ESTIMATING STRUCTURAL MODELS OF SEASONALITY

Robert F. Engle
University of California, La Jolla

INTRODUCTION

The intention of this paper is to define and estimate several classes of models of seasonal behavior. A stochastic model will be formulated for each of several unobserved components that then add to give an observed series. The statistical problem is first to choose the appropriate model from the class, then to estimate the unknown parameters of the model, and finally to estimate the values of the unobserved components. When one of the components is a seasonal, the end result is a seasonally adjusted data series.

There are at least three advantages of this approach to seasonal adjustment. First, with an explicit statistical model of the seasonal process, it is possible to calculate the properties of different methods and the variances of individual component estimates. Second, the method of seasonal adjustment will be tailored to the characteristics of the series, and third, it is possible to incorporate additional information about economic trends and cycles, weather, strikes, and holidays which will help distinguish seasonal from nonseasonal behavior and will provide means for automatically correcting for phenomena that are normally treated as outliers.

This approach to seasonal adjustment was introduced by Grether and Nerlove [14], following the similar discussions in Nerlove [23] [24]. The Grether-Nerlove study assumed that not only the true model but even the parameter values were known. Based on these data, the optimum seasonal adjustment filter was calculated as a signal extraction problem.

Four problems make implementation of this procedure formidable. First is the assumption that the model and parameters are all known or that they are valid for every series. Second, the calculations involved in deriving the filter are quite difficult and are different for each model. Third, the technique is limited to stationary processes and will not easily incorporate causal series or nonstationary means. Finally, the filters are generally infinite in length so there is a truncation problem and a difficulty with the initial values.

Pagan [26; 28; 29], in extending this approach, has estimated the parameters of the Grether-Nerlove model and has suggested the desirability of using Kalman filtering techniques to estimate the unobserved components. However, he did not compare these procedures with alternatives and did not adapt his methods to problems with causal variables.

In the second section of this paper, a class of seasonal unobserved component models without causal variables is defined. The third section describes the Kalman filter and its application to extraction of seasonals. The fourth section suggests several estimation procedures which are tried on a simulated series in the fifth section. Here the data are filtered and compared with the X–11 method and with the true nonseasonal component. The sixth section introduces the class of seasonal, unobserved component models with exogenous variables. The estimation and filtering problems are discussed, and, in the seventh section, several unobserved component models are estimated and compared for seasonally unadjusted monthly retail sales.

A CLASS OF UNOBSERVED COMPONENT SEASONAL MODELS WITHOUT CAUSAL VARIABLES

As the class of integrated autoregressive moving average models (ARIMA) have proved very fruitful for modeling univariate time series (see [4]), they form ideal models for unobserved components. A wide class of models can be derived by assuming that a seasonal variable and a nonseasonal variable each follow an ARIMA model of some form with independent innovations, and that their sum plus, perhaps, a white-noise term, is an observable data series. This additive decomposition applies, without loss of generality, to a purely multiplicative model by letting the components be logarithms of the multiplicative factors. Thus far, the distributional properties of the disturbances have not been specified. However, mixed additive-multiplicative models or models that change over the sample period are more difficult. (See Durbin [7] for an example.) If a data series y is assumed to be additively

The author is indebted to Edwin Kuh at the NBER Computer Research Center for research support on the Troll system, to Kent Wall for advice on the Kalman filtering facilities, and to the colleagues at UCSD for many helpful suggestions. Clive W. J. Granger, Allan Andersen, Terence Wales, and David Hendry all made particular contributions. In addition, Ken Wallis, Arnold Zellner, and my discussants, Philip Howrey and Donald Watts, made very helpful comments on the previous draft.

composed of a seasonal component s, a nonseasonal component x and an irregular component ϵ_y, then in terms of the lag operator L, the model can be written

$$y=s+x+\epsilon_y \tag{1}$$

$$(1-L)^{d_x}A_x(L)x=B_x(L)\epsilon_x \tag{2}$$

$$(1-L^r)^{d_s}A_s(L^r)s=B_s(L^r)\epsilon_s \tag{3}$$

Here, the x process is ARIMA (p_x, d_x, q_x), where p_x and q_x are the orders of the lag polynomials A_x and B_x, respectively. The seasonal s also follows an ARIMA process but in terms of r-period lags where r is the period of the seasonal. A particular member of this class can be characterized by the parameters of each of the processes and the relative variances of the epsilons. The defining characteristic of the seasonal is that it depends only upon the r^{th} lags. This turns out not to be the assumption used in the work of Grether and Nerlove, although there is no reason why one could not expand the specification of the seasonal to include a first-order moving average if there were reasonable a priori reasons to expect that the seasonal would have such a term.

This class of unobserved components autoregressive-moving average models (UCARIMA) will be considered the class of structural models. Corresponding to each structural model (which has unobserved endogenous variables) is a reduced form model that includes only observable variables. This reduced form model will also be an ARIMA model, which can be easily estimated using familiar techniques. However, the statistical problem is to identify the form of the structural relations and to identify and estimate the parameters of these relations. These parameters are required for the construction of the filter that will ultimately estimate the values and distribution of the unobserved components. Unlike the forecasting problem, where only the reduced form parameters are required, the seasonal adjustment process depends upon estimation of the structure.

The reduced form corresponding to the structural relations (1)–(3) is derived by premultiplying the first equation by the operators on the left of equations (2) and (3) and substituting to obtain

$$\begin{aligned}(1-L)^d_x(1-L^r)^d_s A_x(L)A_s(L^r)y=&(1-L)^d_x A_x(L)B_s(L^r)\epsilon_s\\&+(1-L^r)^d_s A_s(L^r)B_x(L)\epsilon_x\\&+(1-L)^d_x(1-L^r)^d_x A_x(L)A_s(L^r)\epsilon_y\end{aligned} \tag{4}$$

The error term of this model is a very complicated function of unknown parameters, which has been called by Pagan [25] a composite error term. However, it has the property that after a certain number, $q_{\max}$, all the autocorrelations will be identically zero. Anderson [3] has proven that any process with this characteristic has an invertible moving average representation of order $q_{\max}$. For the UCARIMA class described here, $q_{\max}$ is given by

$$\begin{aligned}q_{\max}=\max(&d_x+p_x+q_s r,\ d_s r+p_s r+q_x,\\&d_x+d_s r+p_x+p_s r)\end{aligned} \tag{5}$$

as long as all of the variances are nonzero. If one of the variances is zero, then its term is merely eliminated from the maximum in obvious fashion.

Equation (4) is, therefore, a seasonal ARIMA model with differencing of orders d_x and d_s, of the nonseasonal and seasonal, respectively, multiplicative autoregressive errors of order p_x and p_s, and a moving average of order $q_{\max}$. In Box and Jenkins' notation, this is $p=p_x$, $P=p_s$, $d=d_x$, $D=d_s$, $q=q_{\max}$, $Q=0$. It is important to notice that, although the autoregressive part is a multiplicative model, the moving average part is not a multiplicative model of seasonal and nonseasonal parts. The normal multiplicative constraint is not true.

How general is the class of UCARIMA models, or, in other words, is there at least one UCARIMA model corresponding to any observable ARIMA process? If the variances of the innovations are allowed to be zero, then there always exists trivially a member of UCARIMA corresponding to an ARIMA model. If, however, the variances are restricted to be nonzero, there are many ARIMA models without a structural counterpart. In particular, any model with moving average order less than the auto-regressive order would not have an UCARIMA counterpart. The appearance of such models in empirical work suggests that either, through the search for parsimonious models, some nonzero moving average coefficients are set to zero or that the variances of some components may be zero. Alternatively, perhaps other models, such as the causal models (to be discussed later in this paper) are the true generating equations to which the ARIMA model is merely a good approximation.

A second question is whether there may be more than one structural relation corresponding to a particular reduced form model, or, in other words, is the structural model always identified if the reduced form is known? There are many cases where the structural model is underidentified, either through the parameters not being uniquely derivable from the reduced form parameters or that even the form of the structural model is not identified.

Underidentified models are considered by both Pierce [27] and Box, Hillmer, and Tiao [5]. In each case, only one structural formulation would give the observed reduced form, however, several possible parameter values were consistent with this model. Each employed the principle of minimum variance for the seasonal component to identify this parameter.

When ARIMA models of order greater than one or models with three components are used, it is quite common that the reduced form is overidentified. In this case, there are parameter restrictions among the reduced form parameters that could be imposed. If these overidentifying restrictions are imposed, generally improved estimates of the reduced form will be achieved, and a unique structural model can be obtained from the reduced form

parameters. If the restrictions are not imposed, the estimation is less efficient, and the equations relating the reduced form coefficients to the structural parameters will, in general, be inconsistent, leading to no solution. By ignoring some of these equations, of course, an estimate, or several estimates, can be obtained.

In this paper, two overidentified models will be considered. In each case, the estimation problems imposing the restrictions are far greater than those without these restrictions. Furthermore, some of the equations relating the reduced form coefficients to the structural coefficients are far more complicated than the others. Thus, it is natural to ask what loss in efficiency results from not imposing these restrictions and ignoring the complicated transformation equations.

KALMAN FILTERING

Although the primary purpose of this paper is the estimation of the structural parameters, it is useful first to consider the problem of estimating the values of the unobserved components, given the true structural model and values of its parameters. The solution to this problem was initially given by Wiener [34] and Whittle [33] and applied by Grether and Nerlove [14]. Based upon the parameters of the model, the weights of the optimum linear time invariant signal extraction filter can be algebraically calculated. The calculation requires factoring the covariance-generating function that is, in general, a difficult numerical procedure. In addition, the weights, whether one sided or two sided, will, in general, extend to infinity, and, therefore, there are both truncation and initial-value problems.

Pagan [26] has suggested that a computationally simpler way to obtain the Wiener filter is through the use of the Kalman filter. The Kalman filter, originally introduced into engineering by Kalman [16] and Kalman and Bucy [17], is now becoming familiar in economics through work by Taylor [32], Chow [6], and many others. It provides a set of recursive formulas which calculate the mean and variance of the unobserved components at each time conditional on a particular information set. If the information set includes all past and current data on the observable variables, then this is the filtering problem. If current data are not included, it is a forecasting problem, and if future data are included, the problem is called a smoothing problem. Additional information, which is generally assumed to be available, is the mean and variance of the initial conditions. Commonly, an informative or a diffuse prior distribution is used when such data are not available.

The filtering equations are written recursively to give the best linear unbiased estimate of the unobserved component, one period ahead, based on the similar estimate of the component this period. In a time invariant problem, the filter weights are not time invariant, because the initial condition introduces transients. These eventually damp out leaving the time invariant Wiener filter. The Kalman equations, however, are more general than the Wiener formulation of the problem, since the known parameters of the problem are permitted to change over time, leading to a nonstationary output time series and, of course, a time-varying filter. This formulation is especially useful when exogenous variables are included in the model.

A brief statement of the Kalman result may be helpful. Without loss of generality, the models in equations (1)–(3) can be rewritten in the state-space formulation. This formulation increases the dimension of the vector of unobserved components, while reducing the problem to a first-order Markov representation plus noise. Letting w_t be a k-dimensional column vector of unobserved states, ϕ a matrix of transition coefficients, and ϵ_t a vector of disturbances with covariance matrix Q, the state equation is written

$$w_t=\phi w_{t-1}+\epsilon_t \tag{6}$$

and the observation equation is written

$$y_t=Hw_t+\eta_t \tag{7}$$

where y_t is a vector (although, in most of these applications, it is a scalar) of observables, H is a matrix of known constants, and η_t is a vector of disturbances with covariance matrix R. The disturbances are assumed to be Gaussian white-noise processes, and ϵ and η are uncorrelated. If the errors are not white noise, then the dynamics of the system can be expanded to include this process as well.

As an example, consider the system

$$\begin{aligned} y&=s+x+\epsilon_y \\ s&=\alpha s_{-4}+\epsilon_s \\ x&=\beta_1 x_{-1}+\beta_2 x_{-2}+\epsilon_x \end{aligned} \tag{8}$$

For this system, the state vector w must be augmented to convert the transition equation into a first-order system. The matrices have the form

$$w=\begin{pmatrix} x \\ x_{-1} \\ s \\ s_{-1} \\ s_{-2} \\ s_{-3} \end{pmatrix},\quad \phi=\begin{pmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$Q=\begin{pmatrix} \sigma^2_{\epsilon_x} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\epsilon_s} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$R=\sigma^2_{\epsilon_y},\ H=(101000) \tag{9}$$

Notice that although there are only two unobserved components, the state vector is six dimensional to adjust

for the higher order dynamics. If the x equation were generalized to include a first-order moving-average term

$$x=\beta_1 x_{-1}+\beta_2 x_{-2}+\epsilon_x+\gamma\epsilon_{x_{-1}}$$

then the first three rows and columns of the w, ϕ, and Q matrices would become

$$w=\begin{pmatrix} x \\ x_v \\ \epsilon_x \end{pmatrix}, \phi=\begin{pmatrix} \beta_1 & \beta_2 & \gamma \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, Q=\begin{pmatrix} \sigma^2_{\epsilon_x} & 0 & \sigma^2_{\epsilon_x} \\ 0 & 0 & 0 \\ \sigma^2_{\epsilon} & 0 & \sigma^2_{\epsilon_x} \end{pmatrix}$$

In both these examples, the Q matrix is singular, and, in the last, the ϕ matrix is also singular.

The Kalman filter calculates the conditional expectation and variance of w_t given the Gaussian data $y_1 \ldots y_t$, initial conditions w_o and V_o and the parameters of (6) and (7). This is the optimum estimate in terms of minimizing mean-squared error and will be linear in the data. Denoting the conditional mean and variance by $w_{t|t}$ and $V_{t|t}$, the essential simplification of the Kalman filtering equations is to recognize that the optimal estimate at time t depends only on the optimal estimate at time $t-1$, $(w_{t-1|t-1})$ and the new data y_t. The equations have been derived in many sources;[1] thus, only an intuitive explanation for the procedure will be given here.

The proof, however, can be easily motivated by consideration of the joint density of w_t and y_t, conditional on y_1, $\ldots$, y_{t-1}. From normality, the expectation of w_t, conditional on y_t (and y_1, $\ldots$, y_{t-1}), is a linear function of the means of w_t and the deviation of y_t from its mean.

That is

$$w_{t|t}=w_{t|t-1}+K_t(y_t-y_{t|t-1})$$

where K_t is a matrix called the Kalman gain, which depends upon the variances and covariances of the random variables. As all variance matrices are known at the beginning, K_t does not depend upon the data; however, it is dependent upon t, and, therefore, the Kalman filter is a time-varying linear filter.

The equations that determine K_t as well as $w_{t|t}$ and $V_{t|t}$ are—

1. $w_{t|t-1}=\phi w_{t-1|t-1}$
2. $w_{t|t}=w_{t|t-1}+K_t(y_t-Hw_{t|t-1})$
3. $V_{t|t-1}=\phi V_{t-1|t-1}\phi'+Q$
4. $V_{t|t}=V_{t|t-1}-V_{t|t-1}H'(HV_{t|t-1}H'+R)^{-1}HV_{t|t-1}$
5. $K_t=V_{t|t-1}H'(HV_{t|t-1}H'+R)^{-1}$ (10)

For many applications in seasonal adjustment, two-sided filters are desired, since historical data must also be seasonally adjusted and it might be sensible to use subsequent information. It is easy to include a fixed number of future observations in each state estimate. In the example of equation (9), if one wished to calculate $w_{t|t+1}$ so that one future observation would be used in calculating the current estimate, the H matrix could merely be redefined as $H=(0, 1, 0, 1, 0, 0)$. Here, there is not even an increase in dimension of the state vector; however, in general, there would be, and, for long fixed leads, this could be very substantial.

The best estimate, based on all the data, is the conditional expectation of the state, given $y_1, \ldots, y_t$. This optimal smoothing problem has well-known solutions (see [8] for a list of different solutions), but Fraser and Potter [11] and Mehra [18] were the first to note a clever formulation of this result, which is implemented in Cooley and Wall [21]. The estimate $w_{t|t}$ could be combined with the estimate of a reverse Kalman filter that uses only future data. The two estimates are Gaussian, and, because the innovations are white noise, they are independent. These estimates can be optimally combined, based on the variances to yield $w_{t|T}$. This estimator is demonstrated to be numerically equal to the optimal smoother. Because the variances are large at the beginning of the sample period, the reverse filter will initially receive most importance, while at the end of the period, the forward estimates will be chosen. This set of smoothed estimates is a two-sided filter, which optimally wraps itself up at each end of the sample period.

A second extension of the Kalman filter, beyond the formulation in (10), is its application to time-varying systems. The recursive formulas can be used equally if ϕ and H, and even Q and R, have time subscripts. In each place, the current matrix is used. This feature fully sets the Kalman filter apart from the signal extraction problem, which cannot deal easily with any type of nonstationarity. In particular, when exogenous variables are introduced in the sixth section, a time-varying ϕ matrix will be needed, since some of its elements will be the data on the exogenous variables.

ESTIMATION OF UCARIMA MODELS

The estimation of unobserved component autoregressive moving average models is a very complicated process. Several approaches have been suggested that might be separated into full and limited information methods. The most direct approach, following Schweppe [31] and Sarris [20], is to maximize the likelihood under Gaussian assumptions. A direct way to evaluate the likelihood for any set of parameters is available from the Kalman filter output. The log likelihood L is given by Schweppe [31]

$$L=\text{const}+\sum_{t=1}^{T}(\log\det P_t^{-1}-\mu_t' P_t^{-1}\mu_t)$$

$$\mu_t=y_t-Hw_{t|t-1}$$

$$P_t=HV_{t|t-1}H'+R \qquad (11)$$

where μ_t is the innovation at time t and P_t is its covariance matrix.

[1] See, e.g., [6].

The log likelihood can be maximized, using a variety of algorithms. Pagan [28] has applied this technique with a variety of nonlinear optimization methods and finds mixed results, with some preference for a modified Gauss-Newton method but also with satisfactory convergence with the Davidon-Fletcher-Powell algorithm. When the maximum is reached, the estimate is full-information-maximum likelihood, FIML.

An alternative approach, also suggested by Pagan [25], is to begin with the reduced form ARIMA version of the model in equation (4). If all the constraints implicit in the formulation were imposed, the likelihood function would be identical to that in (11), except, perhaps, for the treatment of starting values that could introduce a Jacobian, depending on the particular assumptions used. It might be that some constraints should sensibly be ignored for computational convenience, since only efficiency and not consistency would be impaired. As mentioned previously, the constraints on the moving average parameters are very complicated, and there would be substantial savings on computation from ignoring them. Zero restrictions in the moving averages can, however, easily be imposed, although, generally, these would not be sufficient to identify a unique structure. The suggestion made here is to estimate the moving average terms without constraints but draw the parameter estimates from the autoregressive coefficients. The estimates are, therefore, limited information estimates, since some prior restrictions are not imposed. Pagan's results from the relaxation of the constraints differed little from his full maximum likelihood estimates. Although this could be a result of his particular problems, one would expect, with a large sample, that the estimates would differ little. After all, if they differed substantially, one might reject the specification.

AN EXAMPLE

In an effort to compare various methods on the same data with a known-true decomposition, a quarterly unobserved-components time series was artificially generated, using the following model:

$$y=x+s+\epsilon_y \qquad \alpha=0.95\ \sigma^2_{\epsilon_y}=0.8\ \ \sigma^2_y=29.3$$

$$s=\alpha s_{-4}+\epsilon_s \qquad \beta_1=0.75\ \sigma^2_{\epsilon_s}=0.31\ \sigma^2_s=4.37$$

$$x=\beta_1 x_{-1}+\beta_2 x_{-2}+\epsilon_x,\ \beta_2=0.2\ \ \sigma^2_{\epsilon_x}=2.45\ \sigma^2_x=23.4 \quad (12)$$

where each of the random variables is independent normal[2] with observations from 1950 to 1974. An initial seasonal pattern, with a fourth quarter peak, was used for start values. All calculations were performed on the Troll computer system.

The Kalman filter was used to estimate the parameter values of the transition matrix. The parameter estimates were found in five iterations to be

$$\alpha=0.99,\ \beta_1=0.69,\ \beta_2=0.28$$

corresponding to 0.95, 0.75 and 0.2, respectively. These estimates, however, used the true values to start the nonlinear optimization and used true relative variances. Both of these will be unavailable for real data. On the other hand, the initial state, w_0, in this application, was estimated as a parameter vector, assuming a diffuse prior. In an actual application, one could presumably do better.

Two other estimation procedures were used. The Kalman filter maximum likelihood procedure was used to estimate the two relative variances, as well as the three parameters. For this set of data, the maximization algorithm did not converge, since it drove the seasonal variance to zero as the seasonal autoregressive coefficient went to one. The difficulty in estimating the relative variance in a mixture of normal distributions is well known, since the likelihood function has a singularity where the relative variance goes to zero; this appears particularly in economics in the switching regression problem. Here, the solution usually proposed is to assume that the relative variance is known or, at least, to bound it away from zero. Thus, the assumptions in the estimates of the previous paragraphs seem appropriate.

An alternative solution to this divergence would be to bound the autoregressive parameter away from one. This is sometimes done by including a Jacobian term in the likelihood function, which picks up various assumptions concerning the behavior of the initial values. (See, e.g., [8].)

A second set of estimates were calculated using the ARIMA formulation of equation (4). The model in this example has a multiplicative seasonal autoregressive component with the nonseasonal of order 2 and the seasonal of order 1. The moving average portion has coefficients to order 6.[3] This process was estimated using the ESP version of Charles Nelson's time series package. The results and the associated standard errors are shown in table 1.

Notice that the seasonal *ar* parameter is very close to the true value and the nonseasonal *ar* terms, although not very close, are within two standard deviations of the true values. The fitted model has both autoregressive roots positive, while the true model has a large positive root and a small negative one. In both cases, the spectrum has a sharp peak at low frequencies, but, in the true model, it also has a small rise at the high frequency end of the spectrum. The moving average parameters are generally significant with the first and fourth as the largest, as would be expected. The chi-squared statistic for the test

[2] The random numbers were computed by the SNORM subroutine of the TROLL system, which essentially uses a polar transformation of uniform variates to obtain normal random variables. The uniform numbers are obtained from a linear bicongruential generator that truncates a set of large numbers and then uses a second set to randomly shuffle the first, thereby eliminating all possibilities of serial dependence.

[3] Ken Wallis has pointed out that the third-order term is not zero, as I assumed in an earlier draft.

of whiteness of residuals is abnormally low, suggesting overfitting of the model. Because there are several constraints overlooked in this estimation, such a result is to be expected.

The means of the unobserved component series were estimated for each of three sets of parameter values, using each of two estimation assumptions. Filtered estimates, which use only past and current data[4] and smoothed estimates that optimally use the entire data set, were used with the true coefficients (true Φ) with the Φ matrix estimated by the Kalman filter that provides full information estimates (FIML Φ) and with the coefficients estimated by Box-Jenkins methods that are limited information estimates, since some parameter restrictions are ignored (BJ Φ). In each of these six cases, the estimated seasonal component was subtracted from the original data to obtain a seasonally adjusted series. These series were compared with the known seasonally adjusted series and with the adjusted series produced by the Census X–11 program in an additive mode with no provision for the exclusion of outliers.

For each of the Kalman filter estimates, the initial values were estimated using Rosenberg's [30] algorithm. This basically estimates the initial state by assuming a diffuse prior on the mean and then using all the data to form a posterior. The mean and variance of the posterior are then used to initiate the recursive Kalman equations.

A casual inspection of the output suggested that the X–11 method outperformed all alternatives. However, more specific comparisons do not invariably uphold this result. In table 2, the X–11 output has nearly the highest correlation with the true adjusted series when the whole sample period is considered; however, when the first decade is eliminated, all are about the same. This points out a second characteristic of the output. The Kalman filtered estimates performed very poorly at the beginning of the sample period. This result is traceable directly to the estimation of the initial state. These estimates were invariably far from the true values and were very sensitive to parameter changes. Furthermore, they have a substantial effect on the estimates for many periods. A clear recommendation of this research will be that more informative priors be used for the initial state.

When the deviations from the true adjusted series are decomposed into bias, standard deviation and root mean-squared error, more substantial differences between the estimators appear. The FIML estimates have very substantial biases that, in turn, dominate the *rms* error figures. Presumably, these are due to the seasonal *ar* coefficient which is so close to unity that a bias in the initial state estimate never disappears. In almost all cases, the biases are significantly different from zero, which is rather surprising in that, at least, the true Φ cases should be unbiased.

In practice, one would probably be less concerned with bias than with variance. In seasonal adjustment, it would be reasonable to impose the mean of the unadjusted series on the adjusted series and, therefore, the bias would be corrected. In terms of the standard deviations, several features are prominent. In each case, the smoothed estimate is superior to the filtered, and, in each case, the results, excluding the first decade, are somewhat better than those of the whole sample period. The ranking of the estimators is, however, rather surprising. The Box-Jenkins estimator is generally better than either the true Φ or the FIML Φ, and, between the latter two, the results are ambiguous. In the standard forecasting problem, where the data being forecast are not used in the estimation, one would expect to do better with the true parameters. This is presumably the expectation here, although a case can be made by analogy with forecasting inside the sample period to expect the fitted model to do better.

In terms of standard deviation around the true value, the X–11 performs quite well in this competition. Although it is never the best, it is superior to all the filtered estimates and, at least across the whole sample period, superior to two of the three smoothed estimates, including the one based upon the true parameters. In terms of *rms* error, it is the second best in both sample periods and far superior to several of the other candidates. This suggests that, even though the X–11 is not specifically designed for the series in this example, it does almost as well as the more specific procedures.

In table 3, these estimates are examined to see if a linear combination of the X–11 and another process would provide a superior method of seasonal adjustment to either separately. This criterion has been used by Granger [12] and Nelson [22]. When the whole sample period is used, only the Box-Jenkins estimates contribute to the X–11. When the first decade is eliminated, the importance of the X–11 diminishes sharply and has negative or insignificant weight in each of the smoothed estimates.

The conclusion from this example[5] is that the various methods do relatively equivalently in the latter portion of the sample period. In the early portions, the transients give substantially increased variances to the filtered and smoothed estimates, as well as some appearance of bias. The smoothed estimates perform slightly better, which is not surprising since they use more information, and the methods with estimated parameters appear to perform as well as, if not better than, those with the true values. The Kalman filter does appear to provide a reliable method of seasonal adjustment, at least in terms of variance, but great attention should be given to finding better methods of beginning the procedure. This might substantially improve the estimation capabilities of the filter and allow more information for model identification and diagnostic checking. The nearly comparable performance of X–11,

[4] As will be mentioned, the entire data set is used to estimate the initial state, but, thereafter, only current and past data are used.

[5] In an earlier draft, coherence plots for each series with the true adjusted series were included but gave little grounds for comparison. The causes for this failure are pointed out in the comments by Donald Watts.

however, casts doubt on the extent of the possible improvement. All of these results, of course, are for one realization from one model and are, therefore, only suggestions for further investigation.

MODELS WITH EXOGENOUS VARIABLES

Many of the difficult problems in seasonal analysis have to do with the introduction of additional information about strikes, weather, trading days, holidays, model change dates, etc., as well as information about business cycles, inflation, and other causal influences for a particular series. In order to adapt the analysis of the previous sections of this paper to these fundamentally nonstationary influences, it is only necessary to expand the unobserved component models to include exogenous determinants and, then again, to estimate and infer seasonal patterns. The required tools are, however, somewhat different, especially in the estimation phase.

As an example of the type of problem, consider the seasonal adjustment of fuel oil consumption. Consumption in the winter of 1974 fell below its normal winter peak and, therefore, would appear as a decline in the seasonally adjusted series, suggesting a very large response to the sharp increase in fuel prices. On the other hand, the winter was particularly warm, and, therefore, the seasonal peak was not as high as usual, so perhaps there was very little price responsiveness. Treating weather as a causal variable would help to discriminate between these two possibilities.

This class of models is generated by using transfer function models for each of the unobserved components. Letting Z_x and Z_s be matrices of exogenous variables that determine x and s, equations (1) to (3) can be replaced by

$$y=s+x+\epsilon_y \tag{13}$$

$$A_x(L)x=C_x(L)Z_x+B_x(L)\epsilon_x \tag{14}$$

$$A_s(L^r)s=C_s(L)\,Z_s+B_s(L^r)\epsilon_s \tag{15}$$

where C_x and C_s are matrices of lag polynomials and now B_x and B_s must be interpreted as rational lag polynomials in order to maintain full generality. Differencing operators could still be used to induce stationarity; this is usually not necessary, since the exogenous variables will pick up nonstationarity, leaving the residuals as stationary.

Again, this set of structural equations can be put in the form of a single, more complicated, reduced-form transfer function model in the observed variable y. Premultiplying by the autoregressive operators, this becomes

$$A_x(L)A_s(L^r)y=A_x(L)C_s(L)Z_s+A_s(L^r)C_x(L)Z_x$$
$$+A_x(L)B_s(L^r)\epsilon_s+A_s(L^r)B_x(L)\epsilon_x+A_x(L)A_s(L^r)\epsilon_y \tag{16}$$

This transfer function has both sets of exogenous variables in it and has constraints implicit between the autoregressive coefficients and the transfer function coefficients. The error term is now an ARMA process, since the B's were postulated to be rational functions of L, and there are likely to be constraints between the coefficients of this process and the other parameters of the model.

The estimation problem for this model might appear to be extremely complicated, and, indeed, it is if all the information is to be used efficiently. However, adopting the strategy that it may be reasonable to ignore some constraints, equation (16) can be thought of as a simple dynamic regression problem with a complicated error structure and some nonlinear constraints between regression coefficients.

Methods for estimating regression models with lagged dependent variables and stationary errors were originally developed by Hannan [15] and given time domain interpretation by Amemiya and Fuller [1]. The procedure amounts to estimating the spectrum of the disturbances and then transforming all the data to find a regression with white-noise disturbances. To bypass the complications of the lagged dependent variable, this procedure is applied to the equation obtained when the lagged dependent variable is solved out, leaving an infinite lag distribution in the exogenous variables. More recently, Espasa [10] and Engle [9] have pointed out that maximum likelihood estimates of the lagged dependent variable model can be obtained directly by iterating until the coefficient estimates generate a residual spectrum, which, in turn, generates the same coefficient estimates. This estimate will be exactly maximum likelihood for some assumptions on the initial missing observations and only asymptotically maximum likelihood otherwise. Presumably, a consistent starting guess is not necessary, since any convergence of the iteration is a solution to the likelihood equations. However, it is more likely that the iterations will converge to the appropriate solution if a consistent start is used.

Having ignored the constraints on the nature of the error process, a reasonable second step consists of testing whether these constraints hold in the data. This can be done in many ways. The procedure, followed here, is rather ad hoc and consists of examining the properties of the untransformed residuals to see if they follow the stochastic process implied by the structural model. A more attractive way of formulating this procedure[6] would be to solve from the autoregressive parameters for the implied spectrum of the disturbances. The residuals transformed by this constrained estimate of their spectrum should be white.

This procedure is to estimate the model without imposing constraints on the disturbance process and then test those constraints. If the test fails, the adequacy of the model is questioned, and a new structural relation must be sought. The end result should be the identification of the structural model and estimation of its parameters.

[6] This alternative was stimulated by Watts' comments.

This estimation and testing procedure is less than fully efficient since a priori information is ignored. It is, again, a limited information procedure. As before, there is surely a full-information maximum likelihood estimation procedure for this problem. One potential method of calculating this estimator is to use the Kalman filter to evaluate the likelihood and then to maximize it with respect to all the parameters. However, at this point, there are far more parameters because of the presence of the exogenous variables with their separate lag structures, and, therefore, one might hesitate to maximize such a complicated function in so many dimensions. Alternatively, Pagan [25] has directly calculated the constraints but finds that they help little with his parameter estimates.

Once the identification and estimation steps have been completed, the model can be used to estimate the means of the unobserved components. Again, this can simply be done by the Kalman filter, although now it must also include the exogenous variables. The presence of these variables in the transition equation can either be treated directly or can be treated as a time-varying ϕ matrix, depending on the computer programs available.

AN EXAMPLE WITH EXOGENOUS VARIABLES

As an example, several models of seasonally unadjusted monthly retail sales will be analyzed. The series, titled "U.S. Total, All Stores," series 0A0100, was for 1953 through the first half of 1975. It was deflated by the trading days of the month, including Saturdays and excluding all major holidays. This was then logged, since the seasonal appeared clearly to be multiplicative.

Two models were postulated and fitted, neither of which appears fully adequate in view of the residuals, but they suggest a range of possibilities. The first assumes that the seasonal has a deterministic portion with some spreading of the spectrum around the seasonals. Twelve seasonal dummies, a twelfth-order autoregression, and the timing of Easter were assumed to reduce the seasonal component to white noise. The seasonal pattern is so regular that a purely nondeterministic model, such as that simulated for the previous example, did not appear appropriate. The nonseasonal component was assumed to depend upon a constant and the log of monthly personal income in current dollars (seasonally adjusted). If the series were not seasonally adjusted, a long moving average or computation of a permanent income measure would be appropriate. In this case, the results, using current income or permanent income, were indistinguishable.

The model is, therefore, written

$$\begin{aligned} & y = x + s + \epsilon_y \\ \text{I} \qquad & x = \gamma_o + \gamma_1 \log(PY) + \epsilon_x \\ & s = \alpha s_{-12} + \sum_{j=0}^{10} \delta_j \text{Dec}_{t-j} + \delta_{11}\text{Easter} + \epsilon_s \end{aligned} \tag{17}$$

where Dec is a dummy for the month of December, Easter is a variable which is zero, except in March when it takes on a value between 0 and 12, depending on how many of the 2 weeks of shopping days before Easter occurred in March. The γ_o's are normalized to sum to zero, although this only affects the simulation when they must be separated from y_o.

The regression form of the model is

$$\begin{aligned} y = {} & \alpha y_{-12} + \gamma_1 \log(PY) - \alpha\gamma_1 \log(PY_{-12}) \\ & + \sum_{j=0}^{10} \delta_j \text{Dec}_{t-j} \\ & + \delta_{11}\text{Easter} + \gamma_o + u \\ u = {} & (1 - L^{12})(\epsilon_x + \epsilon_y) + \epsilon_s \end{aligned} \tag{18}$$

There are 15 exogenous variables, a lagged dependent variable, and a nonlinear constraint. The error process is a MA(12), with a negative coefficient that would bias α downwards under ordinary least squares. All autocovariances, except the 12th, should be zero.

The second model assumes that the 12th difference of the seasonal model is white noise, except for the effect of the Easter dummy. The nonseasonal component is, however, assumed to have an AR(1) component. This model is

$$\begin{aligned} & y = x + s + \epsilon_y \\ \text{II} \qquad & x = \beta x_{-1} + \gamma_1 \log(PY) + \gamma_o + \epsilon_x \\ & s = s_{-12} + \delta\text{Easter} + \epsilon_s \end{aligned} \tag{19}$$

that can be written in the regression formulation as

$$\begin{aligned} \Delta_{12} y = {} & \beta\Delta_{12} y_{-1} + \delta\text{Easter} - \beta\delta\,\text{Easter}_{-1} \\ & + \gamma_1\Delta_{12}\log(PY) + \gamma_o + u \\ u = {} & (1 - L^{12})(1 - \beta L)\epsilon_y + (1 - L^{12})\epsilon_x \\ & + (1 - \beta L)\epsilon_s \end{aligned} \tag{20}$$

This is a regression model in the 12th differences of y, with 1 lagged dependent variable, 4 exogenous variables, a nonlinear constraint, and a complicated error term. The error term is 13-order moving average with nonzero autocorrelations only at lags 1, 12, and 13.

Many other models could easily be formulated by taking any acceptable nonseasonal model from the economic literature, where most analysis is performed with seasonally adjusted data, and one of several seasonal models, suggested by Granger [12]. Here, a more complicated model is formulated that consists of the seasonal part of model I and the nonseasonal part of model II. There are consequently autoregressive components from each of the component models. The regression formulation of this model is

$$y=\beta y_{-1}+\alpha y_{-12}-\alpha\beta y_{-13}+\gamma_1\log(PY)$$
$$-\alpha\gamma_1\log(PY_{-12})+\delta_{11}\text{Easter}-\beta\delta_{11}\text{Easter}_{-1}$$

III
$$+\sum_{j=1}^{10}(\delta_j-\beta\delta_{j-1})\,\text{Dec}_{-j}+\delta_o\text{Dec}$$
$$-\beta\delta_{10}\text{Dec}_{-10}+\gamma_o+u$$

$$u=(1-\beta L)\epsilon_s+(1-\alpha L^{12})\epsilon_x$$
$$+(1-\beta L)(1-\alpha L^{12})\epsilon_y \qquad (21)$$

This model now has 17 regressors, 3-lagged dependent variables, many nonlinear constraints, and an error term with a 13-order moving average that has nonzero autocorrelations at 1, 12, and 13 lags.

Several computational methods were used for each of these models, all of which were performed on the Troll computer system. The instrumental variable estimator of Liviatan [18] is consistent, regardless of the error structure. This is easily calculated without the constraints but can also be calculated when the nonlinear constraints are imposed by minimizing the sum of squares in the metric of the projection on the instrument list, as discussed by Amemiya [2] for two-stage least squares.

To achieve a more efficient estimaton, the residual spectrum can be used to correct for the stochastic properties, and this procedure, iterated to convergence, will be maximum likelihood subject to the constraints imposed. Further estimates can, therefore, be computed by iterating from either OLS or INST starting points without the nonlinear constraints or by imposing these constraints within the spectral estimation. The latter was not done in these examples.

The iterative Hannan efficient estimator was calculated using the fast fourier transform with the series extended to 3×2^n for some n. The factor of 3 is important when using monthly data so that spectra are calculated at the exact seasonal frequencies. The iterative procedure calculates the spectrum of the residuals, transforms all the data by taking the fourier transform, dividing by the square root of the spectral estimate, and then taking the inverse fourier transform to obtain a time-domain data vector. The residuals of this process should be white noise if convergence has occurred. If they are not, their spectral estimate is multiplied times the previous estimate and the process repeated. The estimates are said to converge when the scaled log likelihood changes by less than 0.0001.

In table 4, estimates are presented for model I. Six estimates are tabulated that correspond to the various options previously mentioned. The results are, in many cases, quite similar. Of particular interest is the 12th-order autoregressive coefficient. It ranges from 0.2 to 0.42 over the methods. That is the largest variation, but all of these values are small enough that such differences would not substantially affect the behavior of the model. The nonlinear constraint is only roughly satisfied. When it is imposed, the largest values of the autoregressive parameter are coupled with the smallest values of log (PY) and the highest standard error of the regression. One would expect from economic theory that the elasticity of retail expenditure concerning personal income would be one, or maybe slightly less, if the goods are predominantly necessities rather than luxuries. This is precisely observed in the iterated methods, and, altogether, the results are rather encouraging. The main disappointment with these estimates is that the final iterative results differ depending upon the starting point. This is mainly true of the autoregressive coefficient and suggests that the likelihood function is relatively flat regarding this parameter. Perhaps, more efficient algorithms or better criteria for termination would minimize this discrepancy. Generally, the likelihood is slightly larger for the OLS-starting values, and thus, these will be reported subsequently.

In table 5, the iterated Hannan estimates are presented for models II and III. In model II, the first-order autoregression is significant but not very large. Economic theory might suggest that this should be very close to one to construct a measure of permanent income, however, the data support a model that says people consume out of transitory income as well. (This has also been found in other contexts; see [9].) The long run elasticity is about

Table 1. BOX-JENKINS ESTIMATES

$$(1-\beta_1L-\beta_2L^2)(1-\alpha L^4)y+\delta$$
$$=(1-\theta_1L-\theta_2L^2-\theta_3L^3-\theta_4L^4-\theta_5L^5-\theta_6L^6)\epsilon$$

Coefficient	Value	Standard error
a	0.946	0.048
B_1	1.183	.314
B_2	-.296	.329
θ_1	.738	.307
θ_2	-.286	.246
θ_3	-.031	.145
θ_4	.598	.133
θ_5	-.124	.278
θ_6	-.269	.121
δ	-.002	.049

Note: Chi-square (24) = 6.7 with 14 degrees of freedom.

Table 2. COMPARISON OF ESTIMATORS

Measure	X–11	True Φ filtered	True Φ smoothed	FIML Φ filtered	FIML Φ smoothed	BJ Φ filtered	BJ Φ smoothed
	January 1950 to April 1974						
Correlation	0.990	0.920	0.982	0.892	0.987	0.984	0.991
Bias	-.37	1.02	1.29	3.58	4.50	-.57	-.37
Standard deviation	.71	1.96	.97	2.25	.83	.89	.69
Root mean-squared error	.80	2.21	1.61	4.23	4.58	1.06	.78
	January 1960 to April 1974						
Correlation	0.991	0.992	0.992	0.991	0.992	0.990	0.992
Bias	-.40	-.02	.86	2.72	4.22	-.84	-.30
Standard deviation	.73	.87	.72	1.58	.72	.80	.69
Root mean-squared error	.84	.87	1.13	3.14	4.28	1.16	.75

Table 3. REGRESSION COEFFICIENTS IN EXPLANATION OF TRUE ADJUSTED SERIES

Range	Constant	X–11	True Φ filtered	True Φ smoothed	FIML Φ filtered	FIML Φ smoothed	BJ Φ filtered	BJ Φ smoothed
1950 to 1974	0.38	1.01	-0.02	-	-	-	-	-
	[1](4.2)	(25.0)	(-.37)	-	-	-	-	-
	.51	1.08	-	-.09	-	-	-	-
	(2.2)	(8.12)	-	(-.67)	-	-	-	-
	.45	1.01	-	-	-.02	-	-	-
	(2.6)	(29.3)	-	-	(-.57)	-	-	-
	.57	1.04	-	-	-	-.04	-	-
	(.59)	(5.06)	-	-	-	(-.22)	-	-
	.42	.71	-	-	-	-	.29	-
	(6.03)	(8.45)	-	-	-	-	(3.41)	-
	.38	.31	-	-	-	-	-	.71
	(5.60)	(1.68)	-	-	-	-	-	(3.76)
1960 to 1974	.22	.40	.66	-	-	-	-	-
	(2.1)	(2.0)	(3.1)	-	-	-	-	-
	-.98	-.08	-	1.04	-	-	-	-
	(-1.86)	(-.20)	-	(2.62)	-	-	-	-
	-1.35	.51	-	-	.65	-	-	-
	(-2.3)	(3.2)	-	-	(2.9)	-	-	-
	-4.8	-.14	-	-	-	1.10	-	-
	(-2.15)	(-.29)	-	-	-	(2.3)	-	-
	.62	.56	-	-	-	-	.45	-
	(5.15)	(4.21)	-	-	-	-	(3.17)	-
	.33	.09	-	-	-	-	-	.93
	(3.48)	(1.26)	-	-	-	-	-	(2.76)

- Entry represents zero.

[1] Standard errors are shown in parentheses.

Table 4. ESTIMATES OF MODEL I: APRIL 1954 TO JUNE 1975

Variable	OLS	INST	Nonlinear	Nonlinear INST	Iterate from OLS	Iterate from INST
Y_{-12}	[1] 0.38(0.06)	0.22(0.13)	0.42(0.06)	0.40(0.13)	0.30(0.06)	0.23(0.06)
log (PY)	1.22(.08)	1.23(.08)	.83(.006)	.83(.006)	1.00(.11)	1.01(.11)
log (PY_{-12})	-.72(.09)	-.61(.12)	-	-	-.43(.13)	-.37(.13)
Easter	.004(.001)	.004(.001)	.004(.001)	.004(.001)	.003(.0007)	.003(.0007)
December	.09(.01)	.11(.02)	.09(.01)	.09(.02)	.10(.010)	.11(.010)
January	-.10(.01)	-.12(.02)	-.09(.01)	-.10(.02)	-.11(.011)	-.12(.011)
February	-.09(.01)	-.12(.02)	-.09(.01)	-.09(.02)	-.10(.010)	-.11(.010)
March	-.07(.01)	-.08(.02)	-.06(.01)	-.06(.017)	-.08(.009)	-.08(.009)
April	-.06(.01)	-.07(.01)	-.06(.01)	-.06(.012)	-.06(.008)	-.07(.008)
May	-.03(.008)	-.04(.01)	-.03(.008)	-.03(.010)	-.03(.006)	-.04(.006)
June	-.01(.007)	-.02(.008)	-.01(.008)	-.01(.008)	-.02(.005)	-.02(.006)
July	-.05(.009)	-.07(.01)	-.05(.009)	-.05(.01)	-.06(.008)	-.07(.008)
August	-.05(.009)	-.06(.01)	-.05(.009)	-.05(.01)	-.05(.007)	-.06(.007)
September	-.03(.008)	-.03(.009)	-.02(.008)	-.02(.009)	-.03(.006)	-.03(.006)
October	-.04(.008)	-.05(.01)	-.04(.008)	-.04(.01)	-.05(.006)	-.05(.006)
Constant	1.11(.10)	1.37(.23)	.96(.11)	.99(.22)	1.15(.118)	1.27(.12)
Standard Error	.0235	.0239	.0245	.0345	.0200	.0203
Durbin-Watson	1.31	1.24	1.16	1.16	1.99	1.99

- Entry represents zero.

[1] Standard errors are shown in parentheses.

Table 5. **ITERATED ESTIMATES USING OLS STARTING VALUES: APRIL 1954 TO JUNE 1975**

Model II $\Delta_{12}y = 0.29\Delta_{12}y_{-1} + 0.77\Delta_{12}\log(PY) + 0.0005\ \text{Easter}$

[1] (0.06) (0.10) (0.0004)

$+ 0.0001\ \text{Easter}_{-1} - 0.012$

(0.0004) (0.006)

Model III $y = 0.40y_{-1} + 0.26y_{-12} - 0.12y_{-13} + 0.69\log(PY) - 0.31\log(PY_{-12})$

(0.06) (0.06) (0.06) (0.10) (0.10)

$+ 0.003\ \text{Easter} - 0.001\ \text{Easter}_{-1} + 0.09\ \text{Dec} - 0.18\ \text{Jan} - 0.08\ \text{Feb}$

(0.0001) (0.0007) (0.01) (0.02) (0.01)

$- 0.06\ \text{Mar} - 0.05\ \text{Apr} - 0.03\ \text{May} - 0.02\ \text{Jun} - 0.08\ \text{July} - 0.05\ \text{Aug}$

(0.01) (0.008) (0.007) (0.006) (0.010) (0.006)

$- 0.03\ \text{Sept} - 0.06\ \text{Oct} + 0.80$

(0.006) (0.008) (0.14)

SER = 0.0181 , DW = 1.98

Model IV $\Delta_{12}y = 0.37\Delta_{12}y_{-3} + 0.94\Delta_{12}\log(PY) - 0.33\Delta_{12}\log(PY_{-3})$

(0.06) (0.15) (0.15)

$+ 0.003\Delta_{12}\text{Easter} - 0.005$

(0.0005) (0.006)

SER = 0.0208 , DW = 1.99

[1] Standard errors are shown in parentheses.

unity as expected. The constraint on Easter is not supported, but the standard errors are large. Probably, the 12th difference of Easter should be used anyway. The standard error of this model is substantially larger than model I, and, on that ground alone, one might prefer I.

Model III is the same as Model I except with an autoregressive term introduced into the nonseasonal component. The results are basically similar to those found above with significant, but not very large, autoregressive coefficients for both the seasonal and nonseasonal components. The autoregressive and Easter constraints are satisfied almost exactly, while the income constraint is only roughly true. The standard error of this model is substantially lower than the 2 percent of model I.

Part of the diagnostic checking of these models must be examination of the residuals to see if they satisfy the processes assumed for them. For all of these models, there should be a negative autocorrelation coefficient at lag 12, for some a nonzero coefficient at lags 1, 11, and 13, but always zero for the other lags. The most cursory check of the autocorrelations of the raw residuals (the difference between the original data and the fitted values, before transformation to whiten the regression residuals, labeled u) shows this not to be true. In figure 1, the autocorrelations are shown for model I, but they look much the same for the others. There are substantial autocorrelations at 3, 6, and 9 but not for 12. The pattern appears to repeat with diminished amplitude. This suggests a third-order autoregression. Furthermore, the spectrum of the raw residuals has a slight negative seasonal pattern, but with a large positive peak at about 0.35, or roughly 3 months per cycle. Since this is a seasonal harmonic, it seems reasonable to attribute this missing autoregression to the seasonal component.

A third-order autoregression in the twelfth differences of the seasonal, plus the twelfth difference in Easter, are used as the seasonal model with the simple nonseasonal of model I to get model IV. The regression form is, therefore,

$$\text{IV}\qquad \begin{aligned}\Delta_{12}y &= \alpha\Delta_{12}y_{-3}+\gamma_1\Delta_{12}\log(PY)\\ &\quad -\alpha\gamma_1\Delta_{12}\log(PY_{-3})\\ &\quad +\delta\Delta_{12}\text{Easter}+\gamma_0+u\\ u &= (1-L^{12})(1-\alpha L^3)(\epsilon_x+\epsilon_y)+\epsilon_s \qquad (22)\end{aligned}$$

This is now one of the simplest models in that there are only five regressors. There are still a nonlinear constraint and an error process, which has third-, twelfth-, and fifteenth-order moving average terms. The estimates of this model look quite encouraging. The constraint is almost exactly satisfied, and income has nearly a unit elasticity. The standard error is between model I and model II, but the structure is far more parsimonious. The autocorrelation of the residuals looks substantially better with the expected negative term at 12. Various alternative means of respecifying the model to capture the unexplained seasonal part might prove still better.

The technique of model identification is, therefore, an iterative one. The autoregressive parameters are estimated, assuming a general stationary error structure. The residuals are then examined to see if they obey the prescribed process. If not, a respecification of the model is required. The mechanics of this search procedure are yet to be worked out, but this example may be a helpful illustration.

This discussion has suggested several types of unobserved component models with exogenous variables and various estimation methods. The estimation difficulties do not appear particularly great, but some of the specification choices may be difficult. It is quite possible that the seasonally adjusted series would not be very sensitive to the model specification, just as in the simulated series in the fifth section. Only further study of this problem can provide the answer.

CONCLUSION

This paper has proposed a method of seasonal adjustment, based upon specifying and estimating a structural unobserved components model, and then using this model to optimally filter the data to remove the seasonal component. The computational problems in estimating the models using asymptotically efficient procedures are quite severe. In this paper, it has been argued that some constraints on parameters can be relaxed to yield less efficient, but still consistent, structural estimates. In this fashion, model estimation becomes feasible, although the identification problems remain very complicated.

The model is then used to optimally filter the data, using the Kalman version of the Wiener filter. Both a one-sided filter and a two-sided smoother were used on artificial data. Empirically, even when the true model parameters were employed, the X–11 performed nearly as well as the optimal filter. In terms of variance, the differences between results using estimated models and the true model were not great, although one of the estimated models was badly biased. The filtered and smoothed results were generally inferior at the beginning of the sample period, reflecting a badly estimated initial condition. Improvement in the performance would surely be expected from imposing a proper prior on this state.

The relative success of the X–11 in a situation where it should be dominated suggests that, perhaps, the structural approach to seasonal adjustment, based entirely on a single series, should not be recommended. However, when causal variables are introduced, there is no alternative, to the estimation of a structural model and inference about the seasonal characteristics from this model. The second example presents estimation methods and results for a model of monthly retail sales. Here, the timing of holidays and trading days were shown to influence the seasonal component, while personal income influenced the nonseasonal.

In this case, one would expect the Kalman filter to

Figure 1. MODEL I, AUTOCORRELATIONS OF RESIDUALS

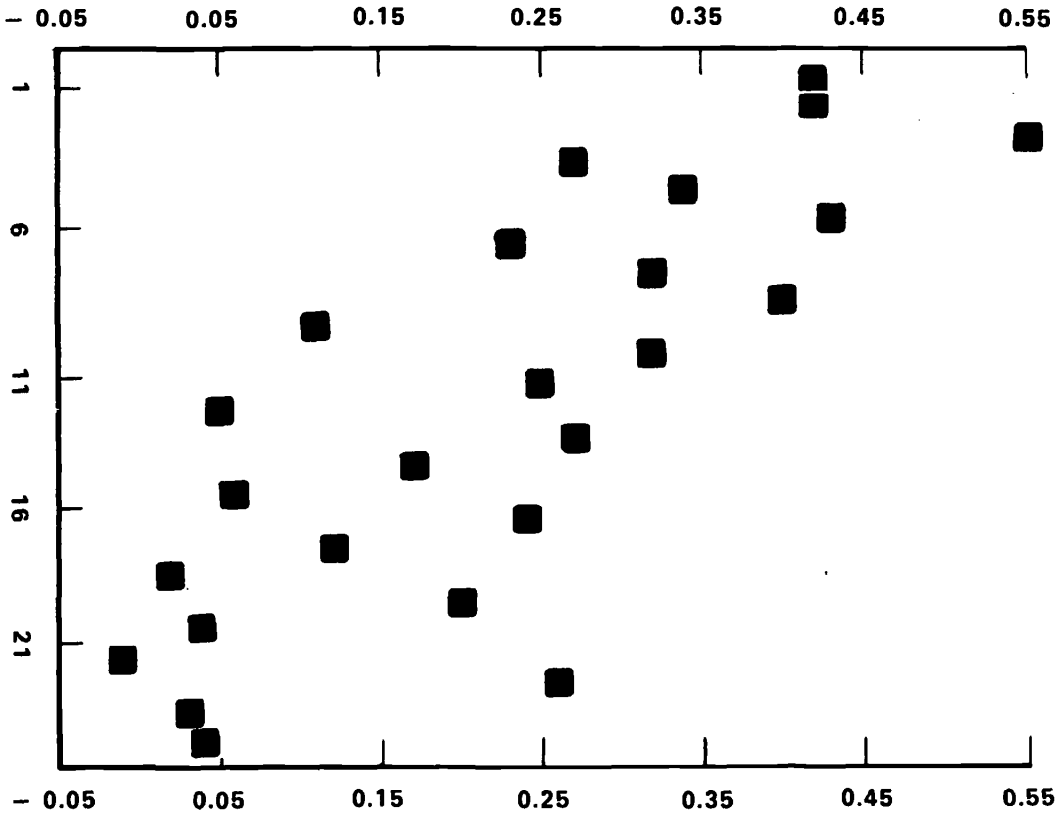
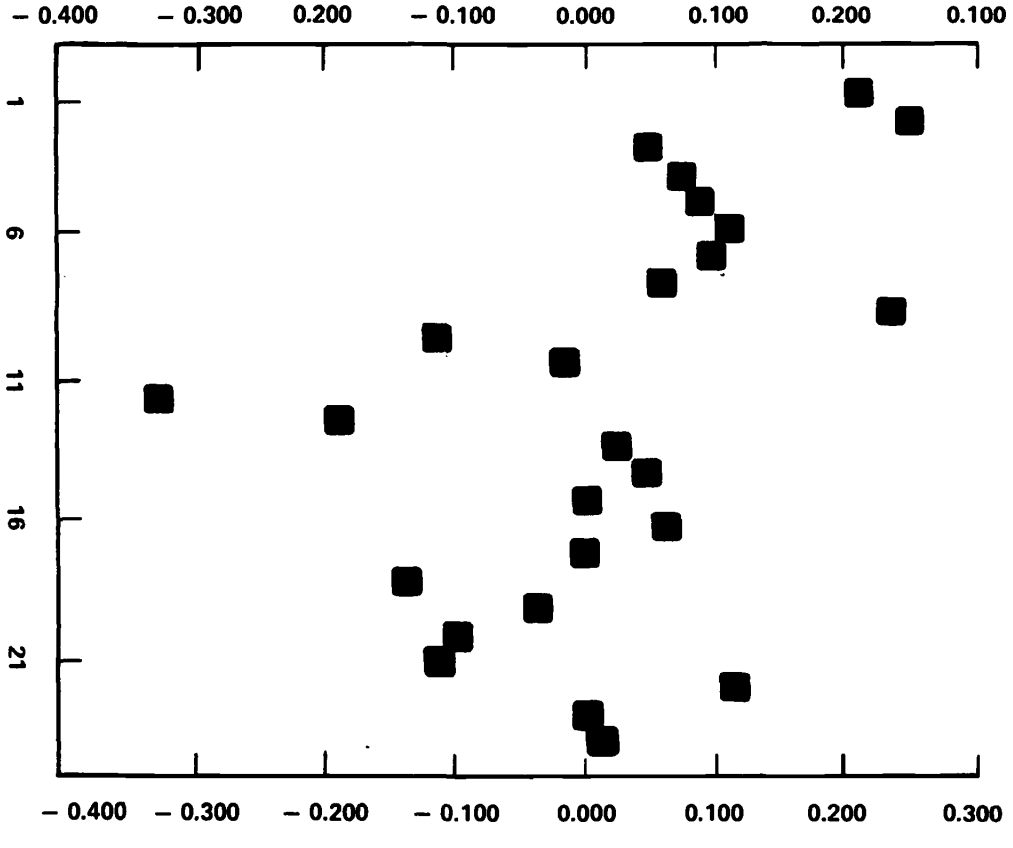


Figure 2. MODEL IV, AUTOCORRELATIONS OF RESIDUALS



more efficiently utilize the information in the causal series and provide superior estimates of the adjusted series than the X–11, which can only, in simplistic fashions, make use of causal information.

REFERENCES

1. Amemiya, T., and Fuller, W. A. "A Comparative Study of Alternative Estimators in a Distributed Lag Model." *Econometrica* 35 (1967): 509–529.

2. ———. "The Nonlinear Limited Information Maximum Likelihood Estimator and the Modified Nonlinear Two-Stage Least-Squares Estimator." Working Paper No. 52. Palo Alto, Calif. Institute for Mathematical Studies in the Social Sciences, Stanford University, 1974.

3. Anderson, T. W. *The Statistical Analysis of Time Series*. New York: John Wiley and Sons, Inc., 1971.

4. Box, G. E. P., and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day, 1976.

5. ———, Hillmer, S. C. and Tiao, G. C. "Analysis and Modelling of Seasonal Time Series." Included in this report.

6. Chow, G. C. *Analysis and Control of Dynamic Economic Systems*. New York: John Wiley and Sons, Inc., 1975.

7. Durbin, J., and Kenny, P. B. "Seasonal Adjustment when the Seasonal Component Behaves Neither Purely Multiplicatively Nor Purely Additively." Included in this report.

8. Engle, R. F. "Band Spectrum Regression." *International Economic Review* 15 (1974): 1–11.

9. ———. "Band Spectrum Regression Methods with Lagged Dependent Variables and Stationary Errors—An Iterated Instrumental Variables Approach," paper presented to Symposium on Current Directions in Applied Time Series Analysis, Tulsa, Okla., 1976.

10. Espasa, A. "The Estimation of the Multiple Regression Model with Stationary Errors and Lagged Endogenous Variables." London School of Economics, 1975. (Unpublished manuscript.)

11. Fraser, D. C., and Potter, J. E. "The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters." *IEEE Transaction on Automatic Control* AS14 (1969): 387–390.

12. Granger, C. W. J. "Seasonality: Causation, Interpretation and Implications." Included in this report.

13. Granger, C. W. J., and Newbold, P. *Forecasting Economic Time Series*. New York: Academic Press, 1977.

14. Grether, D. M., and Nerlove, M. "Some Properties of 'Optimal' Seasonal Adjustment." *Econometrica* 38 (1970): 686–703.

15. Hannan, E. J. "The Estimation of Relationships Involving Distributed Lags." *Econometrica* 35 (1965): 206–224.

16. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* 82D (1960): 33–45.

17. Kalman, R. E., and Bucy, R. S. "New Results in Linear Filtering and Prediction Theory." *Journal of Basic Engineering* 83 (1961): 95–108.

18. Liviatan, N. "Consistent Estimation of Distributed Lags." *International Economic Review* 4 (1963): 44–52.

19. Mehra, R. K. "On a Limiting Procedure in Linear Recursive Estimation Theory." *Journal of Mathematical Analysis and Applications* 26 (1969): 1–8.

20. National Bureau of Economic Research. *A General Algorithm for Simultaneous Estimation of Constant and Randomly-Varying Parameters in Linear Relations*, by A. H. Sarris. Working Paper No. 38. New York: National Bureau of Economic Research, 1974.

21. ———. *A Note on Optimal Smoothing for Time Varying Coefficient Problems*, by T. F. Cooley and K. D. Wall. Working Paper No. 128. New York: National Bureau of Economic Research, 1976.

22. Nelson, C. R. "Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." *American Economic Review* 62 (1972): 902–917.

23. Nerlove, M. "Distributed Lags and Unobserved Components in Economic Time Series." *Ten Economic Studies in the Tradition of Irving Fisher* (W. Fellner et al.). New York: John Wiley and Sons, Inc., 1967.

24. ———. "Lags in Economic Behavior." *Econometrica* 40 (1972): 221–251.

25. Pagan, A. "Efficient Estimation of Models with Composite Disturbance Terms." *Journal of Econometrics* 1 (1973): 329–340.

26. ———. "A Note on the Extraction of Components from Time Series." *Econometrica* 43 (1975): 163–168.

27. Pierce, D. A. "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present." Included in this report.

28. Princeton University, Econometric Research Program. "Estimation of an Evolving Seasonal Pattern as an Application of Stochastically Varying Parameter Regression. Research Memorandum No. 153, by A. Pagan, 1973.

29. ———. "A Note on the Extraction of Components From Time Series." Research Memorandum No. 148, by A. Pagan, 1973.

30. Rosenberg, B. "A Survey of Stochastic Parameter Regression" *Annals of Economic and Social Measurement* 2 (1973).

31. Schweppe, F. C. "Evaluation of Likelihood Function for Gaussian Signals." *IEEE Transactions in Information Theory* 11 (1965): 61–70.

32. Taylor, L. "The Existence of Optimal Distributed Lags." *Review of Economic Studies* 37 (1970): 95–106.

33. Whittle, D. *Prediction and Regulation by Linear Least-Square Methods*. London: The English Universities Press, Ltd., 1963.

34. Wiener, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge and New York: Massachusetts Institute of Technology Press and John Wiley and Sons, Inc. 1950.

COMMENTS ON "ESTIMATING STRUCTURAL MODELS OF SEASONALITY" BY ROBERT F. ENGLE

E. Philip Howrey
University of Michigan

In this paper, Engle investigates a relatively new approach to seasonal adjustment. The distinguishing characteristic of this approach is that an explicit statistical model of seasonal variation is postulated. On the basis of the specified model, optimal methods of parameter estimation, filtering, and smoothing are proposed.

Engle discusses a number of advantages that this model-specific approach to seasonal adjustment has over the more traditional methods of seasonal adjustment that do not depend on a specific model. Since this is a new approach to the problem, it raises a number of important and difficult issues. These comments concentrate primarily on two general problem areas: Criteria for model specification (or what is called identification by Box and Jenkins [3]) and optimal filtering procedures.

MODEL IDENTIFICATION

The first step in Engle's scheme of seasonal adjustment is the specification of a model on which to base subsequent computations. A general class of models is introduced in which the observed series is the sum of three unobserved components: A seasonal, a nonseasonal, and a random irregular time series. Both the seasonal and nonseasonal time series are assumed to be generated by ARIMA processes. Each of these two ARIMA processes is characterized by three parameters: p, the order of the autoregressive part; d, the degree of differencing required to achieve stationarity, and q, the order of the moving average part of the process. Once d is specified, $p+q+1$ additional parameters are needed to complete the specification of each ARIMA process (i.e., p AR parameters, q MA parameters, and the innovation variance of the process). Thus, p_r+q_r+2 parameters are needed to describe the nonseasonal component, p_s+q_s+2 parameters are needed to describe the seasonal component, and one additional parameter describes the irregular component, a total of $p_r+q_r+p_s+q_s+5$ parameters.

In applied work, it is necessary to assign values to (p, d, q) for both the seasonal and nonseasonal components. But, the basic rationale that is given for the introduction of this class of stochastic processes is that ARIMA models have been found to be very useful for modeling univariate time series. It appears then that the theory provides little guidance, if any, in the specification of the appropriate degrees of the AR and MA processes. Thus, it is necessary to utilize the data to determine the length of the lags that are required to yield a model that provides an adequate description of the time series.

Since the modeling aspect is basic in this approach to seasonal adjustment, it is desirable, if not essential, to have some way to determine the appropriate form of the model to be used. A number of identification criteria have been suggested, by time series researchers, to determine or identify the appropriate values of p and q for an observed series. However, little is known about the relative performance of these criteria. In an attempt to determine the ability of several alternative identification criteria to select the correct model, a fairly extensive set of simulation experiments were conducted. The results of a small subset of these experiments are described in this paper to indicate how well these criteria can be expected to perform in actual practice.

In each of the experiments summarized, 100 realizations were obtained from a Guasian ARMA (p, q) process of the form

$$\phi_p(L)\, x_t=\theta_q(L)\, \epsilon_t \qquad (1)$$

where $\phi_p(L)$ is a polynomial of degree p, $\theta_q(L)$ is a polynomial of degree q, and ϵ_t is normally distributed with mean zero and unit variance. Each of the 100 series was then analyzed in an attempt to determine the appropriate values of p and q.

The first step in the analysis of each series is the computation of preliminary estimates of the parameters of the operators ϕ_p and θ_q, using the methods described in Box and Jenkins [3, 201–203] for the nine models corresponding to $p=0$, 1, 2 and $q=0$, 1, 2. These parameter estimates are then used to obtain three statistics that could be used in a straightforward, mechanical way to identify the order of the model. The first of these criteria is similar to the mean-squared prediction error (MSPE) criterion used by Akaike [1]. Modified to handle the moving average case, it is defined as

$$\text{MSPE}=s^2\{1+(p+q)/n\}/\{1-(p+q)/n\} \qquad (2)$$

E. Philip Howrey is a professor of economics at the University of Michigan. *The author is indebted to Beth Anne Hayes for computational assistance. This research was supported, in part, by the National Science Foundation.*

where n designates the number of observations in the series and s^2 is the preliminary estimate of the residual variance as given by Box and Jenkins [3, 203]. A second identification criterion is the unconditional mean-squared error (UMSE) defined by

$$\text{UMSE} = \text{SSE}/(n-p-q) \tag{3}$$

where SSE is the unconditional sum of squared residuals as defined by Box and Jenkins [3, 215–220]. A third criterion that was investigated is based on an approximation due to Grenander and Rosenblatt [4, 267–270]. The absolute efficiency of the predictor based on the ARMA (p, q) process is defined by

$$e = \sigma^2/\sigma^2(p, q) \tag{4}$$

where σ^2 is the innovation variance of the process and $\sigma^2(p, q)$ is the prediction-error variance of the ARMA (p, q) model. Let $f(\omega)$ denote the power spectrum of x_t, and let $\hat{f}(\omega)$ denote the power spectrum implied by the ARMA (p, q) model that is used to predict x_t. Grenander and Rosenblatt show that

$$e = \left\{1+\frac{1}{4\pi}\int [1-\hat{f}(\omega)/f(\omega)]^2\, d\omega - \frac{1}{8\pi^2}\left(\int [1-\hat{f}(\omega)/f(\omega)]d\omega\right)^2\right\}^{-1} \tag{5}$$

An estimate of the absolute prediction efficiency of an ARMA (p, q) model is given by a discrete approximation to this expression. Specifically,

$$\hat{e} = \{1+I_1(p, q)-I_2(p, q)\}^{-1} \tag{6}$$

where

$$I_1(p, q) = \frac{1}{4m}\sum_{s=-m}^{m-1}\{\tilde{f}(\pi s/m)/\hat{f}(\pi s/m) - 1\}^2 \tag{7}$$

$$I_2(p, q) = \frac{1}{2}\left\{\frac{1}{2m}\sum_{s=-m}^{m-1}\tilde{f}(\pi s/m)/\hat{f}(\pi s/m) - 1\right\}^2 \tag{8}$$

In these expressions, $\tilde{f}(\omega)$ is the direct, Parzen-windowed estimate of the spectrum, and $\hat{f}(\omega)$ is the power spectrum implied by the ARMA (p, q), prediction model, namely

$$\hat{f}(\omega) = |\hat{\phi}(e^{-i\omega})|^{-2}\,|\,\hat{\theta}(e^{i\omega})\,|^2 s^2 \tag{9}$$

Finally, an ad hoc degrees-of-freedom adjustment is introduced to yield

$$\text{PEFF} = \hat{e} - (p+q)/n \tag{10}$$

as an estimate of the prediction efficiency (PEFF) of the ARMA (p, q) process.

The results that were obtained for an ARMA (0,0) process are summarized in table 1. The left-hand panel contains the results for 100 realizations of 30 observations each, and the right-hand panel shows the results for realizations of 100 observations. It is seen that, for $n=30$, the mean-squared prediction-error criterion identifies the correct model 77 times in 100 attempts. The unconditional mean-squared error criterion correctly identifies the (0,0) process 52 times, and the prediction efficiency method is correct 64 times. Similar results are obtained with the MSPE and UMSE criteria for $n=100$. Interestingly enough, PEFF performs more poorly with $n=100$ than on the shorter series. These results indicate that, at best, we would be misled into using a more complicated model than necessary almost 25 percent of the time.

Results are shown in table 2 for simulation experiments, based on (1, 0) and (0, 1) processes for which the unknown parameter is 0.9 in both cases. The most apparent feature

Table 1. IDENTIFICATION RESULTS FOR AN ARMA (0,0) PROCESS

Model	Identification criteria (n = 30)			Identification criteria (n = 100)		
(p,q)	MSPE	UMSE	PEFF	MSPE	UMSE	PEFF
(0,0) . .	77	52	64	72	51	38
(0,1) . .	10	13	23	12	9	9
(0,2) . .	6	11	1	6	6	16
(1,0) . .	-	8	9	1	16	18
(1,1) . .	4	6	-	6	4	2
(1,2) . .	-	1	-	1	2	2
(2,0) . .	3	8	3	-	12	15
(2,1) . .	-	1	-	2	-	-
(2,2) . .	-	-	-	-	-	-

- Entry represents zero.

Table 2. IDENTIFICATION RESULTS FOR ARMA (1,0) AND ARMA (0,1) PROCESSES

Model	Identification criteria (1,0) process			Identification criteria (0,1) process		
(p,q)	MSPE	UMSE	PEFF	MSPE	UMSE	PEFF
(0,0) . .	-	-	-	-	-	-
(0,1) . .	-	-	-	19	24	23
(0,2) . .	-	-	-	4	3	6
(1,0) . .	84	65	54	3	3	3
(1,1) . .	7	12	23	17	19	20
(1,2) . .	8	17	16	9	9	8
(2,0) . .	-	4	6	7	1	2
(2,1) . .	1	1	-	34	33	33
(2,2) . .	-	1	1	7	8	5

- Entry represents zero.

of these results is that the identification criteria perform much better for the AR(1) process than for the MA(1) process. In both cases, however, the model is incorrectly identified a minimum of 15 percent of the time.

As a final illustration of the results that can be expected from the application of these identification criteria, two ARMA (1, 1) runs are summarized. In the left-hand panel of table 3, the AR coefficient is 0.9, and the MA coefficient is 0.5. In the right-hand panel, the AR coefficient is 0.5, and the MA coefficient is 0.9. With these slightly more complicated models, the success rate in identification of the correct model falls to about 50 percent or less.

Table 3. IDENTIFICATION RESULTS FOR ARMA (1,1) PROCESSES

Model (p,q)	Identification criteria (ϕ_1= 0.9;θ_1= 0.5)			Identification criteria (ϕ_1= 0.5;θ_1= 0.9)		
	MSPE	UMSE	PEFF	MSPE	UMSE	PEFF
(0,0) . .	-	-	-	2	2	-
(0,1) . .	17	-	-	27	16	17
(0,2) . .	12	-	1	34	31	36
(1,0) . .	3	7	6	-	1	-
(1,1) . .	44	51	41	12	21	22
(1,2) . .	17	10	8	10	12	8
(2,0) . .	2	23	38	2	4	3
(2,1) . .	1	4	2	6	7	10
(2,2) . .	4	5	4	7	6	4

- Entry represents zero.

These experimental results are not intended to be definitive, but rather suggestive, of the identification problem. Other identification procedures could be employed, such as the residual spectrum analysis, suggested by Engle. Undoubtedly sharper parameter estimates and more accurate model specification could be achieved by proceeding from the preliminary parameter estimates to more refined estimation procedures. Nevertheless, these experimental results do indicate the danger in using mechanistic rules to identify the order of an ARMA process. If the model is at all complicated, there is a danger that the model will not be correctly identified. In view of the large number of series that are currently seasonally adjusted, it is difficult to see how any technique that is not largely mechanical and, therefore, relatively free of individual judgment would be feasible to use on a large scale. Unless better criteria than those used here can be developed, it is difficult to have a great deal of confidence that the model identified is the correct specification and, hence, that it provides a useful basis for seasonal adjustment of the time series.

OPTIMAL SMOOTHING

After a model has been identified and the unknown parameters have been estimated, the model can be used to separate the seasonal and nonseasonal components. The smoothing problem involves the computation of the conditional expectation of the unobserved seasonal component of the time series, given the model and the observed series. The rationale for this approach is quite general. Aström [2], e.g., shows that if (1) the loss function is symmetric and nondecreasing for positive arguments and (2) the conditional distribution of the unobserved component has a unimodal density function that is symmetric about the conditional expectation, then the conditional expectation is optimal in the sense that it minimizes expected loss.

The computational schemes that are proposed to smooth a series to extract the seasonal component usually take the model, including the parameters to be given. If the model is correctly specified and the parameter values are correct, the smoothed series is an optimal estimate of the unobserved component. In actual practice, neither the form of the model nor the parameter values are known with certainty. It is not clear that it is appropriate in the face of model uncertainty to proceed as if the model were correct. It is well known that, in a regression context, the best linear unbiased predictor is obtained using the least squares estimates of the regression coefficients. If a similar result holds for the unobserved components model, the case for using the conventional optimal smoothing procedures would be vindicated.

The strength of the optimal smoothing approach to seasonal adjustment is that, if the model is correct, the seasonal series obtained is indeed optimal. However, if the model is not correct, little can be said about the properties of the resulting series. The results summarized in this paper suggest that identification of the appropriate model is not an easy task; there is a nontrivial chance that an incorrect model will be selected. An issue that needs to be resolved is the sensitivity of optimal smoothing procedures to parameter error. There may be suboptimal procedures that perform quite well over a wide range of processess, whereas the performance of optimal methods may be quite sensitive to model error.

The empirical results, presented in Engle's paper, are of considerable interest in this respect. Two sets of parameter estimates are given for the expository seasonal model: One set of estimates of the structural parameters and another set for the reduced form of the model. The structural parameter estimates differ only slightly from the true values; the reduced form estimates vary more widely from the true parameter values. The summary results indicate that the seasonal pattern obtained from the estimated structural form of the model is quite close to the optimal seasonal pattern, based on the true parameter values. This indicates that the smoothing procedure is not sensitive to small errors in the parameters. It would be very interesting to know if the reduced-form model

produces an acceptable seasonal pattern. If it does, then parameter error of the order of magnitude that is likely to be encountered in practice may not be a serious problem for this approach.

In connection with these expository empirical results, the performance of the X–11 method of seasonal adjustment looks extraordinarily good. The optimal smoothing procedure should have a real advantage in this case, since it is given the correct specification of the model. However, over the full sample period, X–11 performs nearly as well without any information on the form of the model. Even over the last half of the sample, X–11 produces a seasonally adjusted series that is highly correlated with the true adjusted series.

In summary, these comments point toward two problem areas. One is the development of criteria to aid in the identification of models of seasonal variation. The other is the investigation of the sensitivity of optimal smoothing procedures to model misspecification. In view of the likelihood of model misspecification, it is desirable to develop seasonal adjustment methods that are not overly sensitive to model error.

REFERENCES

1. Akaike, H. "Fitting Autoregressive Models for Prediction." *Annals of the Institute of Statistical Mathematics* 21 (1969): 243.

2. Aström, Karl J. *Introduction to Stochastic Control Theory*. New York: Academic Press, 1970.

3. Box, George E. P., and Jenkins, Gwilym M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, Inc., 1970.

4. Grenader, U., and Rosenblatt, M. *Statistical Analysis of Stationary Time Series*. New York: John Wiley and Sons, Inc., 1957.

COMMENTS ON "ESTIMATING STRUCTURAL MODELS OF SEASONALITY" BY ROBERT F. ENGLE

Donald G. Watts
Queen's University

It is Engle's intention, in his paper, ". . . to define and estimate several classes of models of seasonal behaviour." He then attempts to illustrate the models and estimation procedures for them, using two examples.

TWO CLASSES OF MODELS

In the section concerning a class of unobserved component seasonal models without causal variables, Engle discusses standard additive models of the form

$$y=\frac{B_x}{A_x}\epsilon_x+\frac{B_s}{A_s}\epsilon_s+\epsilon_y \qquad (1)^1$$

where A and B are transfer function operators and the ϵ are white-noise processes. Schematically, the model can be represented, as in figure 1.

This model can be rearranged to give a covariance-equivalent model (4), as in figure 2. In this representation, as Engle points out, the relationship between σ_ϵ^2 and the parameters of B and $\sigma_{\epsilon_x}^2, \sigma_{\epsilon_y}^2, \sigma_{\epsilon_s}^2, A_x, A_s, B_x, B_s$ is a complicated nonlinear one, as well as the fact that the transfer function B is not a multiplicative one.

These two models constitute Engle's arsenal of UCARIMA models.

In the section concerning models with exogenous variables, Engle adds causal variables Z_x and Z_s (which may be matrices) to the system in figure 1 to produce the models (18) and (19) with the block diagram representation shown in figure 3. As before, the operators A_x^{-1} and A_s^{-1} may be factored out to give a covariance-equivalent representation (20) with the block diagram shown in figure 4. Again, it is recognized that σ_ϵ^2 and the parameters of B are related in an extremely messy fashion to $\sigma_{\epsilon_x}^2, \sigma_{\epsilon_y}^2, \sigma_{\epsilon_s}^2$ and the parameters of A_x, A_s, B_x, C_x and C_s.

The models portrayed in figures 3 and 4 constitute Engle's collection of additive models with exogenous variables.

Now, the models shown in figures 1 and 2 are certainly standard additive models for seasonal time series, of which economists seem highly enamored, even to the extent of extolling the fact that taking logarithms of a multiplicative series can convert it to an additive one.[2] This seems, to me, to introduce many difficulties that must then be overcome using complicated methods, such as those discussed in this paper. I wonder why economists do not try multiplicative models, with their equally valid justification and vastly simpler and more effective analysis procedures, more often?

Upon reflection, I suppose it is because we like to incorporate the effects of exogenous variables in an additive fashion. One can then argue that if a seasonal component is added when its cause is known, why not add it when its cause is unknown? Accepting additive models, then, it still seems to me that the models shown in figures 3 and 4 are ill conceived, because they impose the same dynamics (A_x^{-1}) on all of the regular exogenous variables Z_x and the same dynamics (A_s^{-1}) on all of the seasonal exogenous variables Z_s. Possibly more appropriate models would be of the forms

$$x=\frac{C_x}{D_x}Z_x+\frac{B_x}{A_x}\epsilon_x$$

and

$$s=\frac{C_s}{D_s}Z_s+\frac{B_s}{A_s}\epsilon_s$$

where each exogenous variable Z_x and Z_s has its own denominator. Cross-multiplying these models would then, of course, yield very high-order autoregressive and moving average polynomials, but this seems, to me, to be the price one has to pay for having a realistic model in which each exogenous variable influences the response in its own particular fashion.

THE ESTIMATION PROCEDURES AND EXAMPLES

Having introduced the additive models, Engle next attempts, in the main body of the paper, to apply some methods of estimation for the models chosen. For the UCARIMA models, he follows Pagan (20) to try to

[1] Equation numbers refer to the equations in Engle's paper.

[2] It is important to recognize that transformation should only be done in order to ensure assumptions about the error structure. That is, if the variance of a series is constant and if the error is normal, transforming to produce an additive model will foul up the variance and the normality, thereby invalidating ordinary least squares estimation procedures.

Figure 1. BLOCK DIAGRAM FOR AN ADDITIVE MODEL

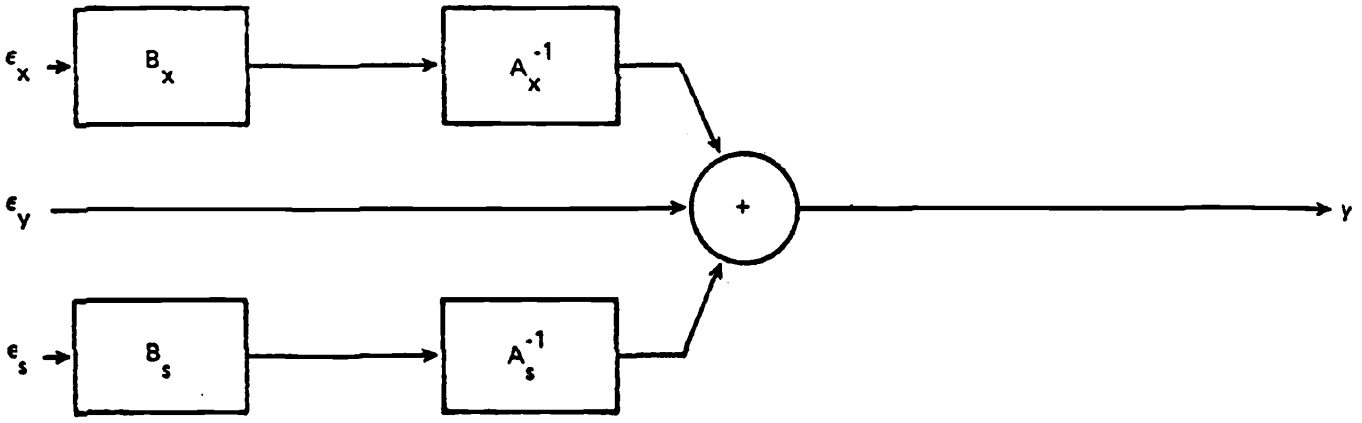


Figure 2. COVARIANCE-EQUIVALENT BLOCK DIAGRAM FOR THE ADDITIVE MODEL

Figure 3. BLOCK DIAGRAM FOR AN ADDITIVE MODEL WITH CAUSAL VARIABLES

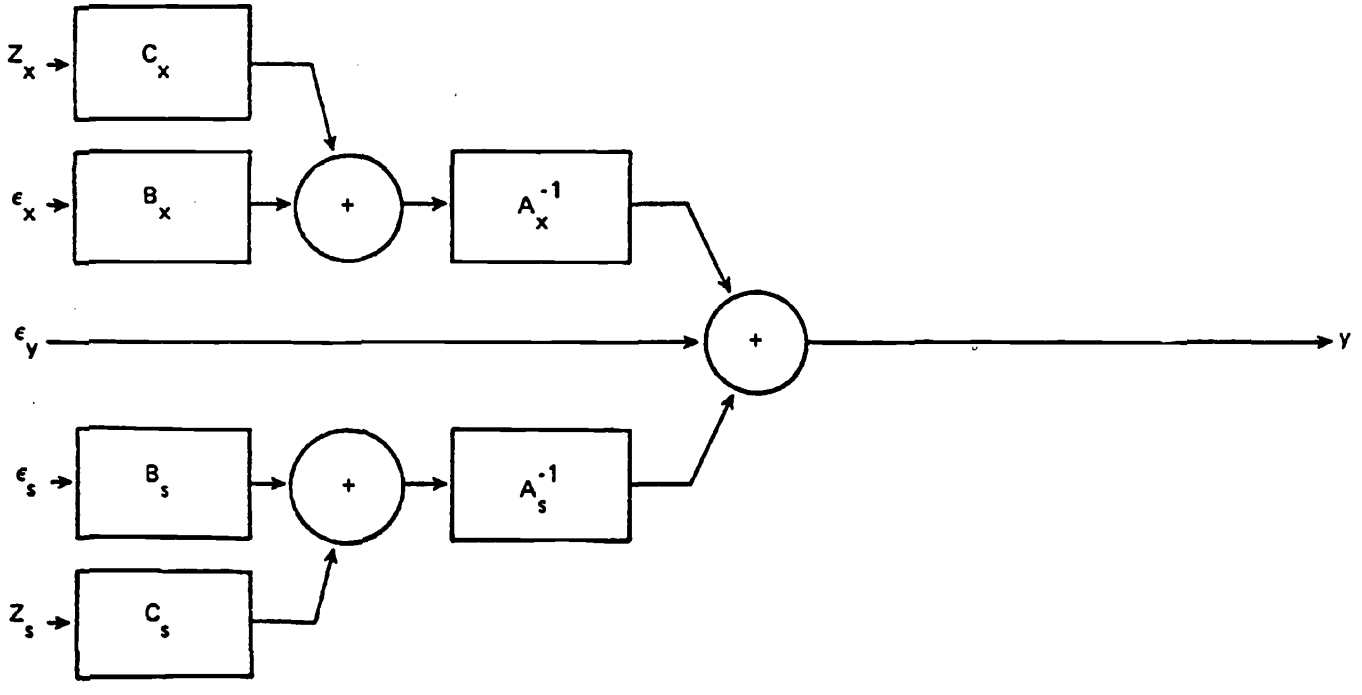


Figure 4. COVARIANCE-EQUIVALENT BLOCK DIAGRAM FOR THE ADDITIVE MODEL WITH CAUSAL VARIABLES

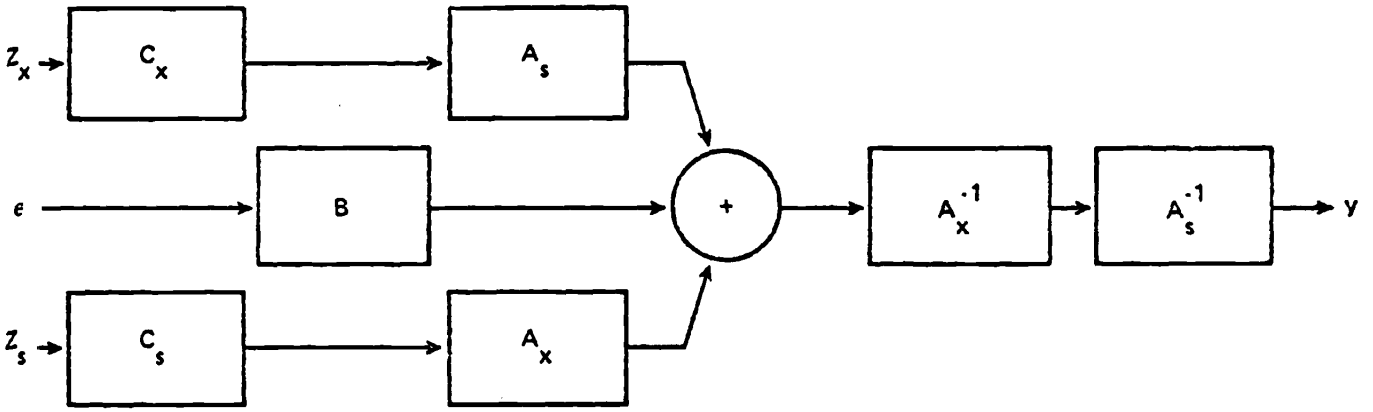
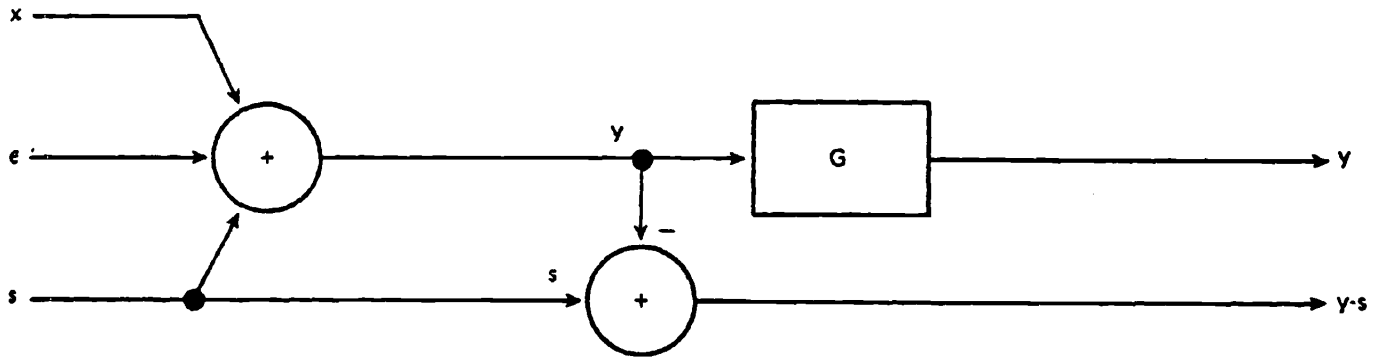


Figure 5. FILTERING AND REMOVING COMPONENTS



maximize the log-likelihood function in order to estimate the parameters in the state-variable formulation, and following Box and Jenkins, he estimates the parameters in an unconstrained ARIMA model. These estimations were performed using a simulated series for which the true form of the model is known.

It would seem that no success can be claimed for the Kalman approach, because estimates were only obtained when the true variance ratios were used and, possibly even more important, when the true parameter values were used for starting values. In this case, the converged values may have simply been constrained to lie near the true values because of the influence of the prior on the Kalman procedure. The unconstrained ARIMA model yielded a good estimate for the seasonal parameter but rather poor values for the regular model parameters. It is clear from table 1 that, in this case, the model is over parameterized, since three of the parameters have nonsignificant t-values.

Using the true and Kalman-approach results (obtained under unrealistic conditions) Engle then apparently used the Kalman filter approach to estimate the seasonal component s and, by subtraction, the seasonally adjusted series $x+\epsilon_y=y-\hat{s}$. He then compared these seasonally adjusted series and the seasonally adjusted series obtained from applications of the X–11 process against the true original composite series $x+\epsilon_y$.

The X–11 process proved to give the highest correlation between the values of the adjusted series. I believe that additional correlations, given in table 2, between various differences of the adjusted series, are unjustified, arbitrary, and potentially misleading.

Professor Engle remarks that for all the seasonally adjusted series, the coherence squared between the true adjusted series and the seasonally adjusted series dips at the seasonal frequency. This dip is entirely expected, because removal of a component (that is subtraction) is not the same as filtering (that is multiplication). In fact, as will be shown, the coherency squared between any filtered version of y and $(y-s)$ is theoretically identical, which explains Engle's inability to "pick one model over the others."

Consider the system shown in figure 5. Now, the spectrum of Gy is $|G|^2\ \Gamma_y=|G|^2(\Gamma_s+\Gamma_x+\Gamma_\epsilon)$, since s, x, and ϵ are independent. Γ denotes the autospectrum and G, the frequency response function. Vertical bars indicate a modulus. Similarly, the spectrum of $y-s=\epsilon+x$ is $\Gamma_x+\Gamma_\epsilon$. The cross-spectrum between $(y-s)$ and Gy is then $G \times$ (cross-spectrum of y and $\epsilon+x)=G(\Gamma_x+\Gamma_\epsilon)$. Hence, the coherency squared is

$$K^2 = \frac{|G(\Gamma_x+\Gamma_x)|^2}{\{|G|^2(\Gamma_s+\Gamma_x+\Gamma_\epsilon)\}\{\Gamma_x+\Gamma_\epsilon\}}$$

$$= \frac{\Gamma_x+\Gamma_\epsilon}{\Gamma_s+\Gamma_x+\Gamma_\epsilon} = \frac{1}{1+\dfrac{\Gamma_s}{\Gamma_x+\Gamma_\epsilon}}$$

independent of the filter G. Now, at the seasonal frequency, $\Gamma_s >> \Gamma_x+\Gamma_\epsilon$ and so $K^2 \approx 0$, while elsewhere $\Gamma_s << \Gamma_x+\Gamma_\epsilon$ and $K^2 \approx 1$, which is how the coherency squared behaves in figure 1, lying near 1 except for dips to 0.3 or 0.4, at frequency 0.25 cycles per year. The coherency squared suffers from bias, however, as shown in Jenkins and Watts, the bias depending mainly on the smoothing window used and the ratio of L/N, the number of covariance lags to the number of observations. For $L=40$, $N=100$, and a Tukey window, the bias is about 0.3, which could explain the nonzero value at the seasonal frequencies.

Turning to the section that deals with a real series with an exogenous variable, we see that Engle abandons the Kalman filter approach and uses statistical methods, proposed by Hannan, and other methods available on the Troll computer system. Several models are fitted, using a barrage of methods, with seemingly satisfactory results to the author that are unsatisfactory to me.

A major criticism can be levied at his procedure which proposes that the autoregressive parameters can be estimated, assuming a general stationary error structure, and the residuals can then be examined to see if they obey the prescribed process. This is, of course, completely antithetical to modern statistical practice that employs careful and clear specification of the error structure and estimation procedures consistent with that error structure, to say nothing of the sampling properties of nonwhite-noise processes and the difficulties these impose. Also, it is not clear just what approach or criterion Engle will use to estimate the autoregressive parameters.

SUMMARY

With regard to the Kalman filter approach, I find it wanting in two major respects. The first is that it requires specification of a state-variable model and the second, that it requires specification of the statistical properties of all the disturbances. In economics, as contrasted with missile guidance, e.g., where does the state-variable model come from? And, from where do the disturbance variances and covariances come? The approach used by Box and Jenkins, which allows the data to lead the way, seems, to me, to be better suited to dealing with economic time series.

Also, concerning multiplicative versus additive models, it would seem that a very worthwhile program could be mounted to consider the problem of discriminating between additive and multiplicative models. This, of course, was not considered by Engle but, perhaps, should have been included since additive models are only one structural form.

Engle's iterative procedure, suggested in the sections on models with exogenous variables and an example with exogenous variables, states that "the autoregressive parameters are estimated, assuming a general stationary error structure. The residuals are then examined to see if they obey the prescribed process." I suggest that, again, the preferred statistical procedure is to develop models

that produce white-noise residuals, which then have the enormous advantage of having simple, well-defined statistical properties.

Indeed, if I may close with a comment concerning seasonal adjustment of series, I have yet to be convinced that seasonal adjustment is the best thing to do to a series. I believe, rather, that the aim of time series model building should be to develop forecasting models that yield white-noise residuals. Nonwhite forecast errors then yield information on turning points, changes in level, changes in parameters and models, etc. And, in addition, such information enjoys the important advantage that probability or confidence levels can be derived, stemming from the properties of white noise.

RESPONSE TO DISCUSSANTS

Robert F. Engle
University of California, La Jolla

The discussants have made a wide range of remarks. I was particularly grateful to Watts for pointing out that the coherence between the true and estimated seasonally adjusted series has a shape independent of the seasonal adjustment filter. Although this result applies for time-invariant linear filters, it must also be approximately true for time-varying filters, such as the Kalman filter, or nonlinear filters, such as the X–11. He, thus, explained why results that could be seen to be widely different gave similar coherence plots.

I take issue, however, with several of his other remarks. Watts asks, in his concluding section, "in economics, as contrasted with missile guidance, for example, where does the state-variable model come from?" The whole point of this paper is to discuss procedures for estimating and testing the state variable or structural models using actual data. It is only by practical experience that we can learn whether techniques developed in other fields have useful economic applications.

Watts criticizes the procedure for estimating "the autoregressive parameters . . . assuming a general stationary error structure," and would prefer "to develop models which produce white-noise residuals." The estimation method employed, however, does just this. It is descended from earlier works of Hannan, Amemiya, and Fuller. However, his point is well taken in a hypothesis testing context, because the ignored constraints can be more easily tested by examining transformed residuals for whiteness than untransformed residuals for the appropriate degree of color. In the revised paper, I have suggested this alternative.

Watts suggests using multiplicative models "with their equally valid justification and vastly simpler and more effective analysis procedures." It is not clear what he has in mind, since he excludes taking logarithms, and since, in general, multiplicative models generate complicated nonlinear statistical procedures.

He asserts that the causal models are "ill conceived, because they impose the same dynamics (A_x^{-1}) on all of the regular exogenous variables Z_x and the same dynamics (A_s^{-1}) on all of the seasonal exogenous variables Z_s." These models, however, do not impose the same dynamics, only the same autoregressive portion. Further generalization, of course, is easy.

Turning now to the comments of Howrey, he first discusses the problem of identification of ARMA models but draws no particular conclusions about unobserved component ARMA models. This identification problem is probably more difficult, but insofar as there are overidentifying restrictions, these might provide additional diagnostic checks of model adequacy.

His second point is concerned with the properties of the estimates of the unobserved components when the state-space model is estimated rather than known a priori. The answer is surely the same as forecasting from any ARMA, nonlinear, or lagged dependent variable model. The parameter estimates will be consistent and, thus, asymptotically, the unobserved component estimates will be unbiased and of minimum variance.