/s_n(\omega)},$$

so that

$$s_{\hat{n}}(\omega) \leq s_n(\omega).$$

The spectrum of $\hat{n}_t$ will be substantially less than the spectrum of n_t over those frequencies where the spectrum of s_t is large relative to the spectrum of n_t. Since this occurs at the seasonal frequencies the spectrum of the adjusted series will contain "dips" at these frequencies. Equivalently, the adjusted series will have negative autocorrelations at the seasonal lags. The "optimal" procedure will tend to "overadjust" for seasonality.

This optimal filter cannot be used for obvious reasons. The parameters of the model and hence the elements of $V(B)$ are rarely known, and a complete realization of y_t is never available. Since the process is stationary $v_j = v_{-j} \approx 0$ for large j implying that the last problem is most serious near the beginning and end of the sample.

Pagan (1975) and Engle (1979) overcome this problem through the use of the Kalman filter and smoother. The Kalman filter produces linear minimum mean square error estimates of n_t using observed data up through time t. The smoother optimally updates these estimates as data beyond time t becomes available. (The Kalman filter and smoother are discussed in detail in Chapter 20 of this Handbook.) To implement the filter the model is written in state space form. Although moving average terms can easily be handled [see Harvey and Phillips (1979)] it is notationally convenient to assume that $b_s(B) = b_n(B) = 1$. With this assumption the model can be written as:

$$y_t = \begin{bmatrix} 1'_{p_n} & 1'_{p_s} \end{bmatrix} \begin{bmatrix} \tilde{n}_t \\ \tilde{s}_t \end{bmatrix},$$

$$\begin{bmatrix} \tilde{n}_t \\ \tilde{s}_t \end{bmatrix} = \begin{bmatrix} \phi_n & 0 \\ 0 & \phi_s \end{bmatrix} \begin{bmatrix} \tilde{n}_{t-1} \\ \tilde{s}_{t-1} \end{bmatrix} + \begin{bmatrix} 1_{p_n} & 0 \\ 0 & 1_{p_s} \end{bmatrix} \begin{bmatrix} n_t \\ \varepsilon_t \end{bmatrix},$$

where 1_k is a k vector with one as its first element and all other elements zero:

$$\tilde{n}_t' = \left[n_t, n_{t-1}, \ldots, n_{t-p_n+1} \right],$$
$$\tilde{s}_t' = \left[s_t, s_{t-1}, \ldots, s_{t-p_s} \right],$$

and

$$\phi_k = \left[\begin{array}{c|c} a_1^k \quad a_2^k \ldots a_{p_k-1}^k & a_{p_k} \\ \hline & 0 \\ & 0 \\ I_{(p_k-1)} & \vdots \\ & 0 \end{array}\right], \quad \text{for } k = n \text{ or } s.$$

As Engle (1979) notes, this formulation has several advantages. Computationally it is easier to implement than the Weiner filter, which requires a factorization of the spectral density of y [see Nerlove, Grether and Carvalho (1979)]. The model is also more general as a slight modification will allow weakly exogenous variables to appear as explanatory variables for n_t and s_t. Models with deterministic components can easily be handled. The filter also insures that the revisions made in n_t at time $t+k$ follow a (time varying) moving average process of order $k-1$. This follows since the revision will be a (time varying) linear function of $e_{t+1}, e_{t+2}, \ldots, e_{t+k}$.

The filter does require a value of the mean and variance of n_0 and s_0 to begin the recursions. In the case under consideration these components are covariance stationary and the correct starting values are just the unconditional means and variances. For non-stationary models the initial values can be estimated as nuisance parameters, as described in Rosenberg (1973) or Engle and Watson (1981b).

Since the parameters of the model are rarely known, they will generally need to be estimated prior to the adjustment process. If ε_t and η_t are assumed to be normally distributed, the parameters can be estimated using the maximum likelihood methods discussed in Chapter 20 of this Handbook. The scoring algorithm presented in Engle and Watson (1981a) and the EM algorithm discussed in Engle and Watson (1981b) have been successfully used in similar models.

There are of course many ways to additively decompose y_t into two uncorrelated components. The parameters of the model will not in general be identified. Identification can sometimes be achieved by assuming specific forms for the processes as in Engle (1979), or by finding a representation which minimizes the variance of the seasonal component as in Pierce (1979).

Some of the other approaches to seasonal adjustment rely on models which have parameters varying in a seasonal manner, such as the cyclo-stationary models investigated by Parzen and Pagano (1978), while others have the amplitude of the seasonal changing with the size of other components, such as the multiplicative and the harmonizable models. Havenner and Swamy (1981) propose a model similar to the deterministic model discussed above, but they allow the regression coefficients to vary stochastically. When some of these models are employed the concept of seasonal adjustment can become rather confused.

The most widely used program for seasonal adjustment is the Census Bureau's X-11. The program consists primarily of a set of symmetric linear filters applied to the data, but also has features which correct for the number of trading days and "extreme" values. For recent data the symmetric filter is inappropriate and special "end weights" are used. Young (1968) presents a symmetric linear filter which approximates the filter used by X-11, and Cleveland and Tiao (1976) present models for which X-11 is approximately optimal. Details on the characteristics of X-11 can be found in Shiskin, Young, and Musgrave (1967) and Kupier (1979). A discussion of the models for X-11 is presented in the survey paper by Pierce.

In practice, the use of seasonally adjusted data can lead to considerable modeling problems. Many techniques, including X-11, will usually insert "over-adjustment problems", such as the above mentioned negative autocorrelations at seasonal frequencies and the relationships between pairs of series can be considerably disturbed, as various studies have indicated. Partly this is due to the use of robust techniques, which attempt to reduce the relevance of outliers. When actual outliers occur, these methods are valuable, but if over-used, as in X-11, the resulting non-linearities that are introduced can have serious consequences for modeling relationships, for parameter estimation, for causality testing and for forecasting.

Godfrey and Karreman (1967) present evidence that the methods of adjustment often used in practice will have no unfortunate effects on low-frequency components (that is components with frequencies lower than the seasonal frequency), but that all other components are badly affected, even non-seasonal higher-frequency components. The original components with frequencies higher than the seasonal frequencies are partly replaced with variables uncorrelated with them, so that coherences between the original non-seasonal components and the corresponding components of the adjusted series are reduced. This suggests that modeling pairs of seasonally adjusted series can lead to difficulties, and Newbold (1981) presents convincing evidence that this does occur. Wallis (1974, 1979) and Sims (1974) have discussed this problem in detail. Their conclusions suggest that in general it is preferable to use seasonally unadjusted data and explicitly model the seasonality.

The question of how to evaluate a seasonal adjustment procedure is not an easy one, partly because the seasonal and non-seasonal components introduced above are not clearly distinguished. A white-noise series has a flat spectrum and, thus, has some power at seasonal frequencies. The seasonal component is thought of as giving extra power at seasonal frequencies, over and above that provided by the non-seasonal component. However, this statement does not provide enough information to ensure a unique decomposition of a given series into seasonal and non-seasonal components. A similar criterion applies to the simple criterion that a series, after adjustment, has no peaks remaining in its spectrum. A clearer criterion is to require that the variance of the seasonal component, or a suitable transformation of it, should be minimized. This criterion can be characterized in either time or frequency domains and in a sense removes no more than necessary to achieve no seasonality. When one knows the correct model, or a reasonable approximation to it, such a criterion can be used to provide a good seasonal adjustment procedure. However, if the assumed model does not approximate the true world, an inappropriate adjustment may occur.

To evaluate an adjustment procedure, it has been suggested that spectral techniques are the most appropriate and that, (a) the adjusted series should have neither peaks nor dips (over adjustment) at seasonal frequencies, and (b) if the adjustment procedure is applied to a non-seasonal series, the cross spectrum between the original and the adjusted series should have a coherence near one and a phase near zero at all frequencies. Although these appear to be sensible criteria, as shown above the "optimal" adjustment method mentioned earlier will not obey them, producing dips in the spectrum at seasonal frequencies or, equivalently, negative autocorrelation at seasonal lags. This merely means that a pair of "sensible" criteria are inconsistent, but it does leave the choice of proper criteria for the selection and evaluation of techniques for further consideration.

The methods discussed above have all been "auto-adjustment," in that just the observed series x_t has been utilized. As one must expect the seasonal components to be, at least partially, the results of various causal variables a sounder approach would be to seasonally adjust in a multivariate context. Thus, if the weather causes the seasonality in Chicago house construction, it should be natural for econometricians to model this relationship. The effects of a severe winter, for example, are then directly allowed for rather than being considered as some vague, unexplained outlier. Of course, it is by no means easy to correctly model the required relationships, particularly as the series involved will all be strongly seasonal and the use of causal adjustment procedures would be far too expensive for the government to use on all of the series that are said to need adjustment. Nevertheless, if an econometrician is anxious to produce a really sound model, it is advisable to use unadjusted, raw data and to build seasonal causal terms in the model. However, even then the data may still need application of a seasonal

adjustment procedure as some causes could be unobservable, but if one does it oneself at least the methods used is under one's own control and need produce less unpleasant surprises than the use of an "off-the-shelf" technique. Further discussion of these points may be found in the papers by Granger and Engle in the volume edited by Zellner mentioned above.

7. Applications

In this section a few examples of the way in which time series techniques have been applied to economic data will be briefly discussed. It would be virtually impossible to survey all of the applications that exist. Two applications that will not be discussed, although they are currently very much in vogue, are testing for causality and the use of Kalman filter techniques for investigating time-varying parameter models, as these are described in Chapters 18 and 20 of this Handbook. Additional applications using frequency domain techniques can be found in Granger and Engle (1981).

The most obvious, and oldest, application is to model a single series to provide what are termed "naive" forecasts against which the forecasts from a full-scale econometric model can be compared. Of course, the comparison is not strictly fair, as the econometric model uses a much larger information set, and also has the "advantage" of being based on an economic theory, but, nevertheless, econometricians have behaved as though they believe that such naive models are worthy forecasting opponents. In fact, the econometric models have found it difficult to beat the time-series forecasts, an example being Cooper (1972), who used only AR(4) models. More recently, the econometric models have performed relatively better, although a more stringent criterion suggested in Granger and Newbold (1977, ch. 8), involving the combination of forecasts, would still probably suggest that there is still room for considerable improvement by the econometric models. It will be interesting to continue to compare forecasts from the two types of model, as each is certainly improving through time.

More natural comparisons are between econometric models and multivariate time series, although the best way to specify the latter is still uncertain. Some examples are the papers by Zellner and Palm (1974), Sargent (1981) and Taylor (1979). No complete comparison of relative forecasting abilities is available at this time. Multivariate time series techniques can also be used to measure the importance, in terms of improved forecasting ability, of adding further variables to the model. An obvious example is to ask how useful is anticipation data. The technique used is the same as that developed for causality testing, as discussed in Chapter 18 of this Handbook. The results are sometimes rather surprising, such as the weak relationships found between some financial series by Pierce (1977). Neftci (1979) investigated the usefulness of the NBER leading indicator for

forecasting the index of industrial production (IIP). He modeled IIP in terms of its own lags and the added various leading indicators to the model. Using post-sample forecasts, he found that the leading indicators did not improve forecasts for "normal" times, but did help during the recession year of 1974. The results thus agree with the NBER claims about the usefulness of this indicator series at turning points, but nothing more. Auerbach (1981) studied the usefulness of the leading indicator series in predicting changes in both IIP and the adult civilian unemployment rate. Based on both in-sample fit and forecasting performance he found the leading indicator series useful, but his in-sample results suggest that it may be possible to choose better (possibly time varying) weights for the components of the leading indicator series.

The ARCH model introduced in Section 2 has been used in a number of applications. Engle (1980, 1982) has shown that there are significant ARCH effects in U.S. and U.K. inflation data, and Engle and Kraft (1981) derive conditional multiperiod forecast variances from an autoregressive model where the disturbance follows an ARCH process. Robbins (1981) estimates a model in which the conditional variance of excess returns for short rates affects the liquidity premium for long rates. Engle, Granger and Kraft (1981) use a multivariate ARCH model to compute optimal time varying weights for forecasts of inflation from two competing models.

The obvious applications of univariate spectral analysis are to investigate the presence or not of cycles in data. Thus, for example, Hatanaka and Howrey (1969) looked for evidence of long swings or long cycles in the economy, by asking if there were peaks in the spectrum corresponding to such cycles. The results were inconclusive, because very long series would be required to find significant peaks, particularly against the "typical spectral shape" background, corresponding to the high power at low frequencies found with ARIMA (p, d, q) models, $d > 0$, which we often observed for the levels of economic macro variables. A related application is to compare the estimated spectral shape with that suggested by some theory. For example, the random-walk theory of stock market prices suggests that price changes should be white noise and thus have a flat spectrum. Granger and Morgenstern (1970) found evidence that was generally in favor of the hypothesis, although a very slight evidence for a seasonal in price changes was occasionally observed. Estimated spectra of a wide range of economic series give no evidence of strict cycles except for the seasonal component. Howrey (1972) calculated the spectra of major variables implied by the Wharton model and compared them to the typical spectral shape, and generally found the econometric model did produce the correct spectral shape.

The power spectrum is obviously useful in consideration of the seasonal, both to find out if a series contains a seasonal component, to measure its strength and also to investigate the effects of seasonal adjustment. One of the very first applications of frequency domains techniques to economic data was by Nerlove

(1964) investigating these aspects of seasonality. He also used the spectrum to define the seasonal component in a similar way to that used in Section 6. He gave clear indication that seasonal adjustment could disrupt the data in an unfortunate manner with the follow-up study by Godfrey and Karreman (1967) providing further illustrations of this problem.

The first application of cross-spectral analysis in economics were by Nerlove (1964) on seasonals and by Hatanaka in Granger and Hatanaka (1964), who considered the leads and strength of the relationship between the NBER leading indicators and the level of the economy. Hatanaka found some coherence at low frequencies, but the leads observed in the phase diagram were less than found by the NBER using less sophisticated methods. A later investigation of leading indicators by Hymans (1973) also used spectral methods. The results threw some doubts on the usefulness of several of the components of the index of leading indicators and using the observed coherence values an alternative weighted index was proposed, which would seem to be superior to that now in use. Most subsequent applications of cross-spectral analysis try simply to measure the extent to which pairs of series are related and whether or not there is evidence for a simple lag. Examples may be found in Labys and Granger (1970). When there is a feedback relationship between the variables, the lag structure cannot be determined, and so difficulties in interpretation frequently occur.

The Fourier transform of a stationary series allows one to look at the different frequency components of the series, at least to some extent. This idea was used in Granger and Hatanaka (1964) to test for stationarity by considering the possibility of the amplitude of the frequency components varying through time. By isolating frequency components in a group of series, the possibility of the relationships between the series varying with frequency can be analyzed. Calling the technique band spectrum regression, Engle (1974) considered a simple time-domain regression, transformed it into the frequency domain and then used a test similar to the Chow test for structure stability, to see if relationships were frequency dependent. The method is an obvious generalization of the familiar decomposition into "permanent" and "transitory" components and has similar interpretational advantages. In Engle (1978) the technique was applied to a variety of wage and price series and it was found, for example, that "the effect on prices of a low-frequency change in wages is much greater than the effect of a high-frequency change".

Spectral techniques have also been used recently by Sargent and Sims (1977), Geweke (1975, 1977), and Singleton (1980) to search for unobserved variables or factors, in a group of series, such as a common "business cycle factor" in a group of macro variables or a "national factor" in a group of regional employment series. The model is a dynamic generalization of the factor analysis model typically applied to cross-section data and postulates that all of the dynamic interrelationships between the series can be accounted for by a small number of

common factors. In the exploratory version of the model, which is useful for determining the number of common factors, standard estimation techniques adapted for complex arithmetic can be applied. Rather than applying these techniques to a covariance matrix, as in the cross-section case, they are applied to the spectral density matrix, frequency by frequency. When there are constraints on the model, as in confirmatory factor analysis, estimation is more difficult as constraints must be imposed across frequency bands. Often these constraints are more easily imposed in the time domain, and Engle and Watson (1981b) discuss time domain estimation and hypothesis testing methods.

8. Conclusion

Because of the way econometrics has been developing in recent years, the distinction between time series methods and the rest of econometrics has become much less clear. It seems very likely that this will continue and the tendency is already being reflected in modern textbooks such as Maddala (1977). It is nevertheless true that many econometricians do not appreciate the theoretical results and techniques available in the time series field, and so a list of some of the textbooks in this field is provided. The first four books concentrate on the frequency domain, and the others are general in coverage or deal just with the time domain (in each group, the books are approximately in order of increasing mathematical sophistication): Granger and Hatanaka (1964), Bloomfield (1976), Koopmans (1974), Priestly (1981), Granger (1980c), Nelson (1973), Box and Jenkins (1976), Granger and Newbold (1977), Fuller (1976), Anderson (1971), Brillinger (1975), and Hannan (1970).

References

Anderson, T. W. (1971) *The Statistical Analysis of Time Series*. New York: Wiley.

Ansley, C. (1979) "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process", *Biometrika*, 66, 59–65.

Ashley, R. A. and C. W. J. Granger (1979) "Time-Series Analysis of Residuals from the St. Louis Model", *Journal of Macroeconomics*, 1, 373–394.

Auerbach, A. J. (1981) "The Index of Leading Indicators: 'Measurement Without Theory', Twenty-Five Years Later", Harvard Institute of Economic Research, Discussion Paper 841.

Bates, J. W. and C. W. J. Granger (1969) "The Combination of Forecasts", *Operations Research Quarterly*, 20, 451–468.

Bloomfield, P. (1976) *Fourier Analysis of Time Series*. New York: Wiley.

Box, G. E. P. and G. M. Jenkins (1976) *Time Series, Forecasting and Control*, Holden Day, San Francisco, revised edition.

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

Cleveland, W. P. and G. C. Taio (1976) "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program", *Journal of the American Statistical Association*, 71, 581–587.

Cooper, R. L. (1972) "The Predictive Performance of Quarterly Econometric Models of the United States" in *Econometric Models of Cyclical Behavior*, ed. by B. G. Hickman, Columbia University Press.
Engle, R. F. (1976) "Band Spectrum Regression Inter". *International Economic Review*, 15, 1–11.
Engle, R. F. (1978) "Testing Price Equations for Stability Across Spectral Frequency Bands", *Econometrica*, 46, 869–882.
Engle, R. F. (1979) "Estimating Structural Models of Seasonality", in Zellner (1979).
Engle, R. F. (1980) "Estimates of the Variance of Inflation Based Upon the ARCH Model", University of California, San Diego, Discussion Paper.
Engle, R. F. and D. Kraft (1981) "Multiperiod Forecast Error Variances of Inflation Estimated from ARCH Models", paper presented to the Conference on Applied Time Series Analysis of Economic Data, Washington, D.C., October, 1981.
Engle, R. F., C. W. J. Granger, and D. Kraft (1981) "Combining Competing Forecasts of Inflation Using a Bivariate ARCH Model", University of California, San Diego, mimeo.
Engle, R. F. and M. W. Watson (1981a) "A One-Factor Model of Metropolitan Wage Rates", *Journal of the American Statistical Association*, 76 (December).
Engle, R. F. and M. W. Watson (1981b) "A Time Domain Approach to Dynamic MIMIC and Factor Models" (revised), Harvard University, mimeo.
Engle, R. F., D. F. Hendry, and J. F. Richard (1981) "Exogeneity", University of California, San Diego Discussion Paper 81-1 (revised).
Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of Inflationary Expectations", forthcoming in *Econometrica*.
Fuller, W. A. (1976) *Introduction to Statistical Time Series*. New York: Wiley.
Geweke, J. (1975) "Employment Turnover and Wage Dynamics in U.S. Manufacturing", Unpublished Ph.D. Dissertation, University of Minnesota.
Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series", in: D. J. Aigner and A. S. Goldberger, eds., *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam, ch. 19.
Godfrey, M. D. and H. Kareman (1976) "A Spectrum Analysis of Seasonal Adjustment", *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, ed. M. Shubik, Princeton University Press.
Granger, C. W. J. (1969) "Prediction with a Generalized Cost of Error Function", *Operations Research Quarterly*, 20, 199–207.
Granger, C. W. J. (1980a) "Long-Memory Relationships and The Aggregation of Dynamic Models", *Journal of Econometrics*.
Granger, C. W. J. (1980b) "Some Properties of Time-Series Data and Their Use in Econometric Model Specification", *Annals of Applied Econometrics* (supplement to *Journal of Econometrics*).
Granger, C. W. J. (1980c) *Forecasting in Business and Economics*. New York: Academic Press.
Granger, C. W. J. and A. P. Andersen (1979) *An Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoeck and Ruprecht.
Granger, C. W. J. and M. Hatanaka (1964) *Spectral Analysis of Economic Time Series*. Princeton University Press.
Granger, C. W. J. and P. Newbold (1974) "Spurious Regressions in Econometric", *Journal of Econometrics*, 26, 1045–1066.
Granger, C. W. J. and P. Newbold (1976) "Forecasting Transformed Series", *Journal of the Royal Statistical Society B*, 38, 189–203.
Granger, C. W. J. and P. Newbold (1977) *Forecasting Economic Time Series*. New York: Academic Press.
Granger, C. W. J. and O. Morgenstern (1970) *Predictability of Stock Market Prices*. Lexington: Heath-Lexington Book.
Granger, C. W. J. and R. Ramanathan (1981) "On the Combining of Forecasts", University of California, San Diego, mimeo.
Gray, Kelley and McIntire (1978) "A New Approach to ARMA Modeling", *Communications in Statistics, B*.
Hannan, E. J. (1969) "The Identification of Vector Mixed Autoregressive Moving Average Systems", *Biometrika*, 56, 223–225.
Hannan, E. J. (1970) *Multiple Time Series*. New York: Wiley.

Hannan, E. J. and B. G. Quinn (1979) "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society B*, 1, 190–195.
Harvey, A. C. and G. D. A. Phillips (1979) "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances", *Biometrika*, 66, 49–58.
Hatanaka, M. and E. P. Howrey (1969) "Low-Frequency Variations in Economic Time Series", *Kyklos*, 22, 752–766.
Havenner, A. and P. A. V. B. Swamy (1981) "A Random Coefficient Approach to Seasonal Adjustment", *Journal of Econometrics*.
Hosking, J. R. M. (1980) "Lagrange-Multiplier Tests of Time-Series Models", *Journal of the Royal Statistical Society*, *B*, 42, no. 2, 170–181.
Howrey, E. P. (1972) "Dynamic Properties of a Condensed Version of the Wharton Model", *Econometric Models of Cyclical Behavior*, Vol. II, ed. B. Hickman, Columbia University Press, 601–663.
Hymans, S. H. (1973) "On the Use of Leading Indicators to Predict Cyclical Turning Points", *Brookings Papers on Economic Activity*, no. 2.
Jorgenson, D. W. (1964) "Minimum Variance, Linear, Unbiased Seasonal Adjustment of Economic Time Series", *Journal of the American Statistical Association*, 59, 681–687.
Jorgenson, D. W. (1967) "Seasonal Adjustment of Data for Econometric Analysis", *Journal of the American Statistical Association*, 62, 137–140.
Joyeux, R. (1979) "Harmonizable Processes in Economics", Ph.D. Thesis, Department of Economics, University of California, San Diego.
Koopmans, L. H. (1974) *The Spectral Analysis of Time Series*. New York: Academic Press.
Kupier, J. (1979) "A Survey of Comparative Analysis of Various Methods of Seasonal Adjustment", in Zellner (1979).
Labys, W. C. and C. W. J. Granger (1970) *Speculation, Hedging and Forecasts of Commodity Prices*. Lexington: Heath-Lexington Books.
LeRoy, S. F. and R. D. Porter (1981) "The Present-Value Relation: Tests Based on Implied Variance Bounds", *Econometrica*, 49, 555–574.
Lucas, R. E., Jr. (1976) "Econometric Policy Evaluation: A Critique", *The Phillips Curve and the Labor Market* (K. Brunner and A. Meltzer, ed.) Vol. 1 of Carnegie-Rochester Conferences in Public Policy.
Maddala, G. S. (1977) *Econometrics*. New York: McGraw-Hill.
Neftci, S. N. (1979) "Lead-Lag Relationships and Prediction of Economic Time Series", *Econometrica*, 47, 101–114.
Nelson, C. R. (1972) "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy", *American Economic Review*, December.
Nelson, C. R. (1973) *Applied Time-Series Analysis for Managerial Forecasting*. San Francisco: Holden Day.
Nerlove, M. (1964) "Spectral Analysis of Seasonal Adjustment Procedures", *Econometric*, 32, 241–286.
Nerlove, M., D. M. Grether, and J. L. Carvalho (1979) *Analysis of Economic Time Series, A Synthesis*. New York: Academic Press.
Newbold, P. (1974) "The Exact Likelihood Functions for a Mixed Autoregressive-Moving Average Process", *Biometrika* 61, 423–426.
Newbold, P. (1981) "A Note on Modelling Seasonally Adjusted Data", *Journal of Time Series Analysis*.
Newbold, P. and C. Ansley (1979) "Small Sample Behavior of Some Precedures Used in Time Series Model Building and Forecasting", mimeo.
Pagan, A. R. (1975) "A Note on the Extraction of Components from Time Series", *Econometrica*, 43, 163–168.
Parzen, E. and M. Pagano (1978) "An Approach to Modelling Seasonally Stationary Time Series", *Journal of Econometrics*, 9 (*Annals of Applied Econometrics*, 1979-1), 137–154.
Pierce, D. A. (1977) "Relationships and the Lack Thereof Between Economic Time Series, with Special Reference to Money and Interest Rates", *Journal of the American Statistical Society*, 72, 11–21.
Pierce, D. A. (1979) "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present", in Zellner (1979).

Pierce, D. A. (1980) "Recent Developments in Seasonal Adjustment", *Proceedings of the IMS Special Time Series Meeting on Time Series*.
Plosser, C. I. and G. W. Schwert (1977) "Estimation of a Non-Invertible Moving Average Process: The Case of Over Differencing", *Journal of Econometrics*, 5, 199–224.
Priestly, M. B. (1980) "State-Dependent Models: A General Approach to Non-Linear Time Series Analysis", *Journal of Time Series Analysis*, 1, 45–71.
Priestly, M. B. (1981) *Spectral Analysis and Time Series*. New York: Academic Press.
Prothero, D. L. and K. F. Wallis (1976) "Modelling Macroeconomic Time Series", *Journal of the Royal Statistical Society A*, 139, 468–500.
Robbins, R. (1981), Unpublished Ph.D. Dissertation, University of California, San Diego.
Rosenberg, B. (1973) "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression", *Annals of Economic and Social Measurement*, 2, 399–428.
Sargent, T. J. (1981) "Interpreting Economic Time Series", *Journal of Political Economy*, 89, 213–248.
Sargent, T. J. and C. A. Sims (1977) "Business Cycle Modelling Without Pretending to Have Too Much A Priori Economic Theory", *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis.
Shiller, R. J. (1981), "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, June.
Shiskin, J., A. H. Young, and J. C. Musgrave (1967) "The X-11 Variant of the Census Method-II Seasonal Adjustment Program", Technical Paper No. 15, U.S. Bureau of the Census.
Singleton, K. J. (1980) "A Latent Time Series Model of the Cyclical Behavior of Interest Rates", *International Economic Review*, 21.
Singleton, K. J. (1980) "Expectations Models of the Term Structure and Implied Variance Bounds", *Journal of Political Economy*, 88.
Sims, C. A. (1974) "Seasonality in Regression", *Journal of the American Statistical Association*, 69, 618–626.
Sims, C. A. (1980) "Macroeconomics and Reality", *Econometric*, 48, 1–48.
Taylor, J. B. (1979) "Estimation and Control of a Macroeconomic Model with Rational Expectations", *Econometrica*, 47, 1267–1286.
Tiao, G. C., G. E. P. Box, G. B. Hudak, W. R. Bell, and I. Chang (1979) "An Introduction to Applied Multiple Time Series Analysis", University of Wisconsin, mimeo.
Wallis, K. F. (1974) "Seasonal Adjustment and Relations Between Variables", *Journal of the American Statistical Association*, 69, 18–31.
Wallis, K. F. (1979) "Seasonal Adjustment and Multiple Time Series", in Zellner (1979).
Watson, M. W. and R. F. Engle (1980) "Testing for Varying Regression Coefficients When a Parameter is Unidentified Under the Null", unpublished manuscript, University of California, San Diego, July, 1980.
Weiner, N. (1949) *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Boston: M.I.T. Press.
Whittle, P. (1963) *Prediction and Regulation*. London: English Universities Press.
Young, A. H. (1968) "Linear Approximations to the Census and BLS Seasonal Adjustment Methods", *Journal of the American Statistical Association*, 63, 445–457.
Zellner, A. (1979) "Seasonal Analysis of Economic Time Series", *U.S. Department of Commerce, Bureau of the Census, Economic Research Report*, ER-1.
Zellner, A. and F. Palm (1976) "Time-Series Analysis and Simultaneous Equation Econometric Models", *Journal of Econometrics*, 2, 17–54.