This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Seasonal Analysis of Economic Time Series

Volume Author/Editor: Arnold Zellner

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/zell78-1

Publication Date: 1978

Chapter Title: Analysis and Modeling of Seasonal Time Series

Chapter Author: George E. P. Box, Steven Hilimer, George C. Tiao

Chapter URL: http://www.nber.org/chapters/c4329

Chapter pages in book: (p. 309 - 344)

ANALYSIS AND MODELING OF SEASONAL TIME SERIES

George E. P. Box
Steven C. Hillmer
George C. Tiao
University of Wisconsin

INTRODUCTION

It is frequently desired to obtain certain derived quantities from economic time series. These include smoothed values, deseasonalized values, forecasts, trend estimates, and measurements of intervention effects.

The form that these derived quantities should take is clearly a statistical question. As with other statistical questions, attempts to find answers follow two main routes: An empirical approach and a model-based approach.

The first appeals directly to practical considerations, the latter to theory. We shall argue that best results are obtained by an iteration between the two and further show how this process has led to a useful class of time series models that may be used to study the questions mentioned above.

A SIMPLE EXAMPLE

To illustrate with an elementary example, consider the classical problem of selecting a measure of location M for a set of repeated measurements $Z_1, Z_2, \ldots, Z_n$. Arguing empirically: (1) If it were postulated that a change in any observation Z_j to $Z_j+\delta$ should have a linear effect on the location measure changing M to $M+c_j\delta$, this would imply that M was of the form

$$M=\sum_j w_j Z_j \qquad (1)$$

(2) Suppose it was further postulated that if the set of observations all happened to be equal to some value Z, then M should also be equal to Z. This would imply that

$$\sum_j w_j = 1 \qquad (2)$$

(3) Finally, if there were no reason to believe that any single observation supplied more information about M than any other, then the w_j's might be taken equal so that

$$w_j=n^{-1}, \; (j=1, 2, \ldots, n) \qquad (3)$$

Thus, from this empirical argument, M would be the arithmetic average, $\bar{Z}$, of the data.

From a theoretical or model-based viewpoint, the same quantity $\bar{Z}$ might be put forward if it were believed that the generation of the measurements was realistically simulated by random sampling from a Normal population. From this premise, well-known mathematical argument would lead to $\bar{Z}$.

Iteration Between Theory and Practice

The empirical method and the model-based method of attack are each employed by knowledgeable statisticians and are sometimes thought of as rivals. But they are, we believe, only rivals in the same sense that the two sexes are rivals. In both cases, isolation is necessarily sterile, while interaction can be fruitful.

The model-based approach works only if we can postulate a realistic class of models. But, from where are such models to come? One important source is from empirical procedures that have been found to behave satisfactorily in practice. Suppose, for a particular type of data, practical experience shows that the arithmetic average measures location well. Then, we can be led to the standard Normal model by asking what assumptions would make $\bar{Z}$ a good measure of location.

But, if $\bar{Z}$ can be arrived at empirically, where is the need for a model? The answer is, of course, that the existence of a model acts as a catalyst to further development. In particular, a model allows—

1. Constructive criticism of the original empirical idea. For instance, model-based arguments that recommend $\bar{Z}$ on Normal theory assumptions can also warn of consequences, not easily foreseen from an empirical standpoint, if the population is Cauchy-like or contains an outlier. It can also suggest better functions of the observations in these latter circumstances.
2. Generalization of the idea. For instance, in cases where the Normal theory assumptions are sensible for estimating a mean, they are equally sensible for more general models, leading in particular to the method of least squares estimation.

This research was supported, in part, by a grant from the Wisconsin Alumni Research Foundation and, in part, by the U.S. Army Research Office under grant No. DAAG29-76-G-0304.

History seems to show that most rapid progress occurs when theory and practice are allowed to confront, criticize, and stimulate each other. A brief sketch of some historical developments in time series analysis illustrates this point.

ITERATIVE DEVELOPMENT OF SOME IMPORTANT IDEAS IN TIME SERIES ANALYSIS

Visual inspection of economic time series, such as annual wheat prices, suggests the existence of cycles. When Beveridge [2] fitted empirical models containing linear aggregates of cosine waves plus Normally distributed independent errors, statistically significant cycles, indeed, appeared. However, these cycles had exotic and mysterious periods for which no direct cause could be found. This led Yule [35] to propose a revolutionary kind of model in which it was supposed that time series could be represented as the output from a dynamic system excited by random shocks. The dynamic system was represented by a difference equation, and the random shocks were represented by independent drawings from a Normal distribution.

For illustration, suppose the sequence $\{Z_t\}$ is a time series with mean μ and $\{a_t\}$ is a sequence of independent drawings from the distribution $N(O, \sigma^2)$, which we shall call white noise. Then, an important model proposed by Yule was the second-order autoregressive process

$$z_t=\phi_1 z_{t-1}+\phi_2 z_{t-2}+a_t, \text{ where } z_t=Z_t-\mu$$

With B the back-shift operator, such that $Bz_t=z_{t-1}$, this model may be written

$$(1-\phi_1 B-\phi_2 B^2)z_t=a_t \tag{4}$$

If in (4) the second-degree polynomial (with B treated temporarily as a variable) has complex zeros, then the impulse response of (4) can be oscillatory, and the generated series can exhibit pseudocyclic behavior, like that of the economic data.

More general dynamic characteristics may be obtained with a model of the form

$$\varphi(B)z_t=\theta(B)a_t \tag{5}$$

where the polynomial $\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_P B^P$ is called an autoregressive operator of order P and $\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ is called a moving average operator of order q. The resulting model (5) is called an autoregressive moving average process of order (P, q), or simply an ARMA (P, q) process.

In the decades that followed Yule's proposal, evidence was obtained for the practical usefulness of models of the form of equation (5), and much study was given to them.

Authors, notably Bartlett, Durbin, Hannan, Jenkins, Kendall, Quenouille and Wald, studied their properties including their autocorrelation functions, methods for their fitting, and tests for their adequacy.

Heavy emphasis had, up to this time, been placed on the stationarity of time series, and it was known that, for a stationary process, the zeros of $\varphi(B)$ in (5) had to lie outside the unit circle. However, in the 1950's, operations research workers, such as Holt [14; 23] and Winters [33], required methods for smoothing and forecasting nonstationary economic and business series. Their need sparked a return to empiricism, resulting in the development and generalization of exponential weighting for smoothing and forecasting.

Suppose it is desired to measure the location at current time t of a nonstationary economic time series $\{Z_t\}$. For this purpose, the first two postulates advanced before, resulting in equations (1) and (2), would seem appropriate,[1] so that

$$M_t=\sum_{j=0}^{\infty} w_j Z_{t-j}, \text{ with } \sum_{j=0}^{\infty} w_j=1$$

However, it is sensible to require that the current value Z_t should be given more weight than the penultimate value Z_{t-1}, which, in turn, should have more weight than Z_{t-2}, etc. If the weights are made to fall off exponentially, so that $w_{j+1}=\theta w_j$, where $0<\theta<1$ is a smoothing constant, then, since $\sum w_j=1$, it follows that $w_j=(1-\theta)\theta^j$. The measure M_t was called an exponentially weighted average and has the very convenient property that it can be updated by the formula $M_{t+1}=(1-\theta)Z_{t+1}+\theta M_t$.

The practical usefulness of this measure soon became apparent, and it was developed and generalized by Brown [11; 12] and by Brown and Meyer [13]. One important application was to employ M_t as the one step ahead (lead one) forecast $\hat{Z}_t(1)$ of Z_{t+1}, where the notation $\hat{Z}_t(l)$ refers to a forecast made l steps ahead from time origin t. Thus, in this application

$$\hat{Z}_t(1)=(1-\theta)\sum_{j=0}^{\infty} \theta^j Z_{t-j} \tag{6}$$

The practice-theory iteration proceeded through one more important step when Muth [27] asked what stochastic process would be such that (6) provided a forecast having minimal mean square error. He showed that the required stochastic process was of the form

$$(1-B)Z_t=(1-\theta B)a_t, \quad -1<\theta<1 \tag{7}$$

This is a nonstationary process of the form of (5) with $P=1$, $q=1$, and $\varphi_1=1$. Thus, a root of $\varphi(B)$ is on the unit circle.

The account of further developments follows the approach adopted by Box and Jenkins [3; 4; 5; 6; 7; 8; 9].

Empirical evidence from control engineering also pointed to the importance of nonstationary stochastic models of the form of (5), having roots on the unit circle.

[1] The expressions that follow imply an infinite series of past data. However, since in practice the weights quickly decay towards zero, M_t can be calculated to required accuracy from a fairly short series.

Long before the introduction of James Watt's governor, empirical methods of feedback control were being developed; Mayr [26]. The earliest forms used control in which the adjustment x_t at time t was made proportional to the deviation e_t from target T of some objective function $y_t = T + e_t$. The adjustment function, or controller, was, thus, of the form $x_t = -ke_t$. Adjustments with such controllers lagged behind when trends occurred in the disturbance, and it was soon realized that control could often be greatly improved by adding an integral term. Equivalently, for discrete control with observations and adjustments made at equally spaced times, a summation terms Se_t such that $Se_t = \sum_{i=0}^{\infty} e_{t-j}$ was added yielding a controller

$$x_t = -(k_1 e_t + k_2 Se_t) \tag{8}$$

Since such proportional plus integral controllers have been eminently successful and continue to be widely used throughout the process industries, it is natural to ask "What theoretical setup would make such a controller optimal?" Supposing, as would often be the case, that the dynamic relation between x_t and y_t could be roughly approximated by a first-order system, modeled by the difference equation $y_t = \delta y_{t-1} + \omega x_{t-1}$. It is easily shown that the control equation (8) would be optimal for a disturbance modeled by the stochastic process (7).

Thus, successful empiricism in two widely different fields point to the usefulness of models in which a first difference $\nabla Z_t = (1-B)Z_t$, or possibly a higher difference $\nabla^d Z_t = (1-B)^d Z_t$, could be represented by a stationary model. It, subsequently, turned out that such a class of models had, in fact, been proposed by Yaglom [34]. Such models are, thus, of the form

$$\phi(B)\nabla^d z_t = \theta(B) a_t \tag{9}$$

where $\varphi(B) = \phi(B)(1-B)^d$ and with $P = p + d$. In this model, $\phi(B)$ is a stationary autoregressive operator of degree p, having zeros outside the unit circle, $\nabla^d z_t$ is the d^{th} difference of the series, and $\theta(B)$ is of degree q and has zeros outside the unit circle. Such a model is called an autoregressive integrated moving average process or ARIMA of order (p, d, q). More generally, these developments point to the possible usefulness of models in which one or more of the zeros of $\varphi(B)$ in (5), although not necessarily unity, lie on the unit circle. In what follows, $z_t = Z_t - \mu$, where μ is the mean of the series if stationary and, otherwise, is any convenient origin.

Properties of the ARIMA Models

It now becomes appropriate to test whether these generalizations of models arising from successful empiricism are practical.

Two relevant questions are—

1. What kinds of dependence of a current value z_t on past history can be represented by the model?
2. What kinds of projection or forecast of a time series are possible with the model?

Dependence of z_t on Past Values

Any model of the form of (5) can be thought of as an autoregressive model of possibly infinite order. Thus,

$$z_{t+1} = \sum_{j=0}^{\infty} \pi_j z_{t-j} + a_{t+1} \tag{10}$$

In this model, the one-step-ahead forecast $\hat{z}_t(1) = \sum_{j=0}^{\infty} \pi_j z_{t-j}$ is a linear aggregate of previous observations. Using parlance popular in econometrics, the forecast is obtained by applying a distributed lag weight function[2] to previous observations. Alternatively, the weights π_j can be thought of as defining the memory of the past, contained in the current value z_t.

Now, many forms have been proposed by econometricians for distributed lag-weight functions. How general are those implied by (5)?

This equation may be written

$$\varphi(B)\theta^{-1}(B) z_t = a_t \tag{11}$$

which may be compared with (10), written as

$$\pi(B) z_t = a_t \tag{12}$$

where

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots$$

By equating coefficients in the identity $\varphi(B) = \theta(B)\pi(B)$, the π weights may be calculated correspondingly for any choice of the polynomials $\varphi(B)$ and $\theta(B)$. Also, the nature of these weights may be deduced, and it can be shown that using this form we can represent—

1. A finite set of π weights that are functionally unrelated when we have a pure autoregressive model in which $\theta(B) = 1$.
2. Of more interest, a convergent series of π weights that, after any desired number of initial, unrelated values, follows a function which can be any mixture of damped exponentials, sine and cosine waves satisfying the difference equation $\theta(B)\pi_j = 0$. Convergence is assured by the requirement that the zeros of $\theta(B)$ lie outside the unit circle.

Convergence of the weights seems essential for any sensible memory function. The class of functions included is sufficiently rich to be capable of representing a very wide variety of practical situations.

---

[2] Distributed lag models are usually credited to Irving Fisher [18] but were used earlier by R. A. Fisher [19].

Forms of Projection Implied by the Model

The model in the form (10) for z_{t+l} is

$$z_{t+l}=\sum_{j=0}^{\infty} \pi_j z_{t+1-j}+a_{t+l} \tag{13}$$

Multiplying both sides by $(1+\psi_1 B+\psi_2 B^2+\ldots+\psi_{l-1}B^{l-1})$ and choosing the $l-1$ coefficients $\psi_1, \psi_2, \ldots, \psi_{l-1}$ so that the coefficients of $z_{t+l-1}, z_{t+l-2}, \ldots, z_{t+1}$ are all zero, we can write (13) as

$$z_{t+l}=\sum_{j=0}^{\infty} \pi_j^{(l)} z_{t-j}+\sum_{i=0}^{l-1}\psi_i a_{t+l-i} \tag{14}$$

where $\psi_0=1$. In fact, the ψ's are obtained by equating coefficients in $\pi(B)\psi(B)=1$, where $\psi(B)$ is the infinite series $1+\psi_1 B+\psi_2 B^2+\ldots$.

Now, suppose that (14) is written in the form

$$z_{t+l}=\hat{z}_t(l)+e_t(l) \tag{15}$$

Then $\hat{z}_t(l)$ is the conditional expectation of z_{t+l} given all past history up to time t which we will also write as $[z_{t+l}]_t$. It may be shown (see, e.g., [32]) that this conditional expectation is the minimum mean square error forecast of z_{t+l}, made at origin t.

Consider the forecasts $\hat{z}_t(1), \hat{z}_t(2), \hat{z}_t(3), \ldots, \hat{z}_t(l), \ldots$, all made at origin t. The values together yield a forecast function $f(l, \underset{\sim}{b}^{(t)})$ that constitutes the appropriate projection of the series.

Specifically, (15) may be written

$$z_{t+l}=f(l,\underset{\sim}{b}^{(t)})+e_t(l) \tag{16}$$

where $f(l, \underset{\sim}{b}^{(t)})=\hat{z}_t(l)=[z_{t+l}]_t$ is some function of l, with $\underset{\sim}{b}^{(t)}$ a vector of coefficients that are fixed for any origin t but are updated when the origin is advanced.

For example, if the model (5) were of the form

$$(1-B)^2 z_t=(1-\theta_1 B-\theta_2 B^2)a_t \tag{17}$$

it is readily shown that the forecast function at origin t would be a straight line

$$\hat{z}_t(l)=f(l, \underset{\sim}{b}^{(t)})=b_0^{(t)}+b_1^{(t)}l \tag{18}$$

Any series that could be regarded as an outcome of the model (17) should, thus, be projected or forecasted along a straight line.

Writing the model for z_{t+l} in the form of (5) and taking conditional expectations

$$\varphi(B)[z_{t+l}]_t=\theta(B)[a_{t+l}]_t \tag{19}$$

where

$$[z_{t+j}]_t=\begin{cases}\hat{z}_t(j) \text{ if } j>0\\ z_{t+j} \text{ if } j\le 0\end{cases} \text{ and } [a_{t+j}]_t=\begin{cases}0 & \text{if } j>0\\ z_{t+j}-\hat{z}_{t+j-1}(1) & \text{if } j\le 0\end{cases}$$

so that for $l>q$,

$$\varphi(B)[z_{t+l}]_t=0 \tag{20}$$

It is, thus, easily seen that possible forecast functions are—

1. A set of unrelated values, followed by a fixed value equal to the mean when we have a pure moving average model with $\varphi(B)=1$.
2. Of more interest, any mixture of polynomials, damped exponentials and damped sine and cosine waves, possibly preceded by one or more unrelated values.

We see, therefore, that the forecast functions implied by the model are also sensible and of sufficient variety to satisfy many practical needs.

The relationship (16) has a form that is of interest quite apart from forecasting. The function $f(l, \underset{\sim}{b}^{(t)})$ supplies all the information about z_{t+l} that is available up to time t, while $e_t(l)$ represents information that enters the system at a later time. If, in (5), $P\le q$, then $e_t(l)$ is a particular integral, and $f(l, \underset{\sim}{b}^{(t)})$ is the complementary function of (5), i.e., it is the solution of

$$\varphi(B)f(l, \underset{\sim}{b}^{(t)})=0 \tag{21}$$

Adaptive Updating of Forecast Function

It may be shown, by taking conditional expectations, that the coefficients $\underset{\sim}{b}^{(t)}$ in the forecast function are automatically updated as the origin for the forecast is advanced. Thus, for the model $(1-B)^2 z_t = (1-\theta_1 B-\theta_2 B^2)a_t$, for which the forecast function is the straight line $\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}l$, the updating formulas[3] are

$$b_0^{(t+1)}-(b_0^{(t)}+b_1^{(t)})=\lambda_0 a_{t+1} \tag{22}$$

and

$$b_1^{(t+1)}-b_1^{(t)}=\lambda_1 a_{t+1} \tag{23}$$

where

$$\lambda_0=1+\theta_2;\ \lambda_1=1-\theta_1-\theta_2 \text{ and } a_{t+1}=z_{t+1}-\hat{z}_{t(1)} \tag{24}$$

Such formulas can be obtained for any stochastic model of the form of (5) and its associated forecast function.

From this, we reach the following conclusions: Time series models of the form of (5)—

1. Yield a sensible and rich class of memory functions, relating the dependence of present on the past.
2. Yield a sensible and rich class of forecast (complementary) functions for describing the future behavior of the series, which depends on the past.

[3] Obviously, $b_0^{(t)}+b_1^{(t)}$ would be the updated intercept if no adjustments in the coefficients were made. The constants λ_0 and λ_1, which are functions of the model memory parameters, determine the changes necessary in the coefficients in the light of the discrepancy $a_{t+1}=z_{t+1}-\hat{z}_t(1)$ between prediction and actuality.

3. Yield a sensible procedure for updating forecast functions.

This serves to show that the proposed models are not arbitrary. Indeed, if the desirable characteristics listed were set out as requirements for a model, it can be shown that we would inevitably be led to the form (5) for the generating stochastic process. There is, thus, a strong prima facie case in favor of this form of model.

Further properties, however, are needed for success, since—

1. It should be possible to build a model from available data. When an actual time series is under study, the appropriate form of the model within the general class and appropriate values for the memory parameters $\underset{\sim}{\theta}=(\theta_1, \ldots, \theta_q)'$ and $\varphi=(\varphi_1, \ldots, \varphi_p)'$ should be deducible from the time series itself. The deduced stochastic model will then automatically determine the form of the forecast and memory functions.
2. Evidence from many applications should show that the models do adequately represent real series, and application to problems, such as forecasting and intervention analysis, Box and Tiao [10], ought to have yielded satisfactory results with reasonable consistency.

Model Building

Models have been built using the following three-stage iterative procedure (identification-fitting-diagnostic checking):

1. Identification seeks a tentative model form, worthy to be entertained. In particular, it should suggest appropriate orders for the polynomials $\phi(B)$ and $\theta(B)$, indicate needed transformation of the series, and indicate appropriate orders of differencing. It is usually accomplished chiefly by graphical inspection of the time series and of computed auxiliary sample functions, such as the autocorrelation function, partial autocorrelation function, and, in some cases, the spectrum.
2. Fitting involves the estimation, by maximum likelihood, of the memory parameters $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}$ and, where necessary, the parameters of the transformation.
3. Diagnostic checking is intended to show up inadequacies in the model and to suggest remedies. It is usually accomplished by inspection of residuals and of their computed auxiliary functions. When inadequacies are found, a further iterative cycle is initiated.

Evidence From Applications

In recent years, applications of these methods have become increasingly common not only in economics and business but in widely diverse areas, such as environmental studies and educational psychology. Indeed, it is nearly impossible to keep up with the literature. Up to now, these applications have dealt with forecasting and with intervention analysis. This literature seems to show that models of this general class have usually worked successfully over a very wide field of subject matter.

Seasonality

Seasonality is a phenomenon commonly found in economic, environmental, and other time series. Models of the form (5) are, in principle, sufficiently general to represent such series, but, to allow representation in a most parsimonious fashion, special forms of (5) have been worked out and have proved useful.

Seasonal series are such that similarities occur at equivalent parts of a cycle. As an example, consider monthly data for department store sales that might have a seasonal pattern because of Christmas, Easter, holidays, summer vacations, etc. Now, sales for a particular month, e.g., December, might be related from year to year by a model like (2.2) in which B is replaced by B^{12}. Thus,

$$\Phi(B^{12})z_t=\Theta(B^{12})u_t \tag{25}$$

We may suppose that a similar model applies to the other months. However, the residual u_t's from such a model would be expected to be dependent from month to month. If they can be related by a model

$$\varphi(B)u_t=\theta(B)a_t \tag{26}$$

then, on substituting (26) in (25), we obtain the seasonal model

$$\Phi(B^{12})\varphi(B)z_t=\Theta(B^{12})\Theta(B)a_t \tag{27}$$

For illustration, a typical form of a seasonal nonstationary model that has been identified and used successfully in many economic series arises when the functional form (7) represents both monthly and yearly components. It is, thus,

$$(1-B)(1-B^{12})z_t=(1-\theta_1 B)(1-\theta_2 B^{12})a_t \tag{28}$$

Figure 1a shows the π weights for this model when it is written as an infinite autoregressive process of the form of (10)

$$z_{t+1}=\sum_{j=0}^{\infty}\pi_j z_{t-j}+a_{t+1} \tag{29}$$

To see the implications, note that if we use $z_t^{(\theta)}$ to mean an exponential average of the form of (6), then the model may also be written

$$z_{t+1}=z_t^{(\theta_1)}+\{z_{t-11}-z_{t-12}^{(\theta_1)}\}^{(\theta_2)}+a_{t+1} \tag{30}$$

Thus, the contribution to z_{t+1} of all previous data values

Figure 1a. π WEIGHTS FOR MODEL (2.25)

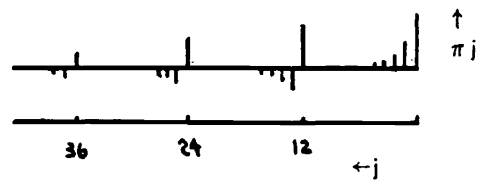


Figure 1b. FORECAST FUNCTION FOR MODEL (2.25)

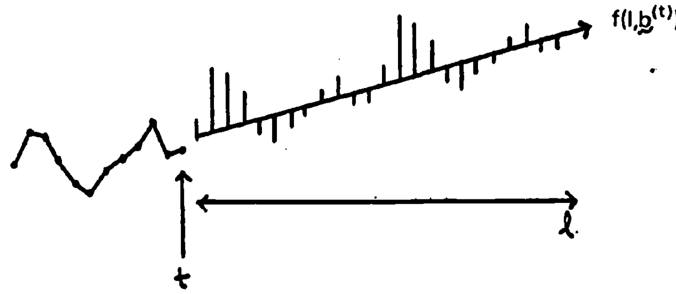
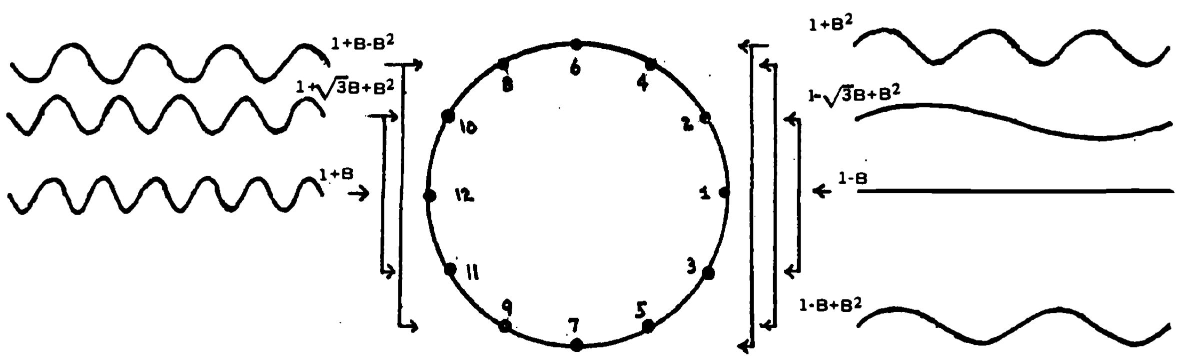


Figure 1c. SEASONAL COMPONENTS OF $(1-B^{12})$ ASSOCIATED WITH ROOTS ON THE UNIT CIRCLE



takes the appealingly sensible form of a crude forecast $z_t^{(\theta_1)}$, which is an exponential average taken over previous months, corrected by an exponential average of discrepancies between similar crude forecast and actuality for the same month, discounted over previous years.

More generally, any model of the form of (27) may be written

$$\frac{\varphi(B)}{\theta(B)}\frac{\Phi(B^{12})}{\Theta(B^{12})}z_t=a_t \qquad (31)$$

or

$$R(B)Q(B^{12})z_t=a_t$$

where

$R(B)=1-R_1B-R_2B^2-\ldots$ and $Q(B^{12})=1-Q_1B^{12}-Q_2B^{24}-\ldots$

Now write

$z_t^{(R)}$ for the weighted sum $R_1z_t+R_2z_{t-1}+\ldots$

and

$z_t^{(Q)}$ for the weighted sum $Q_1z_t+Q_2z_{t-12}+\ldots$

Then

$$z_{t+1}=z_t^{(R)}+\{z_{t-11}-z_{t-12}^{(R)}\}^{(Q)}+a_{t+1} \qquad (32)$$

which may be interpreted as before but with monthly and seasonal weights following a more general, and not necessarily, exponential pattern.

The particular model (28) written in the form of (16) becomes

$$z_{t+l}=b_0^{(t)}+b_1^{(t)}l+b_{0.m}^{(t)}+e_t(l) \qquad (33)$$

The typical appearance of the forecast function is sketched in figure 1b. In this expression, the forecast function consists of an updated straight line plus seasonal adjustment factors $b_{0.m}^{(t)}$, such that $\sum_{m=1}^{12} b_{0.m}^{(t)}\neq 0$. These factors are automatically adjusted as each new piece of data comes to hand and are weighted averages of past data.

Alternatively, (33) may be written

$$z_{t+l}=b^{(t)}+b_1^{(t)}l+\sum_{j=1}^{5}\left\{b_{1j}^{(t)}\cos\left(\frac{2j\pi}{12}\right)+b_{2j}^{(t)}\sin\left(\frac{2j\pi}{12}\right)\right\}+e_{t+l} \qquad (34)$$

In this form, the seasonal component contains a complete set of undamped sinusoids, adaptive in amplitude and phase with frequencies 0, 1, 2, ..., 6 cycles per year. These components are associated with the 12 roots of unity on the unit circle produced by the operator $(1-B^{12})$ and indicated in figure 1c. Thus, the complementary function is a solution of

$$(1-B)(1-B^{12})f(l,\underset{\sim}{b}^{(t)})=0$$

or equivalently of

$$(1-B^2)(1+B+B^2+\ldots+B^{11})f(l,\underset{\sim}{b}^{(t)})=0$$

or equivalently of

$$(1-B)^2(1-\sqrt{3B}+B^2)(1-B+B^2)(1+B^2)(1+B+B^2)(1+\sqrt{3B}+B^2)(1+B)f(l,\underset{\sim}{b}^{(t)})=0,\ l>13 \qquad (35)$$

More elaborate models produce appropriately more elaborate weight functions and forecast functions.

Summary

We have attempted to show that, as a result of the practice-theory iteration extended over many decades and carried on by many different investigators, a class of stochastic models capable of representing nonstationary and seasonal time series has evolved. When these models have been employed for forecasting and intervention analysis, they have worked well. There is no reason to believe that they would be any less useful for a model-based attack on problems, such as smoothing and seasonal adjustment. These problems are now considered.

SMOOTHING AND SEASONAL ADJUSTMENT

Like other problems, smoothing and seasonal adjustment can be tackled from either an empirical or a model-based point of view. Also, like other problems, it is fairly certain that an iteration between the two approaches in which each stimulates the other is likely to be most fruitful. Inspired empiricism, as we have seen, first produced exponential smoothing and its generalizations. It has also produced valuable methods for seasonal adjustment, exemplified, in particular, by the census X-11 procedure, Shiskin et al [29].

The additive version of the census procedure assumes that an observed time series $\{z_t\}$ can be written

$$z_t=S_t+p_t+e_t \qquad (36)$$

where S_t is the seasonal component, p_t is the trend component, and e_t is the noise component. Specific symmetric filters of the form

$$M(\delta,k)=\sum_{j=-k}^{k}\delta_j z_{t-j} \qquad (37)$$

with $\delta_j=\delta_{-j}$, are employed to produce estimates $\hat{S}_t$, $\hat{p}_t$ and $\hat{e}_t$ of these unobserved components. For the majority of economic data met in practice, the weights δ_j's used for computing $\hat{S}_t$ and $\hat{p}_t$ are shown in figure 3, given later in this section. The procedure has been widely used in Government and industry and found to produce sensible results.

The empirical success of this procedure motivated Cleveland and Tiao [16] to ask the question: Are there stochastic models for S_t, p_t and e_t for which the census procedure would be optimal? In general, if each of these components follows a model of the form (5), then, the minimum mean square error estimates of p_t and S_t are, respectively, the conditional expectations $E(p_t|z)$ and $E(S_t|z)$. They showed that if the components follow the models

$$(1-B)^2 p_t=(1+0.49B-0.49B^2)c_t$$

$$(1-B^{12})S_t=(1+0.64B^{12}+0.83B^{24})b_t \qquad (38)$$

and $\{c_t\}$, $\{b_t\}$, and $\{e_t\}$ are three independent Gaussian white-noise processes, such that $\sigma_b^2/\sigma_c^2=1.3$ and $\sigma_e^2/\sigma_b^2=14.4$, then the conditional expectations $E(p_t|z)$ and $E(S_t|z)$ will be of the symmetric form (37), with weights almost identical to the corresponding weights of $\hat{p}_t$ and $\hat{S}_t$ for the census procedure. This finding makes it possible to assess, at least partially, the appropriateness of the census procedure in a given instance. Specifically, expression (38) implies that the overall model for z_t is

$$(1-B)(1-B^{12})z_t=\theta(B)a_t \qquad (39)$$

where $\theta(B)$ is strictly a polynomial in B of degree 25, but the two largest coefficients are $\theta_1=0.34$ and $\theta_{12}=0.42$. This model is broadly similar to the one in (28), although the moving average structure is different. The implication is that the use of the census procedure for seasonal adjustment would be justified in situations where the series can be adequately represented by the model (39). On the other hand, if the model for a series $\{z_t\}$ were found to be vastly different from (39), then the appropriateness of the census decomposition would be in doubt.

The significance of this consideration is that it links the empirically successful census decomposition procedure to a model (39), for which this appropriateness can be verified for any given set of data. It should be borne in mind that, since only the series $\{z_t\}$ is available, any smoothing or seasonal adjustment procedure in the framework of (36) is necessarily arbitrary. On the other hand, whatever procedure one uses must at least be consistent with a model of z_t which can be built from the data. A procedure satisfying this minimum requirement will be called a model-based decomposition procedure.

We now consider what procedures would be produced using the stochastic model in (5). We will suppose that, by using past values of the series, a model has been carefully built in exactly the same way as for any other time series application, i.e., by an iterative sequence involving identification of a model worthy to be entertained, efficient fitting of the tentative model, and diagnostic checking of the fitted tentative model. Based on such a model for the observed series $\{z_t\}$, we shall then derive smoothing and seasonal adjustment procedures, illustrate how these derived procedures behave with actual data, and compare the results with other methods.

It is helpful for the development of ideas to first consider the simpler smoothing problem when there are no seasonal components and then to build onto this to develop seasonal adjustment methods.

Trend Plus Noise

We make the following assumptions, referred to as assumption I:

1. $\{T_t\}$ is an observed time series following the known model $\varphi(B)T_t=\eta(B)d_t$, with d_t being independent drawings from $N(O, \sigma_d^2)$.
2. $\varphi(B)$ is a polynomial in B of degree p with zeros on the outside of the unit circle, $\eta(B)$ is a polynomial of degree u with zeros outside the unit circle, and $\varphi(B)$ and $\eta(B)$ has no common zeros.
3. $T_t=p_t+e_t$, with $\{p_t\}$ and $\{e_t\}$ being independent of each other.
4. $\{p_t\}$ is a stochastic process following some ARIMA model.
5. $\{e_t\}$ is a Gaussian white-noise process, with mean O and variance σ_e^2.

An estimate of p_t is required from the time series $\{T_t\}$. When the models for p_t and e_t are given, the solution to this problem has been derived by Weiner [30], Kolmogorov [24], and Cleveland and Tiao [16]. In the analysis of economic time series, however, it is reasonable to assume that only the model for T_t is known. It is important, therefore, to consider how the models for p_t and e_t are restricted by this knowledge. On assumption I, the following results are readily proved (see, e.g., Cleveland [15]):

1. The autoregressive part of the model for p_t is the polynomial $\varphi(B)$.
2. The model for p_t is $\varphi(B)p_t=\alpha(B)c_t$, where $\alpha(B)$ is a polynomial in B of degree less than or equal to $\max(p, u)$, and c_t are independently distributed as $N(O, \sigma_c^2)$.
3. $\sigma_d^2\eta(B)\eta(F)=\sigma_c^2\alpha(B)\alpha(F)+\sigma_e^2\varphi(B)\varphi(F)$, $\qquad$ (40)

   where $F=B^{-1}$. Obviously, numerous combinations of σ_c^2, σ_e^2 and $\alpha(B)$ will satisfy equation (40).

In the remainder of this section, we sketch the results set out in more detail in Hillmer [22]. A model for p_t is called an acceptable model if, given the model for T_t—

1. $\alpha(B)$ satisfies equation (40) for some $\sigma_c^2\geq 0$ and $\sigma_e^2\geq 0$.
2. The zeros of $\alpha(B)$ lie on or outside the unit circle.

It is easy to obtain the following results:

1. Every given model for T_t has at least one acceptable model for p_t.
2. Given the model for T_t the possible values of σ_e^2 are bounded. We call this bound K^*.

3. Given the model for T_t, then every σ_e^2 in the range $0 \le \sigma_e^2 \le K^*$ determines a unique acceptable model for p_t.

Result (1) follows from letting $\sigma_e^2=0$ in which case $p_t=T_t$ with probability one. Result (2) follows from the fact that, for a model to be acceptable, we require $\sigma_c^2\alpha(B)\alpha(F)\ge 0$ for all $|B|=1$. Then, from equation (40), $\sigma_e^2 \le \frac{\sigma_d^2\eta(B)\eta(F)}{\varphi(B)\varphi(F)}$ for all $|B|=1$. For result (3), if $0\le \sigma_e^2\le K^*$, then $\sigma_d^2\eta(B)\eta(F)-\sigma_e^2(B)(F)=g(B)$ is nonnegative for B on the unit circle. Therefore, $g(B)$ determines a unique moving average polynomial $\alpha(B)$.

When $\sigma_e^2=0$, $T_t=p_t$, and there is no smoothing. When $\sigma_e^2=K^*$, on the other hand, the variance of the added white noise is maximized.

An illustrative example—For illustration, consider the rate of change in the consumer price index. Box and Tiao [10] have studied this time series and found that the model

$$(1-B)T_t=(1-0.84B)d_t,\ \sigma_d^2=0.0019$$

adequately describes its behavior from 1953 to 1971. If we assume that this time series follows assumption I, then the model for p_t must be of the form

$$(1-B)p_t=(1-\alpha B)c_t$$

Figure 2 shows the original and smoothed processes for various values of σ_e^2 and α. Also shown are the weight functions, ω_j, implicitly employed when the smoothed value is written in the form

$$\hat{p}_t=\sum_{-\infty}^{\infty}\omega_j T_{t-j}$$

The functions illustrate the case in which the time series available is $T_1, T_2, \ldots T_n$, and t is not close to an end value. The details of how the smoothing was carried out for this example are given later in this section.

While it seems natural in the practical circumstance when σ_e^2 is unknown to choose the variance of the added noise as large as possible, it is also to be noted, as in this example, that a wide range of models for p_t correspond to approximately the same σ_e^2. Furthermore, for this example, models for p_t with α in the range $-1\le\alpha\le 0$ will imply almost identical smoothed estimates.

Smoothing with maximum σ_e^2—We proceed then on the basis that the smoothing to be used should maximize σ_e^2. It is readily shown that—

1. The bound for σ_e^2, K^*, is attainable.
2. The bound K^* occurs when the moving average polynomial of the model for p_t has a zero on the unit circle.

To see (1),

$$K^*=\min_{|B|=1}\left\{\frac{\sigma_d^2\eta(B)\eta(F)}{\varphi(B)\varphi(F)}\right\} \qquad (41)$$

which can be calculated given the model for T_t. Now for (2), if we let $g(B)=\sigma_d^2\eta(B)\eta(F)-K^*\varphi(B)\varphi(F)$, it is then easily seen that $g(B)$ is a covariance generating function that uniquely determines a $\sigma_{c_*}^2$ and $\alpha_*(B)$ such that

$$\sigma_d^2\eta(B)\eta(F)=\sigma_{c_*}^2\alpha_*(B)\alpha_*(F)+K^*\varphi(B)\varphi(F)$$

Furthermore, $\alpha_*(B)$ is uniquely determined such that it has at least one zero on the unit circle and the remainder of its zeros on or outside the unit circle.

Deviation of the smoothing weights and calculation of end values—Therefore, given any model for T_t, we can calculate the maximum σ_e^2 that is consistent with the model for T_t. We shall see that knowing σ_e^2 is sufficient to be able to carry out the smoothing. Cleveland and Tiao [10] have established the following results:

1. When t is not close to the beginning or end of a time series, the smoothed estimate, $\hat{p}_t$, is a symmetric moving average of T_t where the weights, ω_j, are given by the coefficients of B in the generating function

$$\omega(B)=\frac{\sigma_c^2}{\sigma_d^2}\,\frac{\alpha(B)\alpha(F)}{\eta(B)\eta(F)}$$

However, from equation (40), it follows that

$$\omega(B)=1-\frac{\sigma_e^2}{\sigma_d^2}\,\frac{\varphi(B)\varphi(F)}{\eta(B)\eta(F)} \qquad (42)$$

so that knowledge of the model for T_t together with σ_e^2 will enable us to perform the smoothing. A method for computing the weights ω_j from (42) is described in the appendix.

2. For values of $\hat{p}_t$ near the end of the observed time series the following modification is appropriate. Conditional expectations (minimum mean square error forecasts) given the observed series, of a sufficient number of future values are first obtained. The method to obtain the forecasts is given in Box and Jenkins [3]. Then, the forecasted values are used as observations in the formula appropriate for the middle of the series. Consequently, we only need one set of weights. Precisely similar procedures are used for t near the beginning of the series.

Examples—The next two examples will help clarify the ideas that have been presented.

First, suppose T_t follows the model

$$(1-B)T_t=(1-\eta B)d_t$$

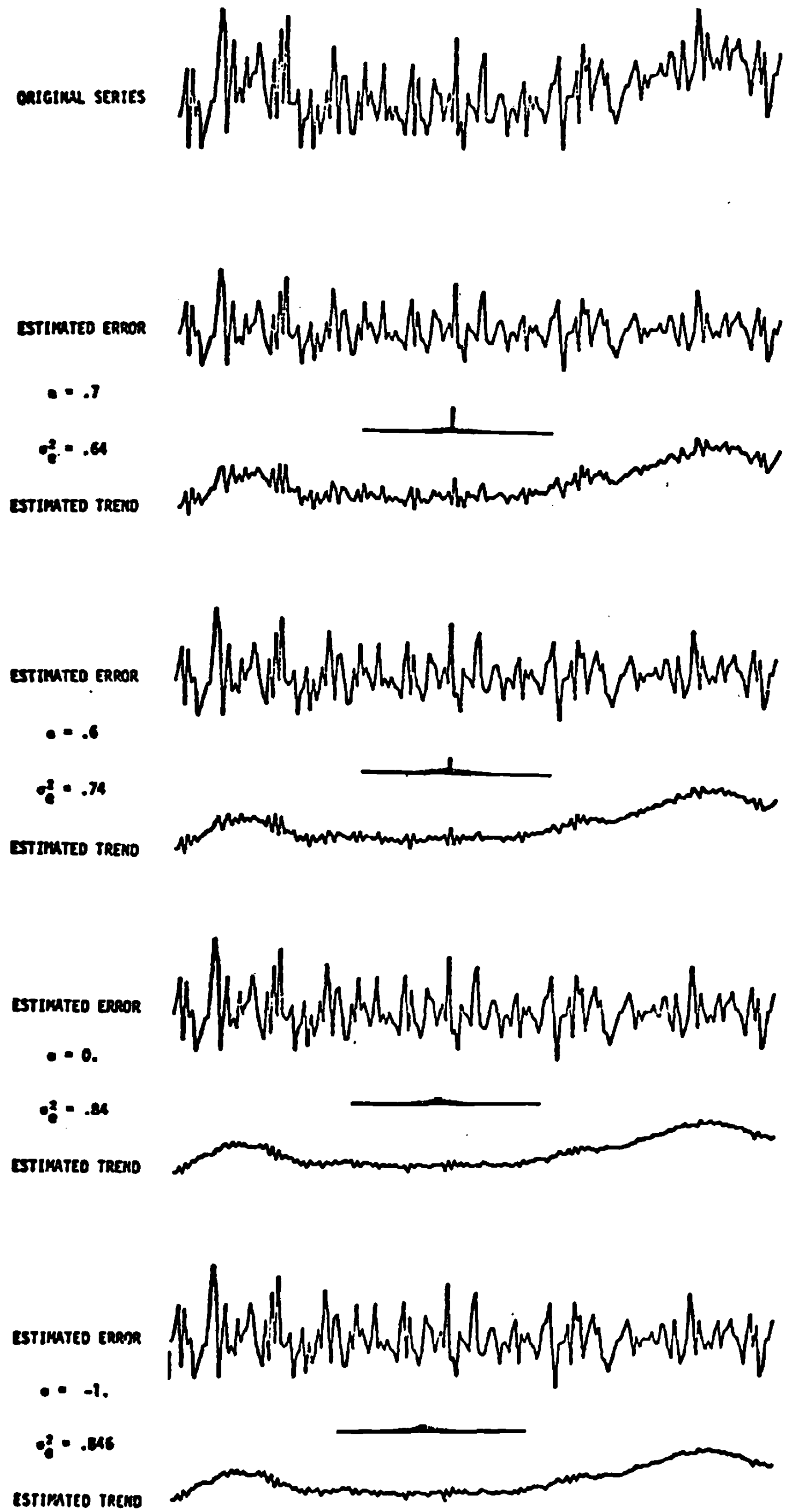
Under assumption I, the model for p_t must be of the form

$$(1-B)p_t=(1-\alpha B)c_t$$

for some α, and

$$\sigma_d^2(1-\eta B)(1-\eta F)=\sigma_c^2(1-\alpha B)(1-\alpha F)+\sigma_e^2(1-B)(1-F) \qquad (43)$$

Figure 2. **SOME POSSIBLE DECOMPOSITIONS INTO TRENDS AND ERROR, CONSUMER PRICE INDEX EXAMPLE**



By setting $B=1/\alpha$ in equation (43) and solving for σ_e^2, we obtain

$$\sigma_e^2=\frac{\sigma_d^2(\alpha-\eta)(1-\alpha\eta)}{(1-\alpha)^2}$$

It is easy to see that the maximum possible σ_e^2 occurs when $\alpha=-1$. This agrees with the result that the maximum σ_e^2 is attained when the zero of $(1-\alpha B)$ is on the unit circle. Consequently, the model for p_t, corresponding to the largest possible σ_e^2, is

$$(1-B)p_t=(1+B)c_t$$

and

$$\sigma_e^2=\frac{(1+\eta)^2}{4}\sigma_d^2$$

From (42), the appropriate smoothing weights are then given by the generating function

$$\omega(B)=1-\frac{(1+\eta)^2}{4}\frac{(1-B)(1-F)}{(1-\eta B)(1-\eta F)} \quad (44)$$

Using the method given in the appendix, we find $\omega_0=\frac{1-\eta}{2}$, $\omega_1=\omega_{-1}=\frac{1-\eta^2}{4}$, and $\omega_j=\omega_{-j}=\eta\omega_{j-1}$ for $j=2, 3, \ldots$. Consequently, for estimating p_t in the middle of the observed T_t series, the smoothed estimate is

$$\hat{p}_t=\frac{1-\eta}{2}T_t+\frac{1-\eta^2}{4}\{(T_{t+1}+T_{t-1})+\eta(T_{t+2}+T_{t-2})+\ldots\} \quad (45)$$

Now, to smooth the observed series at the end, one can proceed to first calculate the conditional expectations of $T_{n+1}, T_{n+2}, \ldots,$ given $T_n, T_{n-1}, \ldots$. These forecasts are: $\hat{T}_{n+1}=(1-\eta)T_n+\eta(1-\eta)T_{n-1}+\eta^2(1-\eta)T_{n-2}+\ \ldots$ and $\hat{T}_{n+j}=\hat{T}_{n+1}$ for $j=2, 3, \ldots$. Then,

$$\hat{p}_n=\frac{1-\eta}{2}T_n+\frac{1-\eta^2}{4}\{(\hat{T}_{n+1}+T_{n-1})+\eta(\hat{T}_{n+2}+T_{n-2})+\ldots\}$$

or

$$\hat{p}_n=\frac{(3-2\eta-\eta^2)}{4}T_n+\frac{1-\eta^2}{4}\{(1-\eta)T_{n-1}+\eta(1-\eta)T_{n-2}+\ldots\} \quad (46)$$

Smoothed values for other observations near the end of the observed series can be obtained in a similar manner.

Second, suppose the model for T_t is

$$(1-\phi_1B-\phi_2B^2)T_t=(1-\eta_1B-\eta_2B^2)d_t$$

Then, the model for p_t must be of the form

$$(1-\phi_1B-\phi_2B^2)p_t=(1-\alpha_1B-\alpha_2B^2)c_t$$

for some unknown $\alpha_1, \alpha_2, \sigma_c^2$ and σ_e^2. From (41), the maximum possible σ_e^2 consistent with the given model for T_t is

$$K^*=\sigma_d^2\min_{|B|=1}\frac{(1-\eta_1B-\eta_2B^2)(1-\eta_1F-\eta_2F^2)}{(1-\phi_1B-\phi_2B^2)(1-\phi_1F-\phi_2F^2)}$$

or, by letting $B=e^{i\omega}$ for $-0\le\omega\le\pi$, we have that $K^*=\sigma_d^2\min_{0\le\omega\le\pi}\{f(\omega)\}$, where

$$f(\omega)=\frac{(1-\eta_1e^{i\omega}-\eta_2e^{2i\omega})(1-\eta_1e^{-i\omega}-\eta_2e^{-2i\omega})}{(1-\phi_1e^{i\omega}-\phi_2e^{2i\omega})(1-\phi_1e^{-i\omega}-\phi_2e^{-2i\omega})}$$

$$=\frac{1+\eta_1^2+\eta_2^2-(\eta_1-\eta_1\eta_2)2\cos(\omega)-2\eta_2\cos(2\omega)}{1+\phi_1^2+\phi_2^2-(\phi_1-\phi_1\phi_2)2\cos(\omega)-2\phi_2\cos(2\omega)}$$

It is straightforward to numerically minimize $f(\omega)$, given specific values of η_1, η_2, ϕ_1 and ϕ_2. Once this is done, K^* can be calculated, and the appropriate smoothing weights can be calculated from the equation

$$\omega(B)=1-\frac{K^*}{\sigma_d^2}\frac{(1-\phi_1B-\phi_2B^2)(1-\phi_1F-\phi_2F^2)}{(1-\eta_1B-\eta_2B^2)(1-\eta_1F-\eta_2F^2)} \quad (47)$$

Seasonal Adjustment

For a nonseasonal T_t, the smoothing method described may be used directly to estimate the trend component p_t in $T_t=p_t+e_t$. Basically, the idea is to choose $\hat{p}_t$ to extract as much white noise from the observed time series as possible.

It is now supposed that the model contains a seasonal component S_t so that, as in (36),

$$z_t=S_t+T_t=S_t+p_t+e_t \quad (48)$$

The problem is to estimate p_t in the presence of the seasonal component S_t and the noise component e_t. To do this, we must ask what the properties of the seasonal component should be. The concept of a seasonal component and, hence, its definition are, to some extent, arbitrary, but it seems reasonable that it satisfies the following conditions:

1. It should be capable of evolving over time.
2. It should be such that for monthly data the sum of twelve consecutive components varies about zero with minimum variance. The minimum variance requirement in (2) arises because variation greater than minimal variation in the twelve monthly sums should properly be reflected in the trend component p_t or the noise component e_t.

To illustrate how the requirements for a seasonal component can be incorporated into a model-based decomposition procedure, we make a preliminary study in this section by considering the particular stochastic model (28)

$$(1-B)(1-B^{12})z_t=(1-\theta_1 B)(1-\theta_2 B^{12})a_t$$

which has been found to provide an adequate representation of many seasonal time series.

We proceed by making the following assumptions:

1. An observed time series is well represented by (28), with θ_1, θ_2 and σ_a^2 known.
2. S_t and T_t in (48) are independent and follow models of the ARIMA class. Then, necessarily (Cleveland [15]), the product of their autoregressive operators is $(1-B)(1-B^{12})$.
3. $T_t=p_t+e_t$ follows assumption I.

As we have seen, the complementary function for this model satisfies

$$(1-B)(1-B^{12})\hat{z}_t(l)=(1-B)^2(1+B+\ldots+B^{11})\hat{z}_t(l)=0$$

or

$$\hat{z}_t(l)=b_0^{(t)}+b_1^{(t)}l+b_{0,m}^{(t)} \text{ with } \sum_{m=1}^{12} b_{0,m}^{(t)}=0 \tag{49}$$

The adaptive trend term, $b_0^{(t)}+b_1^{(t)}l$, satisfies

$$(1-B)^2(b_0^{(t)}+b_1^{(t)}l)=0$$

and the seasonal component satisfies

$$(1+B+\ldots+B^{11})b_{0,m}^{(t)}=0$$

Thus, an appropriate model for the seasonal component is

$$(1+B+\ldots+B^{11})S_t=(1-\psi_1 B-\ldots-\psi_{11}B^{11})b_t \tag{50}$$

and the corresponding model for the trend component is

$$(1-B)^2T_t=(1-\eta_1 B-\eta_2 B^2)d_t$$

where $\{b_t\}$ and $\{d_t\}$ are two independent Gaussian white-noise processes with zero means and variances σ_b^2 and σ_d^2, respectively.

Letting $\psi(B)=(1-\psi_1 B-\ldots-\psi_{11}B^{11})$, $\eta(B)=(1-\eta_1 B-\eta_2 B^2)$, $\theta(B)=(1-\theta_1 B)(1-\theta_2 B^{12})$, and $U(B)=(1+B+\ldots+B^{11})$, we now observe that

$$(1-B)(1-B^{12})z_t=(1-B)(1-B^{12})S_t+(1-B)(1-B^{12})T_t$$

It follows that

$$\theta(B)a_t=(1-B)^2\psi(B)b_t+U(B)\eta(B)d_t$$

and

$$\sigma_a^2\theta(B)\theta(F)=\sigma_b^2(1-B)^2\psi(B)(1-F)^2\psi(F) + \sigma_d^2U(B)\eta(B)U(F)\eta(F) \tag{51}$$

Minimum variance solution—We require a solution such that $\bar{S}_t=U(B)S_t$ has the smallest variance. Let $\sigma_b^2\geq 0$, $\sigma_d^2\geq 0$, $\eta(B)$ and $\psi(B)$ be an acceptable solution, i.e. one which satisfies equation (3.16). Consider the polynomial

$$g(B)=\sigma_b^2\psi(B)\psi(F)-\epsilon^*U(B)U(F)$$

where

$$\epsilon^*=\min_{|B|=1}\left|\frac{\sigma_b^2\psi(B)\psi(F)}{U(B)U(F)}\right| \tag{52}$$

It follows that $g(B)$ is nonnegative for $|B|=1$ and hence there exists a unique $\sigma_{b_*}^2>0$ and $\psi^*(B)$ such that $\sigma_{b_*}^2\psi^*(B)\psi^*(F)=g(B)$. Now, (51) can be written as

$$\sigma_a^2\theta(B)\theta(F)=\sigma_{b_*}^2(1-B)^2(1-F)^2\psi^*(B)\psi^*(F)+U(B)U(F)H(B)$$

where

$$H(B)=\sigma_d^2\eta(B)\eta(F)+\epsilon^*(1-B)^2(1-F)^2$$

Clearly, $H(B)>0$ for $|B|=1$ so that we can determine a unique $\sigma_{d_*}^2>0$ and $\eta_*(B)$ such that

$$\sigma_{d_*}^2\eta_*(B)\eta_*(F)=H(B) \tag{53}$$

Thus,

$$\sigma_a^2\theta(B)\theta(F)=\sigma_{b_*}^2(1-B)^2(1-F)^2\psi_*(B)\psi_*(F) +\sigma_{d_*}^2\, U(B]U(F)\eta_*(B)\eta_*(F) \tag{54}$$

It is shown in Hillmer [22] that the seasonal and trend component model, corresponding to (54), has the desired property that the sum $\bar{S}_t$ has the smallest variance and, further, that the solution is unique.

Thus, to obtain the minimum variance solution, we must first find an acceptable model. As we shall see, for estimating the seasonal component S_t and the trend T_t, it is only necessary to know $\eta_*(B)$. From (51), we see that an acceptable solution is one for which $\sigma_d^2>0$ and

$$\sigma_a^2\theta(B)\theta(F)-\sigma_d^2U(B)U(F)\eta(B)\eta(F) \tag{55}$$

is nonnegative for $|B|=1$ and having four zeros at $B=1$. For the given $\sigma_a^2\theta(B)\theta(F)$, we can then employ numerical methods to obtain an σ_d^2 and an $\eta(B)=1-\eta_1 B-\eta_2 B^2$ having this property. To find the desired $\eta_*(B)$, note that (52) can be alternatively written as

$$\epsilon_*=\min_{|B|=1}\left\{\frac{\sigma_a^2\theta(B)\theta(F)U^{-1}(B)U^{-1}(F)-\sigma_d^2\eta(B)\eta(F)}{(1-B)^2(1-F)^2}\right\} \tag{56}$$

which can calculated using numerical methods. Once ϵ_* is obtained, the required $\eta_*(B)$ can be determined from (53).

Determination of the weight functions for the seasonal and trend components—It remains to obtain the smoothing weights when the models for S_t and T_t are determined. Let the estimated seasonal component be $\hat{S}_t$ and the estimated trend plus noise be $\hat{T}_t$. To estimate a component in the middle of the time series, we have, from Cleveland and Tiao [10], that

$$\hat{S}_t=\sum_{j=-\infty}^{\infty} w_j z_{t-j}=w(B)z_t$$

and

$$\hat{T}_t=\sum_{j=-\infty}^{\infty} h_j z_{t-j}=h(B)z_t \tag{57}$$

The generating function for the seasonal component is

$$w(B)=\frac{\sigma_b^2}{\sigma_a^2}\frac{(1-B)^2\psi(B)(1-F)^2\psi(F)}{\theta(B)\theta(F)} \tag{58}$$

However, since $z_t=\hat{S}_t+\hat{T}_t$, we have that $1=w(B)+h(B)$, so that

$$h(B)=\frac{\sigma_n^2}{\sigma_a^2}\frac{U(B)\eta(B)U(F)\eta(F)}{\theta(B)\theta(F)} \tag{59}$$

We do not need to know $\frac{\sigma_n^2}{\sigma_a^2}$, since this ratio can be obtained from the fact that $h(1)=1$. We suggest calculating $h(B)$, which can then be used to calculate $\hat{T}_t$, then $\hat{S}_t=z_t-\hat{T}_t$.

The smoothed values near the ends of the series are obtained as before, by forecasting the required unobserved values of z_t and using these forecasts in the formula for the center of the series.

The estimated trend plus noise component T_t can now be used to compute the estimate of the trend p_t. Noting that $T_t=p_t+e_t$ and the model for T_t is $(1-B)\,T_t=\eta(B)d_t$, it is readily verified from the results in the subsection on trend plus noise that

$$\hat{p}_t=\left[1-\frac{\sigma_e^2(1-B)^2(1-F)^2}{\sigma_d^2\eta(B)\eta(F)}\right]\hat{T}_t \tag{60}$$

The smoothest estimate of p_t is obtained when σ_e^2 is maximized, i.e., when

$$\sigma_e^2=\min_{|B|=1}\left\{\frac{\sigma_d^2\eta(B)\eta(F)}{(1-B)^2(1-F)^2}\right\} \tag{61}$$

Application to the Times Series of Monthly U.S. Unemployed Males, 20 Years Old and Over

These ideas will be illustrated by applying them to the time series of monthly U.S. unemployed males, 20 years old and over. This series is the largest component of the total unemployed series. The political and economic impact of this series has been especially important recently, and we consider it an important series in which to try out our methods.

By following the model building procedure, sketched in the subsection on model building, we find that the model

$$(1-B)(1-B^{12})z_t=(1-0.75B^{12})a_t,\ \sigma_a^2=0.0037 \tag{62}$$

adequately describes the behavior of the log of the observed series. Therefore, we can apply the seasonal adjustment and smoothing procedures previously outlined to this data. We find—

1. An acceptable solution results when $\eta_1=0.9251$, $\eta_2=0.05$ and $\sigma_d^2=0.700\sigma_a^2$.
2. $\min_{|B|=1}\left\{\frac{\sigma_a^2\theta(B)\theta(F)U^{-1}(B)U^{-1}(F)-\sigma_d^2\eta(B)\eta(F)}{(1-B)^2(1-F)^2}\right\}$ $=0.0376\sigma_a^2$.
3. $\eta^*(B)=(1-0.9798B+0.0034B^2)$, $\sigma_{a_*}^2=0.7780\sigma_a^2$.
4. $\min_{|B|=1}\left\{\frac{\sigma_{a_*}^2\eta_*(B)\eta_*(F)}{(1-B)^2(1-F)^2}\right\}=0.1915\sigma_a^2$

The seasonal and trend weights obtained by our procedure and by the census procedure are plotted in figure 3. It is interesting that the seasonal weights of both procedures, although arrived at quite differently, behave in a similar manner. Notice, however, that the trend weights for the two procedures are different. It should be remembered that our weight functions will adjust appropriately depending on the parameters of the particular series. The census weight function, however, is fixed for all series. The observed time series and the three estimated components derived from our procedure are plotted in figure 4a; the analogous series for the census procedure are plotted in figure 4b. Observe that the estimated seasonal components appear similar to each other. The estimated trend component from the census procedure is smoother than the trend component from our procedure, but this may reflect oversmoothing.

Generalization to other models—We have considered seasonal adjustment when the observed time series follows the model (28). Examination of the complementary function (49) led us to models for the seasonal and trend components that seemed sensible. Because the class of ARIMA models is more general than the model (28), it is doubtful whether all seasonal models in the ARIMA class could be treated in a precisely similar manner. If this is so, it will be because of the too limited nature of the seasonal adjustment concept; this concept may need to be widened. Every model in the ARIMA class will have a corresponding complementary function. Appropriate adjustment, we feel, may turn on suitable factorization of this function. The problem is an important one for future consideration.

Figure 3. TREND AND SEASONAL WEIGHTS FOR THE BUREAU OF THE CENSUS AND MODEL-BASED PROCEDURES

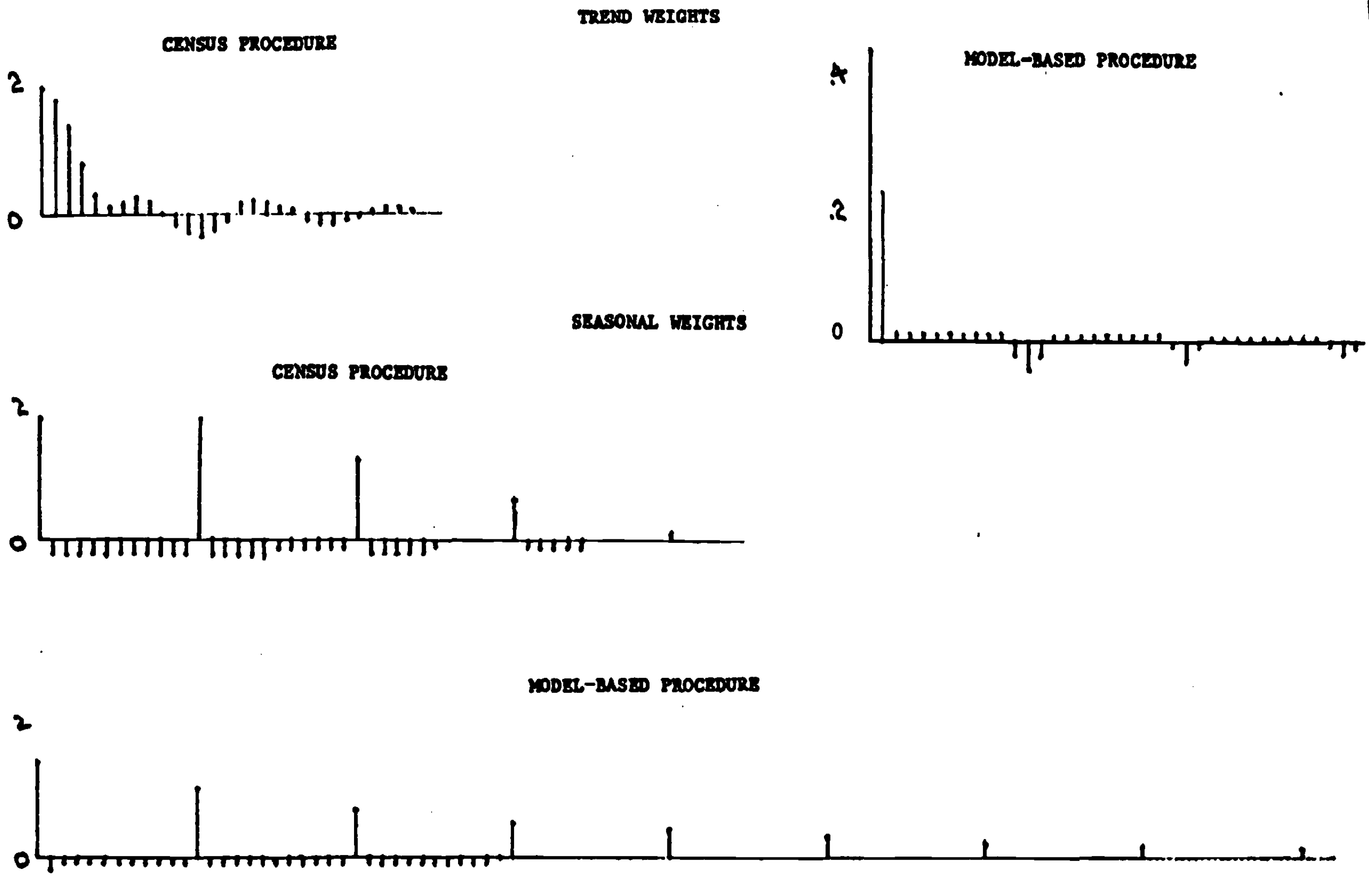
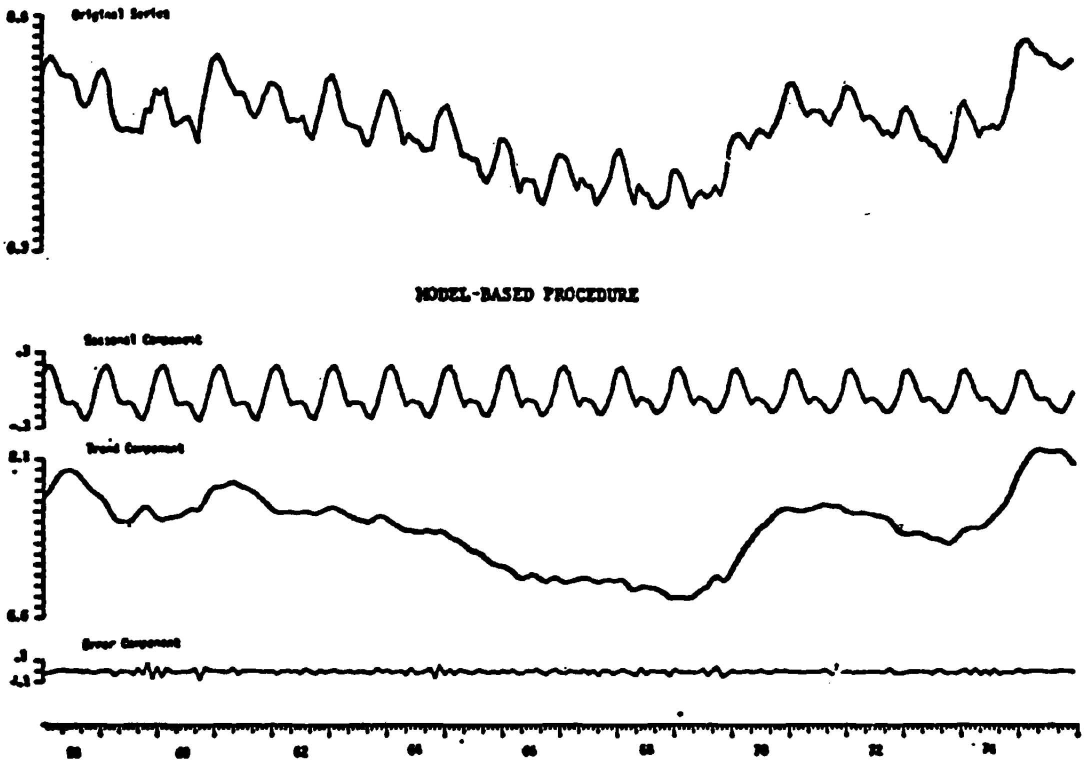
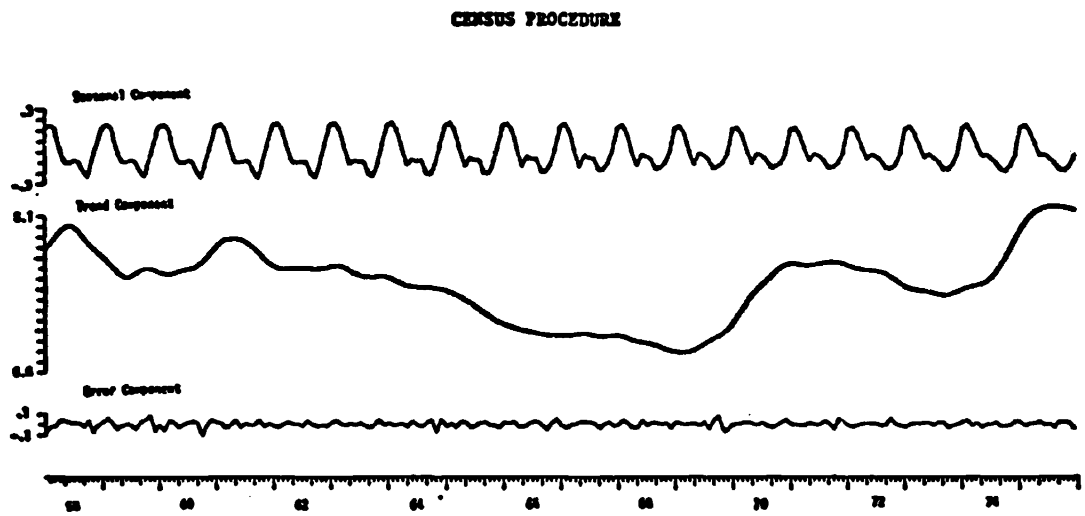


Figure 4a. THE ESTIMATED COMPONENTS FROM THE MODEL-BASED PROCEDURES FOR THE TIME SERIES OF UNEMPLOYED MALES, 20 YEARS OLD AND OVER



(a)

Figure 4b. THE ESTIMATED COMPONENTS FROM THE CENSUS PROCEDURE FOR THE TIME SERIES OF UNEMPLOYED MALES, 20 YEARS OLD AND OVER



(b)

MULTIVARIATE GENERALIZATION OF THE MODELS

When k series $\{z_{1t}\}$, $\{z_{2t}\}$, ..., $\{z_{kt}\}$ are considered simultaneously, it is necessary to allow for dynamic relationships that may exist between the series, for possibilities of feedback and for correlations between the shocks affecting the series. A useful class of models is obtained by direct generalization of (5) to

$$\Phi(B)\underset{\sim}{z}_t=\Theta(B)\underset{\sim}{a}_t \tag{63}$$

In this model,

$$\underset{\sim}{z}_t'=(z_{1t},\ldots,z_{kt}),\; \underset{\sim}{a}_t'=(a_{1t},\ldots,a_{kt})$$

$$\Phi(B)=I-\Phi_1B-\Phi_2B^2-\ldots-\Phi_pB^p$$

$$\Theta(B)=I-\Theta_1B-\Theta_2B^2-\ldots-\Theta_qB^q$$

and the Φ_i's and Θ_j's are $k\times k$ matrices of autoregressive and moving average parameters. $\{\underset{\sim}{a}_t\}$ is a sequence of vector-valued independent random shocks, distributed as multivariate normal $N(\underset{\sim}{O},\Sigma)$ that allows for contemporaneous correlation between the elements $a_{1t},\ldots,a_{kt}$.

While multiple time series models are naturally more complicated to handle and experience in their use is more limited, they provide a potential means of improving on results from univariate models. For example, information about z_{1t}, in addition to that contained in its own past, may be available from other related series $\{z_{2t}\}$, $\{z_{3t}\}$, etc. When this is so, improved forecasts, smoothed values, seasonal adjustments, etc., should be possible.

One difficulty that has previously impeded progress has been the estimation of parameters contained in the models. Initially, therefore, our attack has been directed at this problem, and, in the next section, we describe a practical means of computing exact maximum likelihood estimates.

Estimation

Whittle [31] and Hannan and Dunsmuir [21] have shown, for multivariate ARIMA models with normally distributed errors, that maximum likelihood estimates have desirable asymptotic properties. In particular, the estimates are asymptotically consistent and efficient. In addition, a number of authors have preformed simulations which indicate that maximum likelihood estimates have desirable small sample properties. Several examples are reported in Hillmer [22]. We will proceed to find maximum likelihood estimates of the parameters, assuming that the $\underset{\sim}{a}_t$'s are normally distributed.

Most time series estimation procedures that have been proposed to date are motivated by first considering the likelihood function and, then, by making some simplifying approximations to this function. (See, e.g., Anderson [1], Box and Jenkins [9], and Hannan [20].) While most of the simplifying approximations to the likelihood function have no effect upon the asymptotic estimates, these approximations can have an effect upon the estimates in small or even moderately large samples. In particular, the approximations can have a significant effect when estimating seasonal multiplicative ARIMA models. The likelihood function is usually simplified by ignoring the effect of the ends of the time series and by ignoring the changes that occur in the normalizing determinant. For univariate time series, Box and Jenkins [9] overcame the first problem by proposing an exact method for computing the exponent in the likelihood.

Recent papers, have indicated that further worthwhile improvement can be made in the estimation procedure if the terms in the determinant are not ignored. In the case of univariate time series several procedures have been recently proposed to obtain exact maximum likelihood estimates, see in particular, Ljung [25] and Dent [17].

The problem of estimating the parameters in a multivariate time series model is more difficult than the univariate problem. However, Hillmer [22] has developed procedures that give maximum likelihood estimates for multivariate ARIMA models. In order to illustrate those ideas, we give the details for a first-order moving average model.

First Order Moving Average Model

The multivariate MA(1) model is

$$\underset{\sim}{z}_t=\underset{\sim}{a}_t-\theta\underset{\sim}{a}_{t-1},\; t=1,\ldots n \tag{64}$$

where $\underset{\sim}{z}_t$ is a vector of k-observed time series, $\underset{\sim}{a}_t$ is a vector of unobserved errors following a multivariate normal distribution with mean vector $\underset{\sim}{0}$ and unknown covariance matrix Σ, and θ is a $k\times k$ matrix of parameters. Our objective will be to estimate θ and Σ.

First we derive the likelihood function of θ and Σ given the observations $\underset{\sim}{z}_1,\ldots,\underset{\sim}{z}_n$. Consider the transformation

$$\begin{aligned}\underset{\sim}{a}_0&=\underset{\sim}{a}_0\\ \underset{\sim}{z}_1&=\underset{\sim}{a}_1-\theta\underset{\sim}{a}_0\\ &\;\vdots\\ \underset{\sim}{z}_n&=\underset{\sim}{a}_n-\theta\underset{\sim}{a}_{n-1}\end{aligned}\quad\text{or}\quad \begin{bmatrix}\underset{\sim}{a}_0\\ \underset{\sim}{z}\end{bmatrix}=\begin{bmatrix}I_k & 0\\ D & C\end{bmatrix}\begin{bmatrix}\underset{\sim}{a}_0\\ \underset{\sim}{a}\end{bmatrix} \tag{65}$$

where $\underset{\sim}{z}=(\underset{\sim}{z}_1',\ldots,\underset{\sim}{z}_n')'$ $\underset{\sim}{a}=(\underset{\sim}{a}_1',\ldots,\underset{\sim}{a}_n')'$,

$$D=\begin{bmatrix}-\theta\\ 0\\ \cdot\\ \cdot\\ \cdot\\ 0\end{bmatrix}\quad\text{and } C=\begin{bmatrix}I_k & 0 & \cdots & 0\\ -\theta & I_k & \cdot & \cdot\\ 0 & -\theta & I_k & \cdot\\ \vdots & & \ddots & \vdots\\ \cdot & \cdot & & 0\\ 0\cdots & 0 & -\theta & I_k\end{bmatrix}$$

Equation (65) implies that

$$\begin{bmatrix}\underset{\sim}{a}_0\\ \underset{\sim}{a}\end{bmatrix}=\begin{bmatrix}I_k & 0\\ -C^{-1}D & C^{-1}\end{bmatrix}\begin{bmatrix}\underset{\sim}{a}_0\\ \underset{\sim}{z}\end{bmatrix}$$

$$=\begin{bmatrix}I_k\\ -C^{-1}D\end{bmatrix}\underset{\sim}{a}_0+\begin{bmatrix}0\\ C^{-1}\end{bmatrix}\underset{\sim}{z} \tag{66}$$

Because the $\underset{\sim}{a}_t$ are independently distributed as $N_k(\underset{\sim}{0}, \Sigma)$, the joint probability density function of $\underset{\sim}{a}_0$ and $\underset{\sim}{a}$ is

$$p(\underset{\sim}{a}_0, \underset{\sim}{a}) = \left(\frac{1}{2\pi}\right)^{\frac{k(n+1)}{2}} |\Sigma|^{-\frac{n+1}{2}}$$

$$\exp\left\{-\frac{1}{2}[\underset{\sim}{a}_0', \underset{\sim}{a}'][I_{n+1}\otimes\Sigma^{-1}]\begin{bmatrix}\underset{\sim}{a}_0\\ \underset{\sim}{a}\end{bmatrix}\right\} \quad (67)$$

where $\otimes$ denotes the Kronecker product of two matrices. Noting that the Jacobian of the transformation (66) is unity, expression (67) implies that the joint probability density function of $\underset{\sim}{a}_0$ and $\underset{\sim}{z}$ is

$$p(\underset{\sim}{a}_0, \underset{\sim}{z}) = \left(\frac{1}{2\pi}\right)^{\frac{k(n+1)}{2}} |\Sigma|^{-\frac{n+1}{2}} \exp\left\{-\frac{1}{2}S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma)\right\} \quad (68)$$

where

$$S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma) = \left\{\begin{bmatrix} I\\ -C^{-1}D\end{bmatrix}\underset{\sim}{a}_0 + \begin{bmatrix}0\\ C^{-1}\end{bmatrix}\underset{\sim}{z}\right\}'$$

$$\{I_{n+1}\otimes\Sigma^{-1}\}\left\{\begin{bmatrix} I\\ -C^{-1}D\end{bmatrix}\underset{\sim}{a}_0 + \begin{bmatrix}0\\ C^{-1}\end{bmatrix}\underset{\sim}{z}\right\} \quad (69)$$

By completing the square for $\underset{\sim}{a}_0$ in the quadratic form $S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma)$, we obtain

$$S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma) = S(\hat{\underset{\sim}{a}}_0, \underset{\sim}{z}|\theta, \Sigma) + (\underset{\sim}{a}_0 - \hat{\underset{\sim}{a}}_0)'A_n(\hat{\underset{\sim}{a}}_0 - \underset{\sim}{a}_0) \quad (70)$$

where

$$A_n = [I_k, -D'C'^{-1}][I_{n+1}\otimes\Sigma^{-1}]\begin{bmatrix} I_k\\ -C^{-1}D\end{bmatrix} \quad (71)$$

and

$$\underset{\sim}{a}_0 = -A_n^{-1}[I_k, -D'C'^{-1}][I_{n+1}\otimes\Sigma^{-1}]\begin{bmatrix}0\\ C^{-1}\end{bmatrix}\underset{\sim}{z} \quad (72)$$

Consequently, we have that

$$p(\underset{\sim}{a}_0, \underset{\sim}{z}) = \left(\frac{1}{2\pi}\right)^{\frac{k(n+1)}{2}} |\Sigma|^{-\frac{n+1}{2}}$$

$$\exp\left\{-\frac{1}{2}S(\hat{\underset{\sim}{a}}_0, \underset{\sim}{z}|\theta, \Sigma)\right\}\exp\left\{-\frac{1}{2}(\underset{\sim}{a}_0 - \hat{\underset{\sim}{a}}_0)'A_n(\underset{\sim}{a}_0 - \hat{\underset{\sim}{a}}_0)\right\}$$

By integrating out $\underset{\sim}{a}_0$ in $p(\underset{\sim}{a}_0, \underset{\sim}{z})$, we obtain the probability density function of $\underset{\sim}{z}$. By treating this probability density function as a function of θ and Σ, we obtain the likelihood function of θ and Σ given $\underset{\sim}{z}$ as

$$L(\theta, \Sigma|\underset{\sim}{z}) = |\Sigma|^{-\frac{n+1}{2}}\cdot|A^n|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}S(\hat{\underset{\sim}{a}}_0, \underset{\sim}{z}|\theta, \Sigma)\right\} \quad (73)$$

Observe, from equation (69), that

$$S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma) = \sum_{t=0}^{n}\underset{\sim}{a}_t'\Sigma^{-1}\underset{\sim}{a}_t$$

Furthermore, given θ, $\underset{\sim}{a}_0$, $\underset{\sim}{z}_t$ for $t=1, \ldots n$, we can calculate the $\underset{\sim}{a}_t$'s from the recursive relationship

$$\underset{\sim}{a}_t = \underset{\sim}{z}_t + \theta\underset{\sim}{a}_{t-1}$$

which is defined by the transformation in equation (65). Similarly, we can calculate $S(\underset{\sim}{a}_0, \underset{\sim}{z}|\theta, \Sigma)$, given θ, $\underset{\sim}{z}$ and $\underset{\sim}{a}_0$, by calculating

$$\hat{\underset{\sim}{a}}_t = \underset{\sim}{z}_t + \theta\hat{\underset{\sim}{a}}_{t-1} \text{ for } t = 1, \ldots n$$

and then

$$S(\hat{\underset{\sim}{a}}_0\underset{\sim}{z}|\theta, \Sigma) = \sum_{t=0}^{n}\hat{\underset{\sim}{a}}_t'\Sigma^{-1}\hat{\underset{\sim}{a}}_t \quad (74)$$

The important point to observe from this discussion is that the likelihood function can be evaluated for any given θ and Σ very quickly once A_n and $\hat{\underset{\sim}{a}}_0$ have been calculated. Consequently, we focus our attention upon the calculation of A_n and $\underset{\sim}{a}_0$. Note that

$$C^{-1} = \begin{bmatrix} I_k & & & & \\ & & & & \underset{\sim}{0} \\ \theta & I_k & & & \\ . & \theta & . & & \\ & & . & . & \\ . & & & . & . \\ \theta^{n-1} & \theta^{n-2} & \ldots & \theta & I_k \end{bmatrix}$$

then, from equation (71), we obtain

$$A_n = \Sigma^{-1} + \theta'\Sigma^{-1}\theta + \ldots + \theta'^n\Sigma^{-1}\theta^n \quad (75)$$

One easy way to perform the calculation of A_n is by the recursive calculation

$$A_0 = \Sigma^{-1} \text{ and } A_j = \Sigma^{-1} + \theta'A_{j-1}\theta \text{ for } j = 1, \ldots n \quad (76)$$

Calculation of A_n can be quickly performed on a computer with a minimum of storage. Next, consider the calculation of $\hat{\underset{\sim}{a}}_0$. From (72),

$$\hat{\underset{\sim}{a}}_0 = A_n^{-1}D'C'^{-1}\{I_n\otimes\Sigma^{-1}\}C^{-1}\underset{\sim}{z}$$

By multiplying out the matrices, we have that

$$D'C'^{-1}\{I_n\otimes\Sigma^{-1}\}C^{-1}\underset{\sim}{z} = \theta'A_{n-1}\underset{\sim}{z}_1 + \theta'^2A_{n-2}\underset{\sim}{z}_2 + \ldots + \theta'^nA_0\underset{\sim}{z}_n$$

where the A_j matrices are defined by equation (76). The vector $D'C'^{-1}\{I_n\otimes\Sigma^{-1}\}C^{-1}\underset{\sim}{z}$ can be calculated recursively as follows:

let $\underset{\sim}{g}_0 = \underset{\sim}{0}$ and $A_0 = \Sigma^{-1}$; let $g_j = \theta'\{g_{j-1} + A_{j-1}\underset{\sim}{z}_{n-j+1}\}$ for $j = 1, \ldots n$

then

$$g_n=\theta' A_{n-1}z_1+\theta'^2 A_{n-2}z_2+\ldots+\theta'^{n-1}A_0 z_n$$

To summarize, we have the following algorithm for the calculation of A_n and a_0:

1. Let $A_0=\Sigma^{-1}$ and $g_0=0$.
2. For $j=1, \ldots n$, let $g_j=\theta'\{g_{j-1}+A_{j-1}z_{n-j+1}\}$ and let $A_j=\Sigma^{-1}+\theta' A_{j-1}\theta$.
3. $\hat{a}_0=-A_n^{-1}g_n$.

For any given θ and Σ, we now have a way to calculate A_n and $\hat{a}_0$. Therefore, we can use these values to calculate the exact likelihood function.

We desire to maximize $L(\theta, \Sigma|z)$ with respect to the parameter matrices θ and Σ. We have illustrated a way to evaluate $L(\theta, \Sigma|z)$ for any particular values of θ and Σ. What is needed is some maximization algorithm that will systematically search over the parameter values and find the θ and Σ that will maximize $L(\theta, \Sigma|z)$. There are numerous optimization algorithms that can be used. We have used a nonlinear regression algorithm to estimate the parameters in the multivariate ARIMA examples given in the next section.

Application: The Durable Shipments, Durable New Orders, and Durable Inventories Series

We shall use the three series—

1. Shipments of durable goods
2. New orders of durable goods
3. Inventories of durable goods

to illustrate an analysis of seasonal multivariate models. A useful identification step is to build univariate models for the three series; they are:

1. $(1-B)(1-B^{12})z_{1t}=(1-B)(1-0.75B^{12})a_{1t}$

   $\sigma^2_{a_1}=0.00093$ with z_{1t}=log-durable shipments.

2. $(1-B)(1-B^{12})z_{2t}=(1-0.26B)(1-0.80B^{12})a_{2t}$ (77)

   $\sigma^2_{a_2}=0.00179$ with z_{2t}=log-durable new orders.

3. $(1-0.85B)(1-B^{12})z_{3t}=(1-0.36B)(1-0.73B^{12})a_{3t}$

   $\sigma^2_{a_3}=0.0000268$ with $z_{3t}=(1-B)$ log-durable inventories.

From consideration of the univariate models we tentatively entertained the multivariate seasonal model

$$(I-\Phi_1)(I-\Phi_{12}B^{12})z_t=(I-\theta_1 B)(I-\theta_{12}B^{12})a_t \tag{78}$$

where

$z_t'=(z_{1t}, z_{2t}, z_{3t})$, $a_t'=(a_{1t}, a_{2t}, a_{3t})$ and Φ_1, Φ_{12}, θ_1 and θ_{12} are 3 x 3 matrices. The parameters in this model are estimated as follows:

$$\hat{\Phi}_1=\begin{bmatrix} 0.36 & 0.53 & 0.75 \\ (0.12) & (0.09) & (0.79) \\ -0.29 & 1.30 & -2.00 \\ (0.15) & (0.11) & (0.83) \\ -0.05 & 0.04 & 0.79 \\ (0.01) & (0.01) & (0.07) \end{bmatrix}$$

$$\hat{\Phi}_{12}=\begin{bmatrix} 0.93 & 0.03 & 0.03 \\ (0.06) & (0.06) & (0.21) \\ 0.20 & 0.75 & 0.20 \\ (0.07) & (0.08) & (0.27) \\ 0.01 & -0.01 & 0.93 \\ (0.01) & (0.01) & (0.03) \end{bmatrix}$$

$$\hat{\theta}_1=\begin{bmatrix} -0.22 & 0.31 & -3.22 \\ (0.16) & (0.11) & (0.90) \\ -0.25 & 0.53 & -3.62 \\ (0.24) & (0.16) & (1.07) \\ -0.05 & 0.02 & 0.50 \\ (0.02) & (0.03) & (0.10) \end{bmatrix},$$

$$\hat{\theta}_{12}=\begin{bmatrix} 0.54 & 0.08 & 0.24 \\ (0.11) & (0.08) & (0.43) \\ -0.02 & 0.57 & 0.58 \\ (0.11) & (0.12) & (0.65) \\ 0.04 & -0.01 & 0.76 \\ (0.02) & (0.01) & (0.06) \end{bmatrix}$$

$$\hat{\Sigma}=\begin{bmatrix} 0.000548 & 0.000596 & -0.000010 \\ & 0.001464 & -0.000014 \\ & & 0.000021 \end{bmatrix} \tag{79}$$

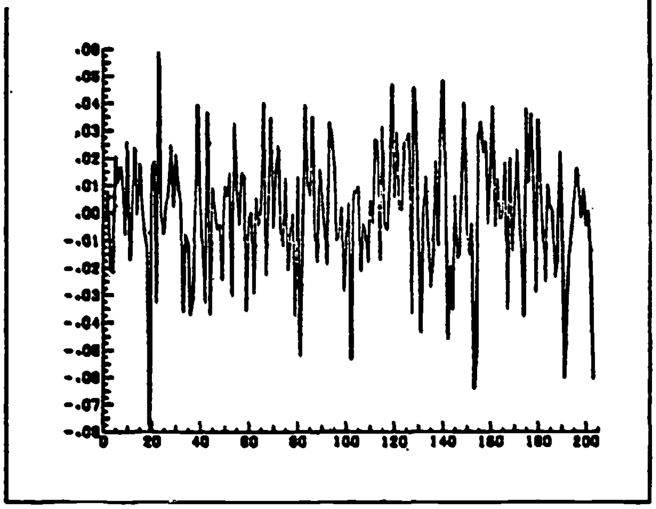
The numbers in the parentheses are the estimated standard errors of the parameter estimates. The residuals from this model are plotted in figure 5, and the autocorrelation and cross-correlation functions of the residuals are plotted in figure 6. Examination of these plots does not suggest any inadequacies in the model.

Inspection of the fitted model reveals a structure that seems to be readily capable of interpretation. This is made clearer if we simplify the model by setting to zero coefficients that are small, compared with their standard errors, and make other minor adjustments.

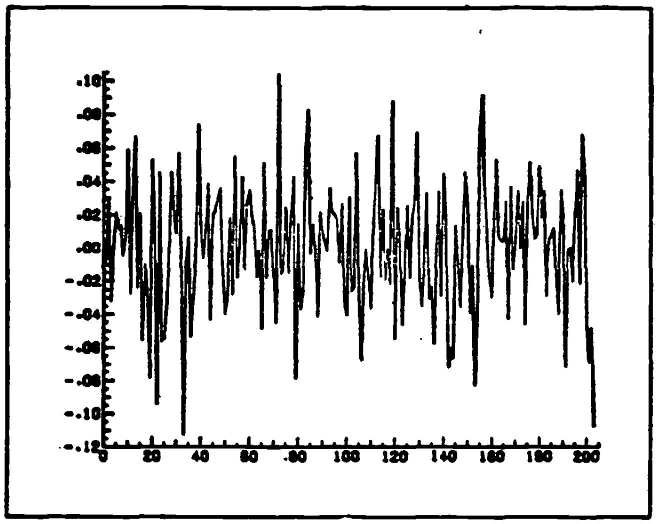
A valuable feature of the multivariate maximum likelihood program is that it permits any chosen coefficient to

Figure 5. **RESIDUALS FROM THE MULTIVARIATE FIT IN THE DURABLES EXAMPLE**

Residuals
Log Shipments



Residuals
Log New Orders



Residuals
(1-B)Log Inventories

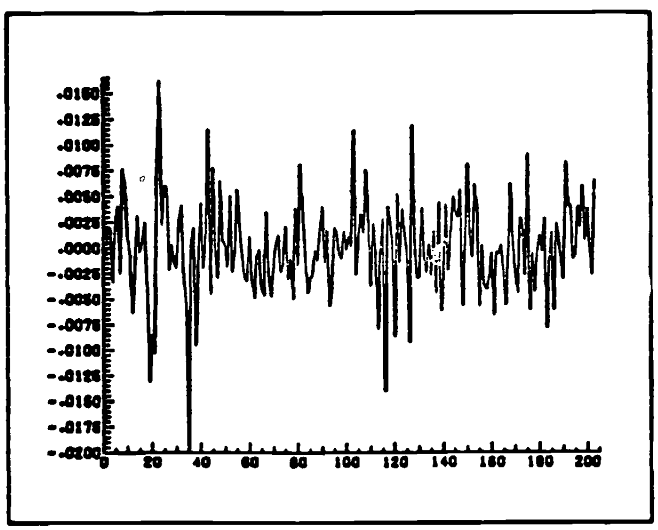
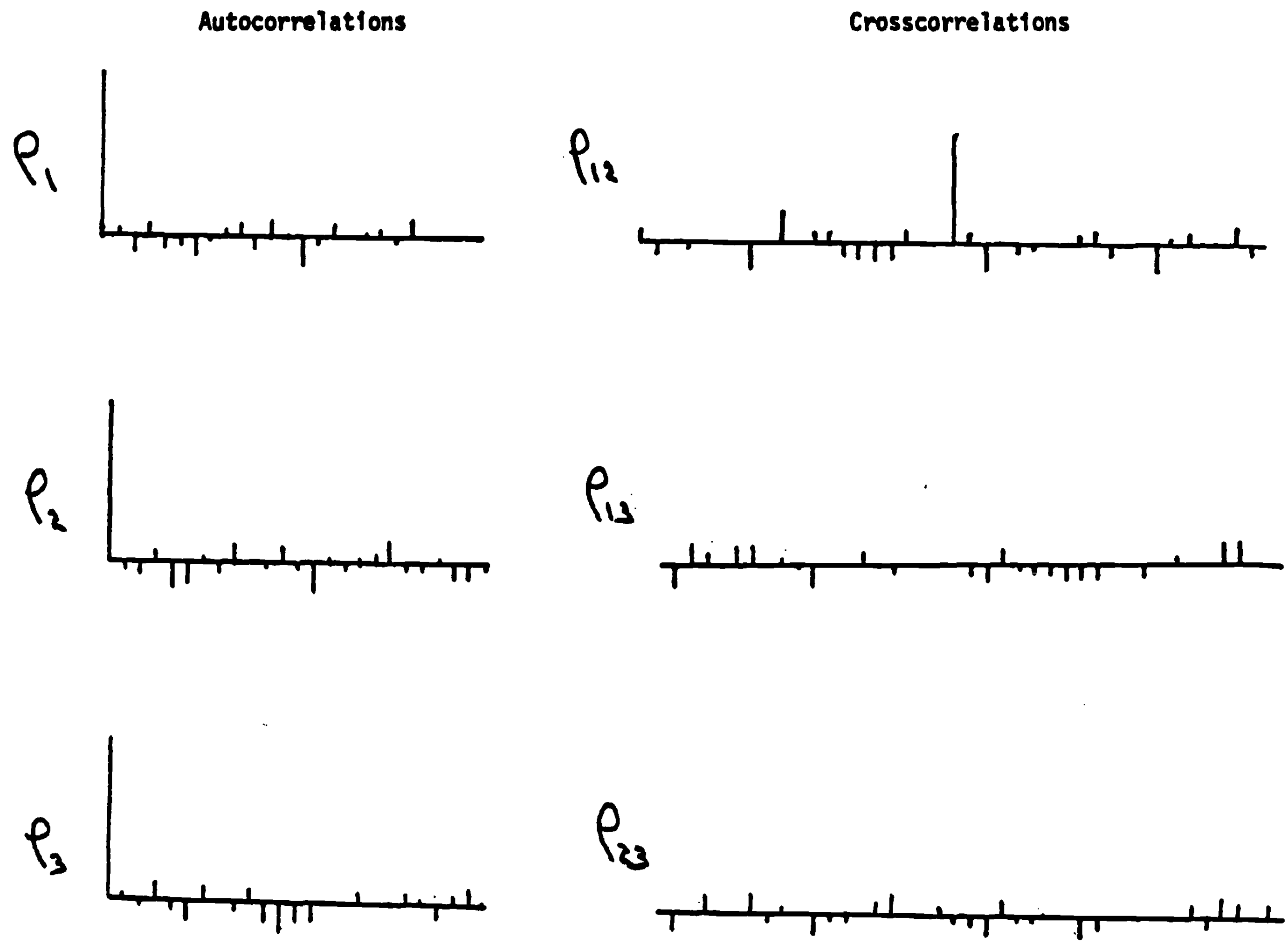


Figure 6. AUTOCORRELATIONS AND CROSSCORRELATIONS OF THE RESIDUALS FROM THE ESTIMATED MULTIVARIATE MODEL FOR THE DURABLES SERIES



be tentatively fixed at any desired value and, in particular, to be set equal to zero. It also allows coefficients to be estimated under a given constraint. For example, two coefficients may be estimated subject to the constraint that they be equal. This permits any desired simplification to be explored and makes it easier to see the implications of a given model. In the present example, the matrices $\hat{\Phi}_{12}$ and $\hat{\theta}_{12}$ are, apart from elements that are not large compared with their standard errors, very nearly diagonal. Also, the diagonal elements in $\hat{\Phi}_{12}$ are large, two of them being close to unity. Now, moderate changes made in $\hat{\Phi}_{12}$ can be very nearly compensated by appropriate changes in $\hat{\theta}_{12}$. In refitting, therefore, Φ_{12} was simplified to be an identity matrix, and θ_{12} was made diagonal. Furthermore, corresponding elements in (1) the first row and second column and (2) the second row and third column of $\hat{\Phi}_1$ and $\hat{\theta}_1$ are not significantly different. This has a particular interpretation that forecasted values may be used directly in the equations. Thus, in refitting, each of these two pairs of elements were tentatively set equal. The results are given in (80). It will be noted that the diagonal elements of $\hat{\Sigma}$, associated with the one step ahead forecast errors of the refitted model, are somewhat larger than those in (79) but still much smaller than the corresponding variances in (77) obtained for the univariate models. No strong evidence of lack of fit was shown in the residual analysis.

$$\hat{\Phi}_1=\begin{bmatrix} \underset{(0.05)}{0.79} & \underset{(0.05)}{0.20} & \cdot \\ \cdot & \underset{(0.06)}{0.97} & \underset{(0.36)}{-0.69} \\ \cdot & \cdot & \underset{(0.04)}{0.90} \end{bmatrix},\ \hat{\Phi}_{12}=\begin{bmatrix} 1. & \cdot & \cdot \\ \cdot & 1. & \cdot \\ \cdot & \cdot & 1. \end{bmatrix}$$

$$\hat{\theta}_1=\begin{bmatrix} \cdot & \underset{(0.05)}{0.20} & \underset{(0.25)}{-1.72} \\ \cdot & \underset{(0.80)}{0.36} & \underset{(0.36)}{-0.69} \\ \cdot & \cdot & \underset{(0.08)}{0.40} \end{bmatrix},\ \hat{\theta}_{12}=\begin{bmatrix} \underset{(0.04)}{0.77} & \cdot & \cdot \\ \cdot & \underset{(0.05)}{0.74} & \cdot \\ \cdot & \cdot & \underset{(0.05)}{0.75} \end{bmatrix}$$

$$\hat{\Sigma}=\begin{bmatrix} 0.000644 & 0.000711 & -0.000007 \\ \cdot & 0.001655 & -0.000002 \\ \cdot & \cdot & 0.000027 \end{bmatrix} \qquad (80)$$

Implications of the model—If we write

$$\underset{\sim}{w}_t=(I-\theta_{12}B^{12})^{-1}(I-\Phi_{12}B^{12})\underset{\sim}{z}_t \qquad (81)$$

then

$$w_{1t}=z_{1t}-\hat{z}_{1,\,t-12},\ w_{2,\,t}=z_{2t}-\hat{z}_{2,\,t-12},\ w_{3t}=z_{3t}-\hat{z}_{3,\,t-12}$$

where, for example,

$$\hat{z}_{1,\,t-12}=0.23(z_{1,\,t-12}+0.77z_{1,\,t-24}+0.77^2z_{1,\,t-36}+\dots)$$

Thus, the w's are deviations from seasonal exponentially discounted averages and have the effect of removing the seasonal component. The relationships between these deviations are now approximately as follows:

$$w_{1t}-0.8w_{1,\,t-1}=a_{1t}+0.2(w_{2,\,t-1}-a_{2,\,t-1})+1.7a_{3,\,t-1} \qquad (82)$$

$$w_{2t}-w_{2,\,t-1}=a_{2t}-0.4a_{2,\,t-1}-0.7(w_{3,\,t-1}-a_{3,\,t-1}) \qquad (83)$$

$$w_{3t}-0.9w_{3,\,t-1}=a_{3t}-0.4a_{3,\,t-1} \qquad (84)$$

We will now consider the implications of these equations.

Inventories—Inspection of $\hat{\Sigma}$ in (80) shows that there is no evidence that a_{3t} is correlated with a_{1t} and a_{2t}; also, (84) does not contain w_{1t} or w_{2t}. Thus, rather remarkably, the inventory series behaves independently of the other two series, and its complementary (forecast) function is very nearly such that

$$(1-B)^3\,(1+B+\cdots+B^{11})\hat{x}_t(l)=0$$

where $\hat{x}_t(l)$ is the log of the (undifferenced) inventories. The solution, thus, involves a quadratic trend component plus a seasonal component. This holds out the possibility that upturns and downturns in inventories might be forecast.

New orders—Again, using a notation adopted earlier, we write $x_t^{(\theta)}$ to mean an exponentially smoothed value with smoothing coefficient θ, $x_t^{(\theta)}=(1-\theta)\sum_{j=0}^{\infty}\theta^j x_{t-j}$.

Thus, equation (83) may be written aproximately as

$$w_{2t}=w_{2,\,t-1}^{(.4)}-\hat{w}_{3,\,t-2}^{(.4)}(1)+a_{2t},\ \sigma^2_{a_2}=0.00166$$

The implication is that the change in new orders from last month's (smoothed) value is just such as will balance the (smoothed) value forecast for last month's change in inventories. Since we are dealing with logged data, it should be remembered that change refers to percentage change.

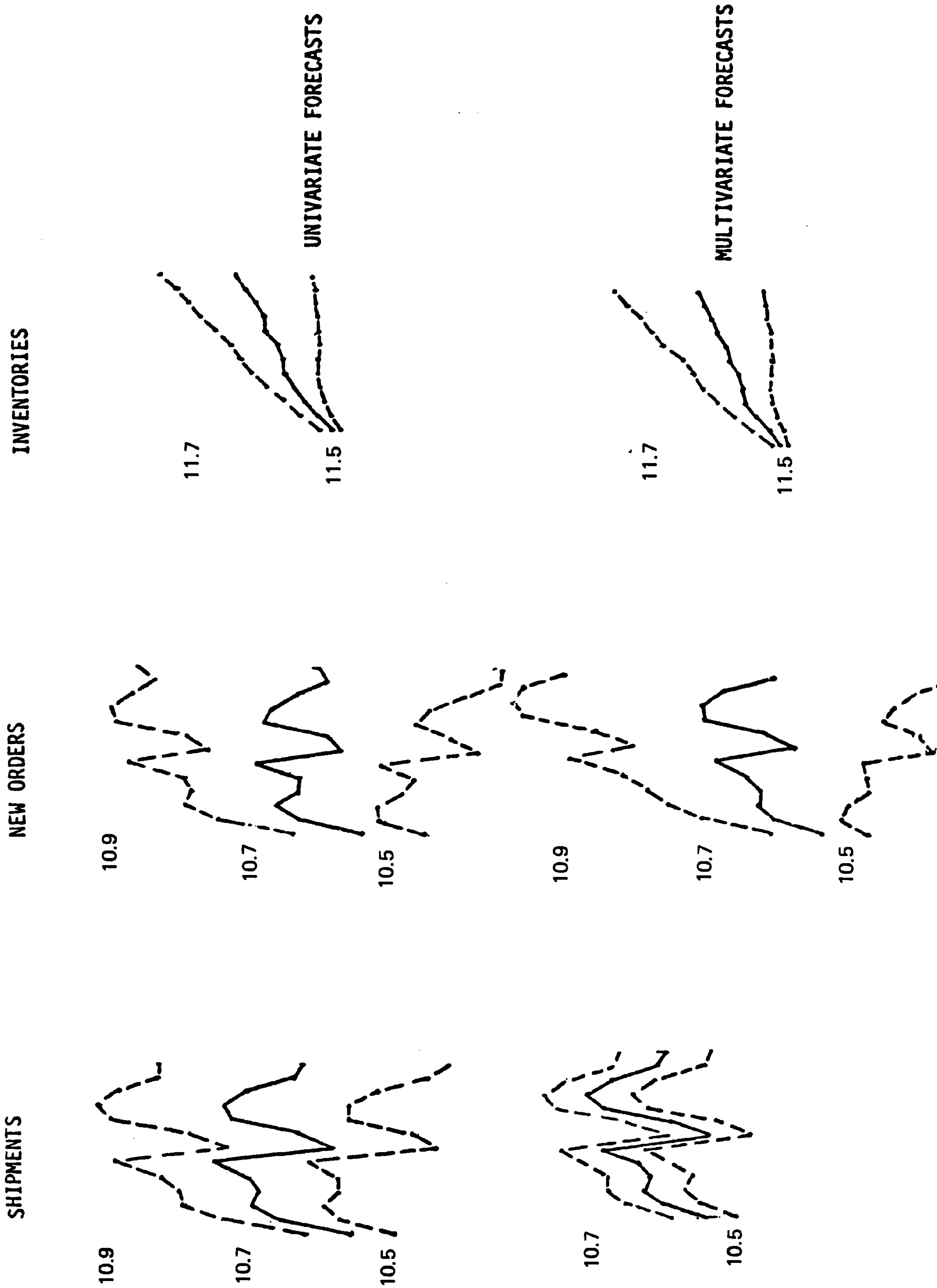
Shipments—Equation (82) may be written

$$\begin{aligned} w_{1t}=w_{1,\,t-1}&+0.2\{\hat{w}_{2,\,t-2}(1)-w_{1,\,t-1}\}\\ &+1.7\{w_{3,\,t-1}-\hat{w}_{3,\,t-2}(1)\}+a_{1t} \end{aligned}$$

Now, whereas a_{3t} appears to be independent of a_{1t} and a_{2t}, a_{1t} and a_{2t} have a correlation of 0.69. Correspondingly, if we write

$$a_{1t}=0.43a_{2t}+a'_{1t},\ \text{with } \sigma^2_{a'_1}=0.000339$$

Figure 7. FORECASTS FROM THE UNIVARIATE AND MULTIVARIATE MODELS FOR THE DURABLES SERIES



then a'_{1t} is uncorrelated with a_{2t}. Thus, we can obtain three equations that have independent errors by taking the third to be

$$w_{1t}=w_{1,\,t-1}+0.2\{\hat{w}_{2,\,t-2}(1)-w_{1,\,t-1}\}+1.7\{w_{3t-1}-\hat{w}_{3,\,t-2}(1)\}$$
$$+0.4\{w_{2t}-\hat{w}_{2,\,t-1}(1)\}+a'_{1t}\ \sigma^2_{a'_1}=0.00034$$

Thus, this month's shipments are increased over last month's when—

1. Forecast orders for last month exceeded shipments.
2. Increases in inventories for last month exceeded that forecasted.
3. Current new orders exceed last month's forecast.

It is gratifying to see that this model makes economic sense.

Forecasting—As already noted, there is a substantial reduction in the variance of the shipments and new order series when the multivariate model is used. A corresponding increase is found in the accuracy of the forecasts, as shown in figure 7. Therefore, by multivariate extensions of the arguments used in this paper, it is to be expected that even more precise methods for smoothing and seasonal adjustment should be derivable for sets of related series.

SUMMARY

It is argued that most rapid progress in statistical methods occurs when the empirical approach and the model-based approach iteratively interact. Such progress has led to the useful and rich class of ARIMA models that may be fitted to a wide variety of time series. We illustrate, in this paper, how a fitted model can then determine appropriate techniques for smoothing and seasonal adjustment of a particular series.

The development of methods for convenient computation of exact maximum likelihood estimates for multivariate extensions of these models makes the multivariate ARIMA models more accessible. Using the shipments-order-inventories series, we have illustrated how such multivariate models permit the analysis of complex relationships, allow more accurate forecasts, and have the potential for improving smoothing and seasonal adjustment methods still further.

REFERENCES

1. Anderson, T. W. "Maximum Likelihood Estimation of Parameters of Autoregressive Processes with Moving Average Residuals and Other Covariance Matrices with Linear Structure." *Annals of Statistics*. 3 (1975): 1283–1304.
2. Beveridge, W. H. "Wheat Prices and Rainfall in Western Europe." *Journal of the Royal Statistical Society* 85 (1922): 412–459.
3. Box, G. E. P., and Jenkins, G. M. "Some Statistical Aspects of Adaptive Optimization and Control." *Journal of the Royal Statistical Society*, ser. B, 24 (1962): 297–331.
4. ———. "Further Contributions to Adaptive Control; Simultaneous Estimation of Dynamics: Non-Zero Costs." *Bulletin of the International Statistical Institute 34th Session* (1963): 943–974.
5. ———. "Mathematical Models for Adaptive Control and Optimization." *American Institute of Chemical Engineering Institute of Chemical Engineering Symposium Series* 4 (1965): 61–68.
6. ——— and Bacon, D. W. "Models for Forecasting Seasonal and Non-Seasonal Time Series." *Advanced Seminar on Spectral Analysis of Time Series*. Edited by B. Harris. New York: John Wiley and Sons, Inc., 1967.
7. ———. "Some Recent Advances in Forecasting and Control I." *Applied Statistics* 17 (1968): 91–109.
8. ———. "Discrete Models for Forecasting and Control," *Encyclopedia of Linguistics, Information and Control*. Calcutta: Pergamon Press, 1969, pp. 162–167.
9. ———. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day, 1970.
10. Box, G. E. P., and Tiao, G. C. "Intervention Analysis with Applications to Economic and Environmental Problems." *Journal of the American Statistical Association* 70 (1975): 70–79.
11. Brown, R. G. *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill, 1959.
12. ———. *Smoothing, Forecasting and Prediction of Discrete Time Series*. Englewood Cliffs, N.J.: Prentice-Hall. 1963.
13. ——— and Meyer, R. F. "The Fundamental Theory of Exponential Smoothing." *Operations Research* 9 (1961): 673–685.
14. Carnegie Institute of Technology. "Forecasting Trends and Seasonals by Exponentially Weighted Moving Averages." Office of Naval Research Memorandum No. 52, by C. C. Holt, 1957.
15. Cleveland, W. P. "Analysis and Forecasting of Seasonal Time Series" Ph.D. dissertation, University of Wisconsin, Statistics Department, 1972.
16. ——— and Tiao, G. C. "Decomposition of Seasonal Time Series: A Model for the Census X–11 Program." *Journal of the American Statistical Association* 71 (1976): 581–587.
17. Dent, W. "Computation of the Exact Likelihood Function of an ARIMA Process." Technical Report No. 45, University of Iowa, Statistics Department, 1975.
18. Fisher, I. "On Unstable Dollar and the So-Called Business Cycle." *Journal of the American Statistical Association*, 20 (1925): 179.
19. Fisher, R. A. "The Influence of Rainfall on the Yield of Wheat at Rothamsted." *Philosophical Transactions of the Royal Society of London*, B, 213 (1924): 89–142.
20. Hannan, E. J. *Multiple Time Series*. New York: John Wiley, 1970.
21. ——— and Dunsmuir, W. "Vector Linear Time Series Models," 1975. (Unpublished manuscript.)
22. Hillmer, S. C. "Time Series: Estimation, Smoothing and Seasonal Adjusting," Ph.D. dissertation, University of Wisconsin, Statistics Department, 1976.
23. Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A. *Planning Production, Inventories and Work Force*. Englewood Cliffs, N.J.: Prentice-Hall, 1960.
24. Kolmogorov, A. "Interpolation and Extrapolation von Stationären Zufälligen Folgen." *Bulletin Academy of Science, U.S.S.R. Series Math* 5 (1941): 3–14.
25. Ljung, G. "Studies in the Modeling of Discrete Time Series." Ph.D. dissertation, University of Wisconsin, Statistics Department, 1976.
26. Mayr, O. *The Origins of Feedback Control*. Cambridge, Mass.: M.I.T. Press, 1970.
27. Muth, J. F. "Optimal Properties of Exponentially Weighted Forecasts of Time Series with Permanent and Transitory Components." *Journal of the American Statistical Association* 55 (1960): 299–306.

28. Nelder, J. A. and Mead, R. "A Simplex Method for Function Minimization." *The Computer Journal* 7 (1965): 308–313.

29. U.S. Department of Commerce, Bureau of the Census. *The X-11 Variant of Census Method II Seasonal Adjustment Program*, by Julius Shiskin, Allan H. Young, and J. C. Musgrave. Technical Paper 15. Washington, D.C.: Government Printing Office, 1967.

30. Weiner, N. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: John Wiley, 1949.

31. Whittle, P. "Gaussian Estimation in Stationary Time Series" *Bulletin Institute International Statistics* 33 (1961): 1–26.

32. ———. *Prediction and Regulation*. New York: D. Van Nostrand, 1963.

33. Winters, P. R. "Forecasting Sales by Exponentially Weighted Moving Averages." *Management Science* 6 (1960): 324–342.

34. Yaglom, A. M. "The Correlation Theory of Processes Whose nth Difference Constitute a Stationary Process." *American Mathematical Society Translations*, ser. 2 and 8 Providence, R.I.: American Mathematical Society, 1955.

35. Yule, G. U. "On the Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers." *Philosophical Transactions*, A, 226 (1927): 267–298.

APPENDIX

EXPANSION OF THE GENERATING FUNCTION $\phi(B)\phi(F/\eta(B)\eta(F)$

For the convenience of the reader, we will sketch a method by which the weight functions ω_j's in (42) may be determined. Note that $\phi(B)$ and $\eta(B)$ are polynomials in degree p and u, respectively. With no loss in generality we may write

$$\phi(B)=1-\phi_1 B-\ldots-\phi_r B^r,\ \eta(B)=1-\eta_1 B-\ldots-\eta_r B^r$$

where $r=\max(p, u)$. First set

$$\frac{\phi(B)}{\eta(B)}=C(B), \text{ where } C(B)=1+C_1B+C_2B^2+\ldots$$

and solve for the C_j's by matching coefficients of B^j in

$$\phi(B)=\eta(B)C(B) \tag{A.1}$$

Specifically,

$$C_1=\eta_1-\phi_1,\ C_j=\eta_j-\phi_j+\sum_{i=1}^{j-1} C_{j-i}\eta_i,\ j=2, \ldots, r \tag{A.2}$$

and for $j>r$, the C_j's can be recursively computed from the relation

$$C_j=\sum_{i=1}^{r} C_{j-i}\eta_i \tag{A.3}$$

Next, set

$$\frac{\phi(F)}{\eta(F)}C(B)=X(B, F)$$

where

$$X(B, F)=X_0+\sum_{j=0}^{\infty} X_j(B^j+F^j)$$

Equivalently,

$$\phi(F)C(B)=\eta(F)X(B, F) \tag{A.4}$$

Note that the largest degree of F on the left hand side of (A.4) is r. By matching coefficients of F^j in (A.4), we find that

1. For $j=0, 1, \ldots, r$

$$\underset{\sim}{\eta}\underset{\sim}{X}=\underset{\sim}{C}\underset{\sim}{\phi} \tag{A.5}$$

where

$$\underset{\sim}{X}'=(X_0, \ldots, X_r),\ \phi'=(1, -\phi_1, \ldots, -\phi_r)$$

$$\underset{\sim}{C}=\begin{bmatrix} 1 & C_1 & \cdot & \cdot & C_r \\ & \cdot & \cdot & & \cdot \\ & & \cdot & \cdot & \\ & & & & \cdot \\ & & & \cdot & C_1 \\ \underset{\sim}{0} & & & & 1 \end{bmatrix} \quad \underset{\sim}{\eta}=\begin{bmatrix} 1 & -\eta_1 \ldots & -\eta_r \\ -\eta_1 & & -\eta_r \\ \cdot & & \cdot \\ \cdot & \cdot & \\ \cdot\ \cdot & & \underset{\sim}{0} \\ -\eta_r & & \end{bmatrix}$$

$$+\begin{bmatrix} 0 & & & & \underset{\sim}{0} \\ \cdot & 1 & & & \\ \cdot & -\eta_1 & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & -\eta_{r-1} & \cdot\ \cdot & -\eta_1 & 1 \end{bmatrix}$$

which define a set of $r+1$ linear equations in $r+1$ unknowns $X_0, \ldots, X_r$, so that

$$\underset{\sim}{X}=\underset{\sim}{\eta}^{-1}\underset{\sim}{C}\underset{\sim}{\phi} \tag{A.6}$$

2. For $j>r$, the X_j's can be computed recursively from the relation

$$X_j=\sum_{i=1}^{r} X_{j-i}\eta_i \tag{A.7}$$

Finally, the weights ω_j's in (42) are

$$\omega_0=1-\frac{\sigma_e^2}{\sigma_a^2}X_0 \text{ and } \omega_j=-\frac{\sigma_e^2}{\sigma_a^2}X_j,\ j=1, \ldots \tag{A.8}$$

COMMENTS ON "ANALYSIS AND MODELING OF SEASONAL TIME SERIES" BY GEORGE E. P. BOX, STEVEN C. HILLMER, AND GEORGE C. TIAO

George A. Barnard
University of Essex

I wish to make clear, at the outset, that I have no claim to expertise in the matter of time series analysis. On the rare occasions when I have referred to stochastic processes in published papers, I have taken care to base my remarks on models to which none of the existing theory applies. As the domain of existing theory extends, it becomes a little more difficult to do this—but only a little more difficult. It is, perhaps, worthwhile to point out that, since the universe, in its evolution, can be regarded as the exemplification of a stochastic process, a full treatment of the theory of such processes still has a very long way to go before it can be said to be complete. Such a reflection is not merely philosophical. It is relevent to the issues we are here to discuss in that attempts to claim universality for any one method of analysis are bound to fail. The iterative process of empirical analysis, interacting with model building so well emphasized by the authors of this paper, should converge when related to a particular case, on the basis of given data; but, in relation to the set of all possible cases, it will clearly have an open-ended character.

Such knowledge as I have gained over the years in connection with time series has been almost wholly derived from conversations with George Box and with Gwilym Jenkins. These began in the mid-fifties, when I had the privilege of hearing Box's first thoughts on the subject of nonstationary series. Perhaps he found me a sympathetic listener, because one of the difficulties I had had with the subject, up to that time, was with the near universality of the assumption of stationarity, which conflicted fundamentally with my own evolutionary view of nature and society. I record this here, because it is not clear, from the sketch of history given in the paper, that although it has subsequently turned out that their work was in some respects anticipated by others, such as Yaglom, Box and Jenkins developed their theoretical and practical approach almost entirely independently. I believe that the coherence of the various aspects and methods, which is such a strong feature of their work and that of their school, is, to a large extent, due to this.

Up to now, the lessons in time series analysis that I have had from Box and Jenkins have come in doses sufficiently small and well expounded that I have been able to digest them without undue effort. Perhaps, because the present instalment represents the work of several of Box's coworkers at Madison, it comes here as an advance in the theory of such magnitude that its implications will take me, at least, a considerable time to assimilate. The traditional approach to seasonality and smoothing has been to process the series in question through a set of filters to remove, as far as possible, the high frequency noise, and the excess amplitude of frequencies associated with seasonality and, then, to set up a model for the residual trend. What the authors propose here is to turn this process upside down—to model the process as it stands and, then, express the model as the sum of the traditional three components. That something of this kind needs to be done, at least in some cases, especially in economic contexts, is indicated by the experience of most workers in the field, that even the highly sophisticated seasonality adjustments, such as the census X–II program or the mixed-multiplicative and additive-adaptive methods studied on the other side of the Atlantic, always seem to leave some element of seasonal pattern in each series to which they are applied. And, it is common sense to think that the way in which each series reacts to and remembers seasonal factors is bound to have some degree of individuality about it, so that methods which, in principle, assume the same type of reaction to seasonal influences are bound to be limited in application.

In so far as our current ideas and practice with seasonal adjustments are based on dating, we are, in effect, assuming that the factors that are influencing the series we are concerned with and whose effect we wish to separate out, are directly and simply related to the position of the earth in its orbit round the sun. This will be true of series in which the date, by itself, has major significance—one thinks primarily here of the influence of, e.g., Christmas on sales figures. But, already with the date of Easter, we have to consider not only the position of the earth relative to the sun, but also its position in relation to the Paschal moon, so that series (such as the Irish consumption of Guinness Stout) for which Easter is an important festival will exhibit peaks that sometimes occur in March and sometimes in April. And, in the many series for which the major influence of this kind is the weather, the patterns will still be more complicated. Most series will be influenced by a mixture of such factors, in proportions peculiar to the particular series, and, perhaps more important, the degree to which these influences are remembered by the system in question will vary, so that an approach, such as that taken in this paper, would certainly have advantages.

Whether the advantages would, in all cases, be worth the extra effort, involved in dealing with each series individually, is a point worth discussion.

My own understanding of the history of the exponentially weighted moving average is that the idea occurred to my friend Arthur A. Brown, working for Arthur D. Little, who was working on stock control problems. It occurred to him, waiting for a plane at O'Hare International Airport, that the EWMA would require, in a computerised setup, only two storage locations per series to be forecast. I say this, partly by way of comment on the historical section of the author's paper, but more especially because the subsequent history of the associated model would suggest that an optimistic attitude to the simplicity of the real world (something that one is trained as a mathematician to despise—mostly to one's detriment) has a lot to be said in its favour. Correspondingly, while, as I have said, I have little doubt that the approach to seasonality, proposed by the authors, will have, in principle, many advantages, the differences between the final results of the census treatment of the unemployed males series and that of the authors' is perhaps arguably small enough to be ignored for most "practical" purposes.

I put "practical" in quotation marks, because I wish it to be understood in the sense in such contexts where there is usually an element of the short run involved in the "practical." In the long run, the difference of principle involved in the difference of approach can have enormous practical consequences. Because, as the authors note, it raises the whole question of why we attempt to deseasonalise our series, and what we understand ourselves to be doing when we so treat them.

One approach to this question would suggest that we can draw a distinction between factors, such as the motion of the earth around the sun or variations in the weather, that are exogenous to the social system and that are altogether beyond our control, and, on the other hand, factors, such as tax rates, bank rates, and perhaps social attitudes, that also are exogenous to most aspects of the social system but which are, to some extent, within our control. Usually, it is assumed that the factors beyond our control will, in some sense, average themselves out in the future, and what planners and politicians need to concern themselves with are the controllable factors. The recent drought in Western Europe is exceptional in that it has brought into prominence, for practical politicians, the fact that most of their policies are based on an extrapolation of present weather conditions that may, in fact, be unjustified. To help with this, they need to have a model of the times series they are concerned with that has the effects of the uncontrollable factors removed or averaged out. If we understand the purpose of deseasonalisation, in this sense, we shall want to further develop the ideas expressed in this paper, especially in relation to the sketch of the approach to multiple time series. We shall also want to develop the ideas of intervention analysis that Box and Tiao have put forward but which they have not dealt with in the present contribution.

But, perhaps, a simpler notion will serve. It is that those who use deseasonalised and smoothed series regard such series as representing the average, relatively long-term behaviour of the series in question. Smoothing gets rid of the month-to-month random fluctuations, while deseasonalising enables one to get an idea of the movement of the yearly average. If this point of view is adopted, then instead of focussing attention, as we do now, on the periodic motion of the earth, we may begin to pay attention to what perhaps in some countries, is becoming a more important length of time—the 4-, 5-, or 6-year term between elections. Would it be too much to hope that one day statisticians will be able to discount the effects of the loosening of credit and other inflationary measures that are becoming all too common, even in advanced countries in preelection periods, so that politicians will be judged more on the long-term effects of their policies? The common man has much more sense than the credit he is usually given. If he is given the information that he needs to judge the long-term effects of political and economic policies, he can be relied upon to look to the reasonable future, instead of the immediate present.

What I am suggesting is that instead of producing, along with the raw data, the smoothed and deseasonalised values, we should think in terms of producing (for instance) a forecast of the discounted future values of the series that we are concerned with in the notation of their paper, e.g.,

$$\hat{\hat{Z}} = \sum_{l=1}^{\infty} \hat{z}_{t+l}(1-\delta)^{l-1}\,\delta$$

where δ is a discounting factor that might be around 0.008 for a monthly series. Such forecasts, updated from month to month, would surely be a better guide to policy and its effects than the values we now use. It is a feature of the methods proposed by the present authors that they provide a reasonably objective procedure for doing this. The necessity of having such an objective procedure is, perhaps, illustrated by some of the U.K. Government's recent *White Papers,* where assumptions about, e.g., growth rates are embodied without proper stress being laid upon the possibility that such assumptions could prove wrong.

To return to the paper and its details, it is perhaps worth emphasising that one aspect of the model-building process that can make it an essential adjunct to the empirical approach is its capacity to bring to bear information different, in kind, from the basic numerical data. As an example, it could be that someone who knew the processes used in stores in adjusting their prices could bring evidence that the extreme value $\alpha = -1$ for the coefficient of B in the forcing function for the smoothed model of the consumer price index was too large in absolute value, and a value near to $-1/2$ would better accord with practical experience. Examining the curves, given for $\alpha = 0$ and for $\alpha = -1$, shows that the data, by themselves, throw very little light on the values of α within this range.

The authors' somewhat apologetic tone in referring to their use of the Nelder-Mead procedure in solving the equations for the coefficients of the seasonal component of the male unemployment series is surely not appropriate. Methods of function minimisation, developed by Powell and others in recent years, have become so powerful that even for the solution of large sets of linear equations it may be preferable to minimise directly the sum of a residuals rather than use the (unstable) algorithms of pivotal condensation, etc. If I may suggest it, perhaps the authors should draw their problems to the attention of a numerical analyst working in this area; one gets the impression that theory has run ahead of practical problems.

Finally, in reference to the remarks concerning the calculation of the likelihood function and the virtues of maximum likelihood estimates, perhaps, the name of Ian McLeod should be added to that of Kang as having provided a practicable procedure for obtaining more exact maximums and showing, by simulation, that the improvement is worthwhile. And, it should, perhaps, be mentioned that Barnard and Winsten showed, using the data cited by Whittle in the paper referred to, that examination of the whole likelihood function threw a much clearer light on the estimate and its error than would come from calculations based on asymptotic results. The problem of exhibiting, in a usable form, the likelihood function for many parameters is not yet solved; but, good progress is being made. There really is no substitute for this, in the case of most economic time series, since the effective lengths of series are typically much too small for asymptotic results to be of relevance.

Since the form and content of my remarks is typical to an opening of discussion on my home round at the Royal Statistical Society, may I follow the custom there and conclude by moving, with much pleasure, a hearty vote of thanks to our three authors for a most rich and stimulating paper.

COMMENTS ON "ANALYSIS AND MODELING OF SEASONAL TIME SERIES" BY GEORGE E. P. BOX, STEVEN C. HILLMER, AND GEORGE C. TIAO

Emanuel Parzen
State University of New York at Buffalo

I would like to express my thanks to the authors for their important and seminal paper that will contribute greatly towards clarifying what is meant by seasonality in time series and developing useful methods for its analysis and modeling.

Given observed time series (or, in general, statistical data) that we seek to model, we are confronted with the basic questions: What do we mean by a model? and where do models come from? One can distinguish between two types of models that can be given names as follows:

Type I: Fundamental or structural
Type II: Technical or synthetic

I will use the adjectives "structural" and "synthetic."

A structural model comes from subject-matter theoretical considerations, and the statistician's role is usually to estimate its parameters or test its fit to data.

A synthetic model comes from statistical and probabilistic considerations, either from empirical data analysis procedures that have been found satisfactory, in practice, or from representation theory of random variables.

In time series analysis, both statistical and probabilistic considerations yield a simple general definition of what we mean by a model: It is a transformation (often a linear filter) F on the data $\{z_t\}$ that whitens the time series $\{z_t\}$. In symbols, let $\{a_t\}$ be a white-noise series such that

$$\{z_t\} - \boxed{F} - \{a_t\}$$

then the transformation F is the model.

The essence of the successful Box-Jenkins approach to time series analysis seems to be to develop empirical procedures for discovering a filter F (in the univariate case of the form $\theta^{-1}(B)\varphi(B)\nabla^d$) for which one can write conclusions similar to that which the authors write following their model in equation (79): "The residuals from this model are plotted in figure 5 and the autocorrelation and cross-correlation functions of the residuals are plotted in figure 6. Examination of these plots does not suggest any inadequacies in the model."

The models (means) to be used in time series analysis should depend on the intended ends or applications (thus, the eternal conundrum: Can one find means robust against ends?). I believe one may distinguish six basic types of applications of time series analysis:

1. Forecasting (or extrapolation)
2. Spectral analysis (or interpolation, by harmonics)
3. Parametrization (or data compression)
4. Intervention analysis (significant changes in forecasts or parameters)
5. Signal plus noise decomposition
6. Control

There is no doubt that ARIMA models are important and can provide means for all these ends. However, other ways of formulating equivalent models should not be discarded from (or fail to be incorporated into) the time series analyst's bag of tools. In particular, spectral and state space representations are often indispensable means.

I believe there can be no disagreement with the conclusion stated at the end of the section on the iterative development of some important ideas in time series analysis that ARIMA models are "a class of stochastic models capable of representing nonstationary and seasonal time series" and with the assertion that a successful seasonal nonstationary model for economic time series with both monthly and yearly components is given by their equation (28), which we shall call model I:

$$(I-\underline{B})(I-\underline{B}^{12})\, z_t=(I-\theta_1\underline{B})(I-\theta_2\underline{B}^{12})\, a_t$$

A question that I believe should be investigated is whether model I, by itself, can yield seasonal adjustment procedures, or is it necessary to pass (algorithmically and conceptually) from model I to the traditional representation of a time series z_t as having a seasonal component S_t and trend-noise component T_t, which we call model II:

$$z_t=S_t+T_t=S_t+p_t+e_t$$

where $T_t=p_t+e_t$ is the sum of a trend p_t and white noise e_t. The authors' main aim in the section on smoothing and seasonal adjustment is to outline approaches to pass from model I to model II. Whether this trip is necessary will be discussed.

Some technical comments on the section on smoothing and seasonal adjustment are the following: In estimating p_t from T_t, the authors consider only two-sided filters that use both past and future values of T_t; would it not be more appropriate to use one-sided filters that use only past values? Then, to estimate p_t from T_t, one would use Kalman filtering techniques. In regard to estimating models for p_t and e_t, an important reference seems, to me, to be Pagano [1].

The aim of the section on multivariate generalization of the models is to develop multivariate seasonal models such as that given by equation (75), which we call model III:

$$(\mathrm{I}-\Phi_1 \underset{\sim}{B})(\mathrm{I}-\Phi_{12}\underset{\sim}{B}^{12})\underset{\sim}{z}_t=(\mathrm{I}-\theta_1\underset{\sim}{B})(\mathrm{I}-\theta_{12}\underset{\sim}{B}^{12})\underset{\sim}{a}_t$$

where $\underset{\sim}{z}_t$ is a k vector and Φ_i and θ_j are $k\times k$ matrices. The time available to me for discussion prevents my commenting further on the section concerning multivariate generalization of the models other than to note my view that its results are pioneering and impressive and should stimulate much further research.

I believe this paper is important and seminal, because the problems considered by Box, Hillmer, and Tiao in the third and fourth sections are the problems at the heart of the problem of seasonal adjustment. In the remainder of my discussion, I would like to outline, from the point of view of my own approach to empirical time series analysis, why and how it might suffice to obtain seasonal adjustment procedures directly from a suitable reinterpretation of models, such as I (and, more generally, III).

A fundamental decomposition of a time series $\{z_t\}$ can be given in terms of its one-step-ahead infinite-memory predictors

$$z_t^{\mu}=E[z_t|z_{t-1}, z_{t-2}, \ldots]$$

and its one-step-ahead infinite-memory prediction errors or innovations

$$z_t^{\nu}=z_t-z_t^{\mu}$$

Then

$$z_t=z_t^{\mu}+z_t^{\nu}$$

The innovation time series $\{z_t^{\nu}\}$ is white noise and, indeed, $\underset{\sim}{a}_t=z_t^{\nu}$

For a nonstationary time series, the modeling problem is first to find the whitening filter (that transforms $\{z_t\}$ to $\{z_t^{\nu}\}$) and, second, to interpret it as several filters in tandem:

D_0: A detrending filter that, in the spectral domain, eliminates the low-frequency components corresponding to trend.

D_λ: A deseasonal filter that, in the spectral domain, eliminates the components corresponding to a periodic component with period λ or to the harmonics with frequencies that are multiples of $\dfrac{2\pi}{\lambda}$.

g_∞ or Π: An innovations filter that transforms the series $z_t^{(\text{stat})}=D_0D_\lambda z_t$, representing a transformation of z_t to a stationary series, to white noise.

The time series modeling problem is, thus, to find the filter representation

	Detrend	Deseasonal			Innovations	
z_t	D_0	D_{λ_1}	$\cdots$	D_{λ_k}	g_∞	$a_t=z_t^{\nu}$

where we admit the possibility of several different periods $\lambda_1,\ldots,\lambda_k$ (e.g., in monthly data λ values are often 12 and 3, in daily data λ values are often 7 and 365, and in hourly data λ values are often 24 and 168).

Given this decomposition, one can form various derived series:

$z_t^{(0)} = D_0z_t$, the detrended series.
$z_t^{(\lambda)} = D_\lambda z_t$, the seasonally adjusted series.
$z_t^{(0,\lambda)} = z_t^{(\text{stat})}=D_0D_\lambda z_t=D_\lambda D_0 z_t$, the detrended seasonally adjusted series.
$z_t^{\nu} = z_t^{(\text{white})}=g_\infty D_0 D_\lambda z_t$, the innovations series.

Instead of universal detrending and seasonal adjustment procedures, what is being suggested are filters accomplishing the same ends that are custom tailored for each series.

Such decompositions seem to be crucial to the study of the relations between time series $Y_1(\cdot)$ and $Y_2(\cdot)$. To study their relations, it seems clear that if one relates $Y_1(\cdot)$ and $Y_2(\cdot)$ without filtering, one will often find spurious relationships. It has been suggested, therefore, that one attempt to relate $Y_1^{\nu}(\cdot)$ and $Y_2^{\nu}(\cdot)$, the individual innovations of each series. What remains to be examined is the insight to be derived from relating $Y_1^{(\lambda)}(t)$ and $Y_2^{(\lambda)}(t)$, the seasonally adjusted series or $Y_1^{(\text{stat})}(t)$ and $Y_2^{(\text{stat})}(t)$, the detrended and seasonally adjusted series.

The question remains of how to find, in practice, the detrending and deseasonal filters. To seasonally adjust for a period λ in data, several possibilities are available that may be interpreted as seasonal adjustment filters.

A filter with the same zeroes in the frequency domain as some usual procedures, which is recursive (acts only on past values) and yields a variety of filter shapes (in the frequency domain) between a square wave and a sinusoid, is the one-parameter family of filters

$$D_\lambda(\theta)=\frac{I-B^{\lambda}}{I-\theta B^{\lambda}}$$

where the parameter θ is chosen (usually by an estimation procedure) between 0 and 1. When $\theta=0$, the filter is denoted ∇_λ and called λ-th difference.

To understand the role of the filter $D_\lambda(\theta)$, denote it for brevity by D and rewrite it, writing $I-B^{\lambda}=I-\theta B^{\lambda}-(1-\theta)B^{\lambda}$, we obtain

$$D=I-\frac{(1-\theta)B^{\lambda}}{I-\theta B^{\lambda}}=I-\{(1-\theta)(B^{\lambda}+\theta B^{2\lambda}+\theta^2 B^{3\lambda}+\dots)\}$$

Then, the output $z_t^{(D)}=Dz_t$ of a filter D with input z_t can be written

$$z_t^{(D)}=z_t-(1-\theta)\{z_{t-\lambda}+\theta z_{t-2\lambda}+\dots\}$$

In words, $z_t^{(D)}$ is the result of subtracting, from z_t, the exponentially weighted average of $z_{t-\lambda}$, $z_{t-2\lambda}$, ...

It seems to me open to investigation whether the filter D (of mixed autoregressive moving average type) is superior to the approximately equivalent autoregressive filter

$$D'=(I-B^{\lambda})(I+\theta B^{\lambda})=I-(1-\theta)B^{\lambda}-\theta B^{2\lambda}$$

whose output

$$z_t^{(D')}=D'z_t$$

can be written

$$z_t^{(D')}=z_t-(1-\theta)\,z_{t-\lambda}-\theta z_{t-2\lambda}$$

It appears to me that the role of moving averages in Box-Jenkins ARIMA models is to build filters of the type $D_{\lambda}(\theta)$. Thus, the ARIMA model

$$(I-B)(I-B^{12})z_t=(I-\theta_1 B)(I-\theta_{12}B^{12})a_t$$

should be viewed as the whitening filter

$$D_1(\theta_1)\,D_{12}(\theta_{12})\,z_t=\left(\frac{I-B}{I-\theta_1 B}\right)\left(\frac{I-B^{12}}{I-\theta_{12}B^{12}}\right)z_t=a_t$$

I should like to emphasize that the output of the filter $D_1(\theta_1)$ $D_{12}(\theta_{12})$ is often not white noise but is only a stationary time series. For purposes of one-step-ahead prediction, it is often not important to differentiate between the case that $D_{12}D_1z_t$ is white noise or not, since most of the predictability is obtained by finding a suitable transformation to stationarity of the form D_1D_{12}. (Next, we will discuss naive prediction as a transformation to stationarity.)

A moral can be drawn from the foregoing considerations. To find a transformation of a nonstationary time series to stationarity, it may suffice to apply pure differencing operators, such as $I-B$ and $I-B^{12}$. However, the transformation of the residuals to the innovations series should be expressed, if possible, in terms of factors corresponding to the filters $I-\theta_1B$ and $I-\theta_{12}B^{12}$, since such factors enable us to interpret the overall whitening filter

as a series of filters in tandem

which can be interpreted as helping to provide solutions to the seasonal adjustment problem.

NAIVE PREDICTION AND TRANSFORMATIONS TO STATIONARITY

To predict a time series z_t, one can often suggest a naive predictor of the form

$$z_t^{\text{naive}}=z_{t-\lambda_1}+z_{t-\lambda_2}-z_{t-\lambda_1-\lambda_2}$$

The prediction error of this predictor is given by

$$\tilde{z}_t=z_t-z_t^{\text{naive}}=z_t-z_{t-\lambda_1}-z_{t-\lambda_2}+z_{t-\lambda_1-\lambda_2}=(I-B^{\lambda_1})(I-B^{\lambda_2})z_t$$

In words, taking λ_1-th and λ_2-th differences is equivalent to forming the naive prediction errors.

A criterion that z_t be nonstationary is that it be predictable (in the sense that the ratio of the average square of $\tilde{z}_t$ to the average square of z_t is of the order of $1/T$). When $\tilde{z}_t$ is stationary (nonpredictable), one models it (see app.) by an approximate autoregressive scheme

$$\hat{g}_{\hat{m}}(B)\tilde{z}_t=a_t$$

which can be used to form $\tilde{z}_t^{\mu}$, the best one-step-ahead predictor of $\tilde{z}_t$.

The best one-step-ahead predictor of z_t is given by

$$z_t^{\mu}=z_t^{\text{naive}}+\tilde{z}_t^{\mu}$$

to prove this, note the identity

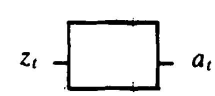
$$z_t=z_t^{\text{naive}}+\tilde{z}_t$$

and form the conditional expectation of both sides of this identity with respect to z_{t-1}, z_{t-2}, ...

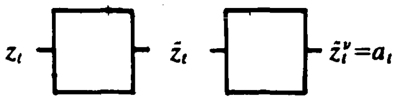
A remarkable fact is the equality of the prediction errors of z_t and $\tilde{z}_t$

$$z_t^{\nu}=z_t-z_t^{\mu}=\tilde{z}_t-\tilde{z}_t^{\mu}=\tilde{z}_t^{\nu}$$

It follows that, to find the whitening filter



for a nonstationary time series z_t (which includes almost all time series with seasonal components), it suffices to apply any one-sided filter (which, in practice, would be suggested by an ad hoc deseasonalizing procedure) whose output $\tilde{z}_t$ is stationary. The tandem filter



then yields the whitening filter. While the filter leading to $\tilde{z}_t$ is not unique, the overall filter leading to a_t is unique.

The final seasonal adjustment procedure is a filter D_λ that comes from interpreting the overall whitening filter as a series of filters in tandem, which can be interpreted as detrending and deseasonalizing filters. An illustration of the application of this approach to real data is contained in [2].

REFERENCES

1. Pagano, M. "Estimation of Autoregressive Signals Plus White Noise." *Annals of Statistics* 2 (1974): 99–108.

2. Parzen, Emmanuel. "An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands," *Proceedings on Forecasting Methodology for Time-of-Day and Seasonal Electric Utility Loads.* Special Report 31. Washington, D.C.: Electric Power Research Institute, 1976.

3. ———. "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes." In *Multivariate Analysis IV*. Edited by P. Krishnaiah. New York: Academic Press, 1976.

APPENDIX

ESTIMATION OF THE WHITENING FILTER OF A STATIONARY TIME SERIES

A rigorous definition of the whitening filter can be given when the time series $\{z_t\}$ is stationary zero mean and has a continuous spectral density function $f(\omega)$, where, for mathematical convenience, we use the definition

$$f(\omega)=\frac{1}{2\pi}\sum_{v=-\infty}^{\infty} e^{-iv\omega}R(v),\ -\pi\leq\omega\leq\pi$$

$$R(v)=E[z_t z_{t+v}]$$

Assume that $\log f(\omega)$ and $f^{-1}(\omega)$ are integrable; then there is a frequency transfer function

$$g_\infty(z)=1+\alpha_1 z+\alpha_2 z^2+\ldots+\alpha_m z^m+\ldots$$

such that

$$f(\omega)=\frac{1}{2\pi}\sigma_\infty^2|g_\infty(e^{i\omega})|^{-2}$$

Further, g_∞ is a whitening filter

$$g_\infty(B)\,z_t=a_t$$

where $\{a_t\}$ is the innovation series (white noise) with variance

$$\sigma_\infty^2=E[|a_t|^2]$$

We call g_∞ the ARTF (autoregressive transfer function). It is the same as the transfer function $\Pi(B)$, defined in the authors' equation (12).

To every finite memory $\underline{m}$, one can define finite memory one-step-ahead prediction errors

$$z_t^{\nu,m}=g_m(B)z_t$$

where

$$g_m(z)=1+\alpha_{1,m}z+\ldots+\alpha_{m,m}z^m$$

is the polynomial of degree $\underline{m}$ minimizing

$$\int_{-\pi}^{\pi}|g_m(e^{i\omega})|^2 f(\omega)\,d\omega$$

among all polynomials of degree $\underline{m}$ with constant coefficient equal to 1. The memory $\underline{m}$ mean square prediction error is denoted by

$$\sigma_m^2=E[|z_t^{\nu,m}|^2]$$

The autoregressive spectral approximator

$$f_m(\omega)=\frac{1}{2\pi}\sigma_m^2|g_m(e^{i\omega})|^{-2}$$

may be shown to converge to $f(\omega)$ as $\underline{m}$ tends to ∞, as does σ_m^2 to σ_∞^2 and $g_{\underline{m}}(\underline{z})$ to $g_\infty(\underline{z})$.

Estimators of these quantities from a finite sample $\{z_t, \underline{t}=1, \ldots, \underline{T}\}$ can be constructed as follows: Define

1. The sample spectral density

$$f_T(\omega)=\frac{1}{2\pi T}\left|\sum_{t=1}^{T} z_t e^{it\omega}\right|^2$$

2. Sample covariances

$$R_T(v)=\int_{-\pi}^{\pi} e^{iv\omega}f_T(\omega)\,d\omega$$

3. Sample order m autoregressive coefficients $\hat\alpha_{j,m}$, $j=1, \ldots, m$, as the solution of the normal (or Yule-Walker) equations

$$\sum_{j=1}^{m}\hat\alpha_{j,m}R_T(k-j)=-R_T(k),\ k=1,\ldots,m$$

4. Sample memory m mean square prediction error

$$\hat\sigma_m^2=R_T(0)+\sum_{j=1}^{m}\hat\alpha_{j,m}R_T(j)$$

5. Sample order m autoregressive transfer function

$$\hat g_m(z)=1+\hat\alpha_{1,m}z+\ldots+\hat\alpha_{m,m}z^m$$

6. Sample order m autoregressive spectral estimator

$$\hat f_m(\omega)=\frac{1}{2\pi}\hat\sigma_m^2|\hat g_m(e^{i\omega})|^{-2}$$

Finally, we estimate $\sigma_\infty^2, g_\infty$, and $f(\omega)$ by $\hat\sigma_{\hat m}^2$, $\hat g_{\hat m}$ and $\hat f_{\hat m}(\omega)$, respectively, where $\hat m$ is chosen by an order-determination criterion. In [3], I have proposed choosing $\hat m$ as the value of m, minimizing the criterion function CAT (criterion autoregressive transfer function)

$$\mathrm{CAT}(m)=\frac{1}{T}\sum_{j=1}^{m}\hat{\bar\sigma}_j^{-2}-\hat{\bar\sigma}_m^{-2}$$

where

$$\hat{\hat{\sigma}}_m^2 = \left(1 - \frac{m}{T}\right)^{-1} \hat{\sigma}_m^2$$

and

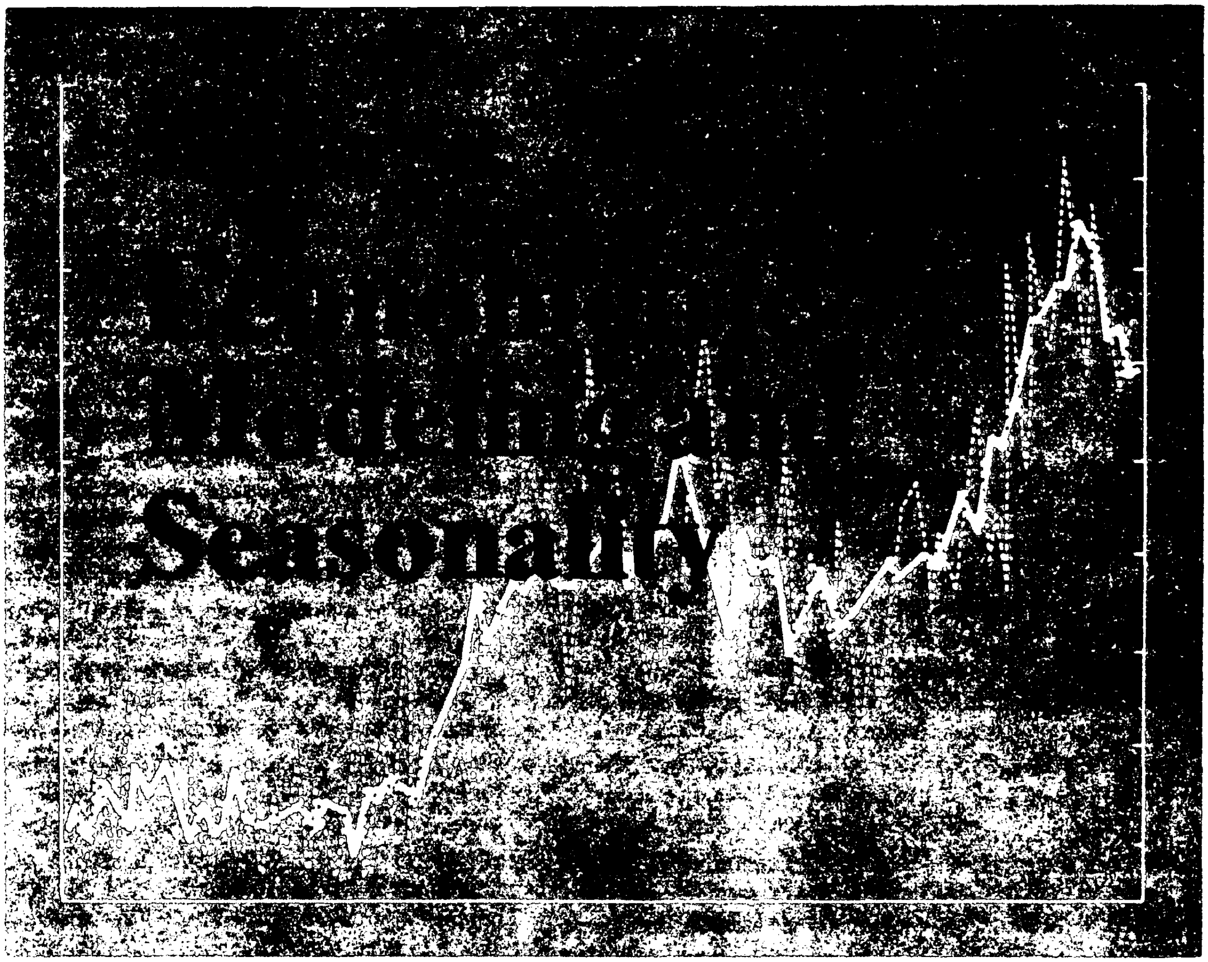
$$\text{CAT}(0) = -1 - (1/T)$$

When $\hat{m}=0$, we say that the time series is white noise.

Having determine the maximum order $\hat{m}$, to help interpret $\hat{g}_{\hat{m}}(z)$, it is useful to use stepwise regression techniques to determine the significantly nonzero autoregressive coefficients. As an example on monthly data, if one had determined that $\hat{m}=13$, it would be of interest to determine whether $\hat{g}_{13}(z)$ were approximately of the form

$$\hat{g}_{13}(z) = (1-\theta_1 z)(1-\theta_{12} z^{12})$$

In my approach to empirical time series analysis, the identification stage is not accomplished chiefly by graphical inspection of the time series and of computed auxiliary sample functions, such as the autocorrelation function, partial autocorrelation function, and spectrum. Rather, the infinite parametric ARTF g_∞ is directly estimated and parsimoniously parametrized.


Seasonality