MODELS FOR X-11 AND 'X-11-FORECAST' PROCEDURES FOR PRELIMINARY AND REVISED SEASONAL ADJUSTMENTS

Kenneth F. Wallis

NUMBER 198

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

MODELS FOR X-11 AND 'X-11-FORECAST' PROCEDURES
FOR PRELIMINARY AND REVISED SEASONAL ADJUSTMENTS

Kenneth F. Wallis

NUMBER 198

For presentation at the ASA-CENSUS-NBER Conference on Applied Time
Series Analysis of Economic Data
Washington, DC, October 1981

Revised, December 1981

1. INTRODUCTION

Procedures for the seasonal adjustment of economic time series have typically been evaluated by studying their effect on a sample of actual time series. Recent proposals for amendments and extensions to existing methods have also been evaluated in the same way. Perhaps this approach is thought to be inevitable given that "there seems to be no ideal process of evaluating a method of adjustment" (Granger, 1978, p.55). In contrast, however, this paper continues a line of research in which the properties of the procedures themselves are studied, in the abstract. It is hoped that this will improve our general understanding of the performance of the existing methods and their extensions, and help to explain the results of the previous empirical studies.

The particular procedure considered is the U.S. Bureau of the Census Method II, Variant X-11 (Shiskin et al., 1967), which is widely used and is generally held to give satisfactory results in the seasonal adjustment of historical data. Our analysis proceeds by linear filter methods. The basic framework of a set of "time-varying" linear filters is presented by Wallis (1982), and further properties of these filters and their components are considered in the present paper. The use of linear methods implies that attention is restricted to the performance of X-11 in additive mode (in which seasonal components are estimated as average differences from, not average ratios to, the trend-cycle), neglecting the option of graduating extreme irregular values.

The problem of first adjusting the current month's observation and then revising the seasonally adjusted figure as time goes by and more data become available has received much attention recently. A specific suggestion is that the subsequent revisions in the adjusted values might be reduced by applying X-11 not to the observed series alone but to a series augmented by forecasts of future values. This has been put forward in association with various forecasting methods (compare Dagum, 1975, 1979; Geweke, 1978; Kenny and Durbin, 1982), but since in all these methods the forecasts are calculated as linear combinations of observed values, our analysis can also encompass these "X-11-FORECAST" procedures. Many statistical agencies run the X-11 program only once a year, and at that time project seasonal factors for the adjustment of the next twelve months' data, to be used as the data become available. Kenny and Durbin (1982) and Wallis (1982) have argued that this practice should be replaced by running the program every month, and it is not considered in the present paper.

The linear filter representation of the X-11 seasonal adjustment procedure is presented in Section 2. The filters range from the purely one-sided moving average implicit in the preliminary adjustment of the most recent observation, through a number of asymmetric moving averages, to the symmetric moving average implicit in the adjustment of historical data. The relations with forecast-augmented procedures are considered in Section 3, and following a result given by Geweke (1978) and Pierce (1980) we define the property of internal consistency of a set of linear filters. The polynomial regression origins of smoothing filters are considered in Section 4, and it is shown that whereas sets of filters constructed in this way are internally consistent, the Henderson moving averages used in X-11, for which a new interpretation is given, are not. In Section 5 forecast-augmented procedures are used to provide models for which the X-11 procedure already minimises the mean squared error of revisions, thus illuminating the empirical comparison of various methods. Particular attention is paid to important work by Kenny and Durbin (1982) and Dagum (1975, 1979). The relation to optimal signal extraction methods is briefly considered in Section 6.

2. THE X-11 LINEAR FILTERS

The original observable monthly series, denoted x_t, is the input to a linear filtering operation, of which the output is denoted y_t. In the present context y is the seasonally adjusted value of x, but the basic framework can be applied to any problem of signal extraction, interpolation, extrapolation and smoothing by linear filter methods. The X-11 program comprises a sequence of moving average or linear filter operations, but their net effect can be represented by a single set of moving averages. For a date sufficiently far in the past, the final or historical adjusted value $y_t^{(m)}$ is obtained by application of the symmetric filter $a_m(L)$,

$$y_t^{(m)} = a_m(L)x_t = \sum_{j=-m}^{m} a_{m,j}x_{t-j} \quad ,$$

where L is the lag operator and $a_{m,j} = a_{m,-j}$. This filter is described variously as a 2m+1-term moving average, or as a symmetric filter of half-length m. For current and recent data this filter cannot be applied, and truncated asymmetric filters are employed:

$$y_t^{(i)} = a_i(L)x_t = \sum_{j=-i}^{m} a_{i,j}x_{t-j} \quad .$$

For the filter $a_i(L)$, $i=0,1,\ldots,m$, the subscript i indicates the number of "future" values of x entering the moving average, that is, the number of negative powers of L that appear, or the negative of the lower limit of summation in the expression $\Sigma a_{i,j}x_{t-j}$. Equivalently, the superscript on y indicates that $y_t^{(i)}$ is the adjusted value of x_t calculated from observations at times $t-m, t-m+1,\ldots, t,\ldots,t+i$, and $y_t^{(0)}$ is the first-announced or preliminary seasonally adjusted figure. For the X-11 filters considered here the value of m is 84 (it is assumed that at the stage at which the program chooses a 9-, 13- or 23-term Henderson moving average to estimate the trend-cycle, the 13-term average is chosen). Thus the X-11 program can be represented as a set of 85 linear filters, with respectively 0,1,...,84 coefficients of future values. There is no attempt to compensate for the truncation of the weights applied to future data by increasing the number of past values entering the moving average, thus with few exceptions all filters involve 84 past values, as indicated by the constant upper limit of summation in the above equation, although the remote weights are very small. In Figure 1 the weights and transfer functions of three filters of particular interest, $a_0(L)$, $a_{12}(L)$ and $a_{84}(L)$, are reproduced from Wallis (1982), where details of their calculation can be found; the symmetric filter is also given in Wallis (1974).

At times t+i and t+k two seasonally adjusted values $y_t^{(i)}$ and $y_t^{(k)}$ can be calculated, corresponding to the unadjusted value x_t, and the <u>revision</u> is defined as

$$r_t^{(i,k)} = y_t^{(k)} - y_t^{(i)}, \quad 0 \leq i < k \leq m.$$

This reflects the information in the "new" data $x_{t+i+1}, x_{t+i+2}, \ldots, x_{t+k}$. The <u>total revision</u> from a given point in time t+i is $r_t^{(i,m)}$. Often statistical agencies run seasonal adjustment programs only once a year, so that the revisions then made are $r_t^{(i,i+12)}$. The first annual revisions $r_t^{(0,12)}$ are of particular interest below.

3. FORECAST-AUGMENTED PROCEDURES AND THE INTERNAL CONSISTENCY OF SETS OF LINEAR FILTERS

In the X-11-FORECAST procedures current data are adjusted not by the one-sided filter $a_o(L)$ but by later filters applied to a series extended by forecasts. In practice twelve such forecasts are used, thus the procedures calculate the preliminary adjusted value not as $a_o(L)x_t$ but as $a_{12}(L)\tilde{x}_t$, where the tilde (~) indicates that the input to the filter is the augmented series $x_1, \ldots, x_t, \hat{x}_{t+1}, \ldots, \hat{x}_{t+12}$. Since the forecasts are calculated as linear combinations of observed values, the result is still a one-sided filter of the original data. Writing the forecasts as

$$\hat{x}_{t+k} = \sum_{j=o}^{\ell} f_{kj} \; x_{t-j} \;, \qquad k = 1, \ldots, 12$$

the new one-sided moving average is

$$\begin{aligned} \tilde{a}_o(L)x_t &= \sum_{k=1}^{12} a_{12,-k} \, \hat{x}_{t+k} + \sum_{j=o}^{m} a_{12,j} \, x_{t-j} \\ &= \sum_{j=o}^{\max(m,\ell)} \left(a_{12,j} + \sum_{k=1}^{12} a_{12,-k} f_{kj}\right) x_{t-j}. \end{aligned}$$

Dagum (1975, 1979) bases the forecasts on ARIMA models fitted to the observed series, while Kenny and Durbin (1982) use autoregressive models fitted by stepwise regression methods to the first differences of the series.

Support for X-11-FORECAST procedures is provided by a result given by Geweke (1978) and Pierce (1980), namely that the asymmetric filter for which the total revision has smallest mean square is given by the application of the symmetric filter $a_m(L)$ to a series extended to the extent necessary by <u>optimal</u> linear forecasts. Similarly the one-sided filter for which the first annual revision $r_t^{(0,12)}$ has smallest mean square is $\tilde{a}_o(L)$ above, provided that the forecasts $\hat{x}_{t+k}$ are optimal. (Many of the empirical evaluations compare the total revisions of various procedures, even though only twelve forecasts are employed in the X-11-FORECAST procedures; we return to this below.) In Geweke's words, "the best linear forecast of any given linear combination of x_t's is the given linear combination of the forecasted x_t's."

Given a symmetric moving average these results provide a way of constructing asymmetric moving averages for the adjustment of recent data, based on optimal forecasting equations for a given x-process. We say that a set of linear filters constructed in this way is internally consistent, with respect to the given x-process. Likewise a given set of linear filters can be examined for its internal consistency by asking what forecast function and hence what x-process is implied if $a_i(L)$ is equivalent to the application of $a_m(L)$ to a series extended by m-i forecasts, and is this x-process the same for all possible pairs. By considering two intermediate adjusted values $y_t^{(i)}$ and $y_t^{(k)}$, $0 \leq i < k < m$, and repeatedly applying Geweke and Pierce's result, in effect repeatedly taking conditional expectations, we see that an internally consistent set of filters minimizes revisions throughout the whole sequence of adjustments. Thus if the filters $a_i(L)$ and $a_k(L)$ minimize the mean square of the revisions $r_t^{(i,m)}$ and $r_t^{(k,m)}$ respectively, being identical to the application of $a_m(L)$ to a sequence of observations augmented by $m - i$ and $m - k$ forecasts respectively, then $a_i(L)$ also minimizes the mean square of the revisions $r_t^{(i,k)}$ and is identical to the application of $a_k(L)$ to a series extended by $k - i$ forecasts. Note that while the theoretical result refers to an optimal forecast given the infinite past, the practical filters involve a finite past, hence some restriction on the x-processes considered is necessary in order for the equivalences among filters to hold exactly. Finite-order autoregressive models clearly allow correspondences among the practical filters, but moving average models whose infinite autoregressive representations do not converge quickly allow only an approximation to an optimal forecast equivalence.

As an illustration, we examine the 5-term "Henderson" moving averages used for estimation of the trend-cycle component in the quarterly version of the X-11 program (Shiskin et al., 1967, Appendix B, Table 3A):

$$a_o(L)x_t = 0.670x_t + 0.403x_{t-1} - 0.073x_{t-2}$$

$$a_1(L)x_t = 0.257x_{t+1} + 0.522x_t + 0.294x_{t-1} - 0.073x_{t-2}$$

$$a_2(L)x_t = -0.073x_{t+2} + 0.294x_{t+1} + 0.558x_t + 0.294x_{t-1} - 0.073x_{t-2}$$

We seek the x-process and associated forecast coefficients implied by the assumption that the one-sided filter $a_o(L)x_t$ is identical to $a_2(L)\tilde{x}_t$, that is, the symmetric filter applied to the forecast-augmented series $x_{t-2}, x_{t-1}, x_t, \hat{x}_{t+1}, \hat{x}_{t+2}$. For an exact equivalence the forecasts can only involve three observed x-values, so we postulate a third-order autoregressive process, with forecasts given by

$$\hat{x}_{t+1} = \phi_1 x_t + \phi_2 x_{t-1} + \phi_3 x_{t-2},$$

$$\hat{x}_{t+2} = \phi_1 \hat{x}_{t+1} + \phi_2 x_t + \phi_3 x_{t-1}$$

$$= (\phi_1^2 + \phi_2)\, x_t + (\phi_1\phi_2 + \phi_3)\, x_{t-1} + \phi_1\phi_3 x_{t-2} .$$

Equating the coefficients of, respectively, x_t, x_{t-1} and x_{t-2} in $a_o(L)x_t$ and $a_2(L)\tilde{x}_t$ then gives three polynomial equations for the ϕ's, which have four solutions. Writing them in the form $\phi(L) = 1-\phi_1 L-\phi_2 L^2-\phi_3 L^3$, these are

$$(1 - 0.568L - 0.432L^2), \qquad (1 - 4.459L + 3.459L^2),$$

$$(1 - 3.027L - 1.493L^2), \qquad (1 - 4.027L + 2.534L^2 + 1.493L^3).$$

Note that this mathematical problem has multiple solutions even when one is attempting to recover an autoregressive operator that has actually been used to construct the one-sided filter from the symmetric filter. One hopes in general that (at least) one of these solutions has a plausible statistical interpretation. To consider this it is more informative to express $\phi(L)$ in terms of its roots $\phi(L) = \Pi\ (1 - \lambda_i L)$, whereupon the solutions can be written

$$(1-L)\ (1 + 0.432L), \qquad (1-L)(1 - 3.459L),$$
$$(1 + 0.432L)(1 - 3.459L), \qquad (1-L)(1 + 0.432L)(1 - 3.459L).$$

Of these, the first is acceptable as an ARIMA (1,1,0) model, but the remainder are not acceptable from a statistical point of view due to the presence of the explosive root. Thus we conclude that $a_o(L)$ is equivalent to the application of $a_2(L)$ to a sequence of observations augmented by optimal forecasts for series obeying the first model, that is, for such a series $a_o(L)$ minimizes the meansquare of the revisions $r_t^{(0,\ 2)}$.

The problem of non-linearity does not arise when comparing $a_o(L)$ with $a_1(L)$, or $a_1(L)$ with $a_2(L)$, since only one forecast is involved, whose coefficients are (linear in) the ϕ's. Simple calculation gives the model relating $a_o(L)$ to $a_1(L)$ as

$$(1-L)(1 + 0.424L)\ x_t = \varepsilon_t$$

and the model relating $a_1(L)$ to $a_2(L)$ as

$$(1-L)(1 + 0.493L)\ x_t = \varepsilon_t.$$

Although the three models are very similar, their coefficient values do not exactly coincide, and we conclude that these filters are not internally consistent in this sense.

4. INTERNAL CONSISTENCY OF REGRESSION-BASED FILTERS

4.1 Least squares polynomial regression

Filters used to estimate the trend-cycle component prior to estimation of the seasonal component are usually designed to reproduce, locally, a polynomial of prescribed degree d. A least squares estimate of the trend value then emerges as a linear function of the data, whose coefficients are independent of time, and these are the moving average coefficients (see, for example, Kendall, 1973, Ch.3). Given a sample of n observations $x_1, \ldots, x_n$ placed in a column vector $\underset{\sim}{x}$, and an $n\times(d+1)$ matrix $\underset{\sim}{Z}$ whose j^{th} column comprises the $(j-1)^{th}$ powers of the integers $1, \ldots, n$, the trend estimate at time t is

$$y_t = \underset{\sim}{z}_t' (\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}\underset{\sim}{Z}'\underset{\sim}{x}$$

where $\underset{\sim}{z}_t' = (1,t,t^2, \ldots, t^d)$. Writing the coefficients of the observations in a column vector $\underset{\sim}{a}$, with the coefficients of the earliest data entering first, then

$$y_t = \underset{\sim}{a}'\underset{\sim}{x}\ , \qquad \underset{\sim}{a} = \underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}\underset{\sim}{z}_t\ .$$

The coefficients are independent of time in the sense that, on indexing the data $t_o+1, t_o+2, \ldots, t_o+n$ and using these integers to construct $\underset{\sim}{Z}$ and $\underset{\sim}{z}_t$ conformably, the same coefficients are obtained whatever the value of t_o. So when $n = 2m + 1$ it is conventional to let the time index run from $-m$ to m, for arithmetical convenience. (Similarly the use of orthogonal polynomials simplifies the arithmetic without affecting the results.) If n is odd and t is the mid-point of the sample, then a symmetric moving average is obtained. Kendall (1973, Appendix A) presents Cowden's tabulations of moving averages for trend estimation at $t = 0,1 \ldots, n, n+1$, for given values of n and d. Such a table does not correspond to a set of filters in the sense of Section 2, but such a set is easily constructed by putting $t = m+1$ and considering in turn $n = m+1, m+2, \ldots, 2m+1$, and this set is internally consistent with respect to the regression model.

To see this consider the filter applicable to $n \leq 2m + 1$ data points and that applicable to $k < n$ (but $k > d$) data points. For internal consistency the latter must be equivalent to the application of the former to the series $x_1, \ldots, x_k, \hat{x}_{k+1}, \ldots, \hat{x}_n$, where the forecasts are now obtained by polynomial regression. We partition the vector $\mathbf{x}$ and matrix $\mathbf{Z}$ into their first k and last (n-k) rows, thus

$$\mathbf{x}' = (\mathbf{x}_1' \vdots \mathbf{x}_2') , \quad \mathbf{Z}' = (\mathbf{Z}_1' \vdots \mathbf{Z}_2') .$$

The forecasts $\hat{\mathbf{x}}_2$ obtained from polynomial regression based on the data $\mathbf{x}_1$ are

$$\hat{\mathbf{x}}_2 = \mathbf{Z}_2 (\mathbf{Z}_1'\mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{x}_1 ,$$

and applying the n-term moving average to the series $\tilde{\mathbf{x}}' = (\mathbf{x}_1' \vdots \hat{\mathbf{x}}_2')$ gives the trend estimate at time t as

$$\begin{aligned} y_t &= \mathbf{z}_t' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \tilde{\mathbf{x}} \\ &= \mathbf{z}_t' (\mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{Z}_2'\mathbf{Z}_2)^{-1} (\mathbf{Z}_1'\mathbf{x}_1 + \mathbf{Z}_2'\mathbf{Z}_2 (\mathbf{Z}_1'\mathbf{Z}_1)^{-1} \mathbf{Z}_1'\mathbf{x}_1) \\ &= \mathbf{z}_t' (\mathbf{Z}_1'\mathbf{Z}_1)^{-1} \mathbf{Z}_1'\mathbf{x}_1 \end{aligned}$$

which is identical to the application of the corresponding k-term filter to the data $\mathbf{x}_1$.

4.2 The Henderson moving averages

These filters are designed to reproduce a cubic polynomial trend. In obtaining a general expression for the weights of the symmetric filter, Henderson (1916) showed that three alternative smoothing criteria lead to the same result:

(i) minimization of the variance of the third difference of the output series,

(ii) minimization of the sum of squares of the third differences of the moving average coefficients,

(iii) fitting a cubic by weighted least squares, with the sum of squares of the third differences of the weights a minimum.

A more accessible presentation of the general derivation and a demonstration of the equivalence of these criteria is in the Appendix of Kenny and Durbin (1982). A further interpretation can be obtained using the approach of Hannan (1970, pp. 186-7), as follows.

Augmenting the moving average coefficient vector $\underset{\sim}{a}$ with zeros to accommodate the differencing operations, criterion (ii), the sum of squares of the third differences of these coefficients, can be written

$$S = \underset{\sim}{a}'\underset{\sim}{V}\underset{\sim}{a}$$

where $\underset{\sim}{V}$ is a symmetric band matrix with elements on the main diagonal and first three sub-diagonals equal to 20,-15,6 and -1 respectively, all other elements being zero. Hannan shows that minimization of this general expression subject to the condition that the filter reproduces an arbitrary polynomial of degree d yields the solution

$$y_t = \underset{\sim}{a}'\underset{\sim}{x} = \underset{\sim}{z}_t' (\underset{\sim}{Z}'\underset{\sim}{V}^{-1}\underset{\sim}{Z})^{-1} \underset{\sim}{Z}'\underset{\sim}{V}^{-1}\underset{\sim}{x} \ .$$

In Hannan's formulation $\underset{\sim}{V}$ is the covariance matrix of the input series, and he notes that, as is well known, this result corresponds to using the best linear unbiased regression of the input series on the polynomials to estimate the trend value at the relevant time point. In the present case the particular matrix $\underset{\sim}{V}$ is the covariance matrix of the moving average process

$$u_t = (1-L)^3 \varepsilon_t,$$

and so we have a further interpretation of the Henderson filters, namely as generalized least squares polynomial regression subject to this error structure. If n is odd and t is the mid-point of the sample then the coefficients are again symmetric, and using this approach we can readily verify their published values.

By setting $t = m+1$ and considering in turn $n = m+1, m+2, \ldots , 2m+1$ the above expression gives a set of filters of the kind discussed earlier. This set is again internally consistent with respect to the present regression model. The proof of this is a tedious generalization of the proof for the least squares case given in section 4.1, taking account of the need to define $\hat{x}_{k+1}, \ldots , \hat{x}_n$ as the best linear unbiased predictors of $x_{k+1}, \ldots , x_n$ given $x_1, \ldots , x_k$.

On computing sets of filters for m = 4, 6 and 11 in this way we immediately notice that, except for the symmetric filter, the coefficients differ from those used in X-11. The source of the asymmetric moving averages given by Shiskin et al. (1967, Appendix B, Tables 3B-3D) corresponding to the symmetric 9-, 13- and 23-term Henderson moving averages is something of a mystery: they are not due to Henderson (1916), no alternative source is cited, and they are not the result of applying Henderson's criteria to the design of asymmetric filters. Our calculations in general find that the weights given to current data are somewhat larger than those in the tabulated filters used in X-11. Kenny and Durbin (1982) report that the "one-sided version is constructed essentially by assuming that the series to be smoothed can be extended by a straight line fitted by least squares", which is their own empirical finding for the monthly filters. For the quarterly 5-term filters analysed in Section 3 we do not claim that the autoregressive models we have deduced were actually used to construct the asymmetric filters, but we note that the models imply that the forecast is approximately given as the average of the last two observed values, which may be a more plausible interpretation of the actual derivation. Basing asymmetric filters on forecasts obtained from a different model than that embodied in the symmetric filter leads to internal inconsistency. In turn this leads one to expect that improvements over the existing one-sided seasonal adjustment filters could be made for series well-described by the model underlying the symmetric filter and hence satisfactorily adjusted by it.

5. FORECASTING INTERPRETATIONS OF THE RELATIONS BETWEEN SELECTED FILTERS

5.1 Autoregressive models relating $a_o(L)$ to $a_{12}(L)$

The filters under consideration require a finite amount of past data, and in section 3 the possibility was suggested of finding an autoregressive model and associated forecast functions such that an existing filter $a_i(L)$ is already equivalent to the application of $a_k(L)$ to a series augmented by k-i forecasts. For such a model the mean square of the revision $r_t^{(i,k)}$ is already at a minimum. Much research has focussed on the adjustment of current data, and the practical X-11-FORECAST procedures augment the observed series with 12 forecasts, and so first we consider the relation between the X-11 filters $a_o(L)$ and $a_{12}(L)$.

To obtain a general characterization we follow the approach of Kenny and Durbin (1982), in which forecasts are obtained from autoregressive models. They estimate these by regressing the first difference Δx_t on $\Delta x_{t-1}, \ldots, \Delta x_{t-25}$, reducing the set of regressors by the Efroymson stepwise method. Neglecting the restriction implied by the use of the first difference operator, the general model is

$$\phi(L)x_t = \varepsilon_t,$$

where $\phi(L)$ is of degree 26. For this model we construct the forecast coefficients for one-step, two-step, ... , twelve-step forecasts

$$\hat{x}_{t+k} = \sum_{j=o}^{25} f_{kj}\, x_{t-j}, \qquad k=1,\ldots,12$$

by starting with the expression for the one-step forecast

$$\hat{x}_{t+1} = \phi_1 x_t + \phi_2 x_{t-1} + \cdots + \phi_{26} x_{t-25}$$

and then using, sequentially, the expression for the k-step forecast

$$\hat{x}_{t+k} = \phi_1 \hat{x}_{t+k-1} + \cdots + \phi_{k-1} \hat{x}_{t+1} + \phi_k x_t + \cdots + \phi_{26} x_{t+k-26}, \qquad k=2,\ldots,12.$$

We then seek the ϕ's that equate the one-sided filter implied by the application of $a_{12}(L)$ to the forecast-augmented series, namely

$$\tilde{a}_o(L) = \sum_{j=o}^{25} \left[a_{12,j} + \sum_{k=1}^{12} f_{kj}\, a_{12,-k} \right] L^j + \sum_{j=26}^{m} a_{12,j}\, L^j ,$$

to $a_o(L)$. Note that the use of a finite autoregression implies that the later coefficients in the new filter are simply those of $a_{12}(L)$, and so the new filter cannot be exactly equated to $a_o(L)$. (Comparing the coefficients of $a_o(L)$ and $a_{12}(L)$ plotted in Figure 1, we see that there is little difference in the coefficients at lags greater than 25 except at lag 36.) Equating the earlier coefficients in the two filters gives

$$a_{o,j} = a_{12,j} + \sum_{k=1}^{12} f_{kj}\, a_{12,-k} , \qquad j=0,1,\ldots,25,$$

a set of 26 nonlinear equations in the 26 unknown ϕ's which we solve by numerical methods. (A nonlinear least squares algorithm is used: at a solution the residual sum of squares is zero.) As in the simple example of section 3 there are multiple solutions to this numerical problem, and while we cannot claim to have examined them all, indeed we are not sure how many there are, the model given below has the most plausible statistical interpretation and is also suggested by the models relating successive pairs of filters.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
ϕ_i	-.60	-.40	-.26	.13	-.19	.38	-.16	.29	-.13	.12	.11	-.28	-.21
i	14	15	16	17	18	19	20	21	22	23	24	25	26
ϕ_i	.27	.14	-.15	.15	-.20	.23	-.16	.15	-.08	-.09	-.39	.59	.01

To interpret this general autoregression we calculate the roots of the autoregressive operator, $\phi(L) = \Pi\ (1-\lambda_i L)$. There are two real roots, (1+0.982L) and (1+0.018L), and twelve complex conjugate pairs, whose modulus and argument (in radians) are as follows:

mod	arg	mod	arg
1.033	0.10	0.981	0.27
0.958	0.51	0.945	0.82
0.983	1.06	0.965	1.37
0.985	1.57	0.975	1.89
0.982	2.10	0.978	2.42
0.985	2.62	0.988	2.81

Noting that the arguments in the first block of six pairs of roots are approximately $k\pi/6$, $k=0,1,\ldots,5$, these together with the first real root can be approximated as $(1-L)(1-L^{12})$, the differencing operator commonly employed in ARIMA models for seasonal time series. While this solution contains a complex conjugate pair of roots with modulus greater than one, it is one of the least explosive solutions to this problem we have found, and no solution representing a stationary autoregression has been observed. In empirical time series analysis the use of as many as twentysix coefficients to specify a model would be thought excessive, and likewise the relations between the filters may be more readily interpreted in terms of seasonal ARIMA models parameterized more parsimoniously, as considered in section 5.2.

The implication of this result is that, for a series with such autoregressive structure, the existing one-sided X-11 filter $a_o(L)$ minimises the mean square of the first annual revision $r_t^{(0,12)}$, and no X-11-FORECAST procedure of this kind will lead to an improvement. Of course for series having substantially different autoregressive structures, the appropriate X-11-FORECAST procedure will outperform X-11. Attention is then drawn to Kenny and Durbin's anomalous finding that for one group of nine series (those with MCD=1) the X-11 $a_o(L)$ gives better results than the X-11-FORECAST $\tilde{a}_o(L)$. If this finding is not due to sampling variation, then three possible explanations are suggested.

The first possibility, on which we have little further to say, is that the various nonlinearities in the practical procedures vitiate our linear analysis. The second is that the comparison is based, as in other studies (e.g. Kuiper, 1978), on the total revision $r_t^{(0,m)}$ of the two methods and not the first annual revision $r_t^{(0,12)}$. The one-sided filter that minimises the mean square of $r_t^{(0,m)}$ is $a_m(L)$ applied to a series augmented by m forecasts; call this $a_o^*(L)$. On the other hand the filter being considered is $\tilde{a}_o(L)$, namely $a_{12}(L)$ applied to a series augmented by 12 forecasts. If the filters were internally consistent then $a_{12}(L)$ would be equivalent to the application of $a_m(L)$ to a series augmented by m-12 forecasts and $a_o(L)$ could be obtained either by applying this $a_{12}(L)$ to a series augmented by 12 forecasts or as $a_o^*(L)$. However, our earlier results suggest and later results confirm that the X-11 filters are not internally consistent, and we have the possibility that for these series the X-11 $a_o(L)$ is a better approximation to $a_o^*(L)$ than is $\tilde{a}_o(L)$, hence the result.

A third explanation involves the possible non-optimality of the autoregressive forecast, which might apply if the appropriate models for the series had important moving average components. However the form of the seasonal ARIMA models identified by Kenny and Durbin for these series is little different in this respect from those for the series for which X-11-FORECAST methods dominate, and without knowing the coefficient values this possibility cannot be assessed. Using the conventional Box-Jenkins notation $(p,d,q) \times (P,D,Q)_{12}$, the nine series giving the anomalous results have p+d+D = 3, and this distinguishes them from the other series. Five of the nine models are $(0,2,1) \times (0,1,1)_{12}$, and the remaining four also have P=0 and Q=1. Leaving aside questions about the existence of measures such as MCD for models of this kind (Burridge, 1982) and despite the fact that if MCD=1 the X-11 program applies the 9-term Henderson moving average to estimate the trend-cycle component, not the 13-term average used in our linear filter calculations, the results suggest that such models might emerge when we consider equivalences between filters based on ARIMA models.

5.2 ARIMA models relating $a_o(L)$, $a_{12}(L)$ and $a_{84}(L)$

In this section we report the results of a comparable exercise to that of the previous section, except that seasonal ARIMA models replace the finite autoregressive model. That is, for a given seasonal ARIMA model we construct the forecast coefficients and hence the X-11-FORECAST filter coefficients

$$\tilde{a}_{o,j} = a_{12,j} + \sum_{k=1}^{12} f_{kj}\, a_{12,-k}, \qquad j=0,1,\ldots,m,$$

and then seek the model and its associated parameter values such that $\tilde{a}_o(L)$ is as close as possible to $a_o(L)$. For a given model the parameter values are chosen to minimise the unweighted sum of squared deviations between the filter coefficients. The choice among candidate models is partly based on the sum of squares, but we also have in mind the various criteria used to interpret such models when they are fitted to observed time series. For comparative purposes we include in the search the three models available in the automatic version of Statistics Canada X-11-ARIMA (Dagum, 1979), and it turns out that one of these is our preferred model. The results for these three models are as follows.

Estimated model relating $a_o(L)$ to $a_{12}(L)$	Sum of squared deviations
$(1-1.37L + 0.39L^2)(1-L)(1-L^{12})\, x_t = (1-1.49L + 0.76L^2)(1-0.69L^{12})\varepsilon_t$	0.0096
$(1-L)^2\, (1-L^{12})\, x_t = (1-1.32L + 0.63L^2)(1-0.74L^{12})\varepsilon_t$	0.0111
$(1-L)(1-L^{12})\, x_t = (1-0.25L)(1-0.69L^{12})\varepsilon_t$	0.0188

The final column can be compared to the "zero-forecast" sum of squares

$$\sum_{j=o}^{m} (a_{o,j} - a_{12,j})^2 = 0.0477 .$$

We note that the autoregressive operator in the preferred model has a root close to unity, but setting it equal to unity worsens the fit, and we retain the $(2,1,2)\times(0,1,1)_{12}$ model for further comparisons below, where this feature is less pronounced.

Again the interpretation is that for a series obeying the model $(2,1,2) \times (0,1,1)_{12}$, with the above parameter values, the existing X-11 filter $a_o(L)$ is a close approximation to the X-11-ARIMA filter that minimises the mean square of the first annual revision $r_t^{(0,12)}$. In empirical comparisons of the two methods the advantage of X-11-ARIMA will increase as the x-process differs more extensively from that given above. In practice we note that substantial variation can be observed as parameter values change within a given model.

To consider filters that minimise the mean square of the total revision $r_t^{(o,m)}$ we repeat this exercise for the filter $a_o^*(L)$ obtained by applying the symmetric filter $a_m(L)$ to a series augmented by m forecasts:

$$a^*_{o,j} = a_{mj} + \sum_{k=1}^{m} f_{kj} \; a_{mk} , \qquad j=0,1,\dots,m.$$

The preferred model is again $(2,1,2) \times (0,1,1)_{12}$, and the results for the three Dagum models are as follows.

Estimated model relating $a_o(L)$ to $a_m(L)$	Sum of squared deviations
$(1-1.23L + 0.35L^2)(1-L)(1-L^{12})x_t = (1-1.45L + 0.73L^2)(1-0.62L^{12})\varepsilon_t$	0.0114
$(1-L)^2(1-L^{12})x_t = (1-1.33L + 0.58L^2)(1-0.65L^{12})\varepsilon_t$	0.0124
$(1-L)(1-L^{12})x_t = (1-0.21L)(1-0.66L^{12})\varepsilon_t$	0.0218

$$\sum_{j=o}^{m} (a_{o,j}-a_{mj})^2 = 0.0696$$

Results in earlier sections cast doubt on the internal consistency of the X-11 filters, and while the models relating these three filters are not identical, the form of the preferred model is the same in each case, and the coefficients are remarkably similar. To complete the picture we report the coefficient values for the $(2,1,2) \times (0,1,1)_{12}$ model that best approximates $a_{12}(L)$ by the application of $a_m(L)$ to a series augmented by m-12 forecasts:

$$(1-1.08L + 0.21L^2)(1-L)(1-L^{12})\ x_t = (1-1.42L + 0.70L^2)(1-0.60^{12})\varepsilon_t.$$

Once more these values are similar, suggesting that whatever adjusted values empirical comparisons are based on - $y_t^{(o)}$, $y_t^{(12)}$ or $y_t^{(m)}$ - then for series well-described by this model X-11 will produce approximately the same results as X-11-FORECAST methods.

6. OPTIMAL SIGNAL EXTRACTION

The problem of seasonal adjustment is formulated as a problem of optimal signal extraction by Grether and Nerlove (1970) and solved using methods presented by Whittle (1963). It is assumed that the observed series x_t is the sum of two unobserved uncorrelated stationary components, s_t and u_t, respectively the seasonal and non-seasonal components. Assuming that the stochastic structure of s_t and u_t, and hence that of x_t, is known Whittle gives expressions for $s_t^{(k)}$, the minimum mean squared error estimate of s_t given data $\{x_\tau ; \tau \leq t+k\}$. These expressions are linear filters of the data, in general of semi-infinite extent; the case $k=\infty$ gives a symmetric filter. Extensions of these results to models in which the components follow non-stationary ARIMA schemes are given by Cleveland and Tiao (1976) (see also Pierce, 1979), who present a components model for which the X-11 symmetric filter $a_m(L)$ approximates the optimal signal extraction filter. The relations between the filters considered in the previous section are examined in the light of such models by Burridge and Wallis (1981).

The practical difficulty in implementing seasonal adjustment methods based on optimal signal extraction results is that of identifying models for the components from the observed series, given that different components models can lead to the same overall model. Burman (1980) makes some practical proposals in this respect, and implements a seasonal adjustment method based on the resulting decomposition. We are not yet able to incorporate this approach into our comparisons, but anticipate that for series well-described by the models for which the X-11 filters approximate optimal signal extraction filters, little benefit from the new methods will be found in empirical comparisons.

ACKNOWLEDGEMENTS

The computational assistance and helpful comments of Peter Burridge are gratefully acknowledged. This research was supported by a grant from the Social Science Research Council.

REFERENCES

BURMAN, J.P. (1980). Seasonal adjustment by signal extraction. Journal of the Royal Statistical Society A, 143, 321-337.

BURRIDGE, P. (1982). Contribution to the discussion of Kenny and Durbin (1982).

BURRIDGE, P. and WALLIS, K.F. (1981). Unobserved-components models for seasonal adjustment filters. Unpublished paper, Department of Economics, University of Warwick.

CLEVELAND, W.P. and TIAO, G.C. (1976). Decomposition of seasonal time series: a model for the Census X-11 program. Journal of the American Statistical Association, 71, 581-587.

DAGUM, E.B. (1975). Seasonal factor forecasts from ARIMA models. Bulletin of the International Statistical Institute, 46, part 3, 203-216.

_______ (1979). The X-11-ARIMA seasonal adjustment method - outline of the methodology. Catalogue no. 12-564E, Statistics Canada.

GEWEKE, J. (1978). The revision of seasonally adjusted time series. Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp.320-325.

GRANGER, C.W.J. (1978). Seasonality : causation, interpretation, and implications (with discussion). In Seasonal Analysis of Economic Time Series (A. Zellner, ed.), pp.33-55. Washington, DC : Bureau of the Census.

GRETHER, D.M. and NERLOVE, M. (1970). Some properties of "optimal" seasonal adjustment. Econometrica, 38, 682-703.

HANNAN, E.J. (1970). Multiple Time Series. New York: Wiley.

HENDERSON, R. (1916). Note on graduation by adjusted average. Transactions of the Actuarial Society of America, 17, 43-48.

KENDALL, M.G. (1973). Time-Series. London : Griffin.

KENNY, P.B. and DURBIN, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series (with discussion). Journal of the Royal Statistical Society A, 145, 000-000.

KUIPER, J. (1978). A survey and comparative analysis of various methods of seasonal adjustment. In Seasonal Analysis of Economic Time Series (A. Zellner, ed.), pp.59-76. Washington, DC: Bureau of the Census.

PIERCE, D.A. (1979). Signal extraction error in nonstationary time series. Annals of Statistics, 7, 1303-1320.

______ (1980). Data revisions with moving average seasonal adjustment procedures. Journal of Econometrics, 14, 95-114.

SHISKIN, J., YOUNG, A.H. and MUSGRAVE, J.C. (1967). The X-11 variant of the Census Method II seasonal adjustment program. Technical paper no. 15 (revised). Washington, DC: Bureau of the Census.

WALLIS, K.F. (1974). Seasonal adjustment and relations between variables. Journal of the American Statistical Association, 69, 18-31.

______ (1982). Seasonal adjustment and revision of current data : linear filters for the X-11 method. Journal of the Royal Statistical Society A, 145, 000-000.

WHITTLE, P. (1963). Prediction and Regulation by Linear Least-Square Methods. London: English Universities Press.

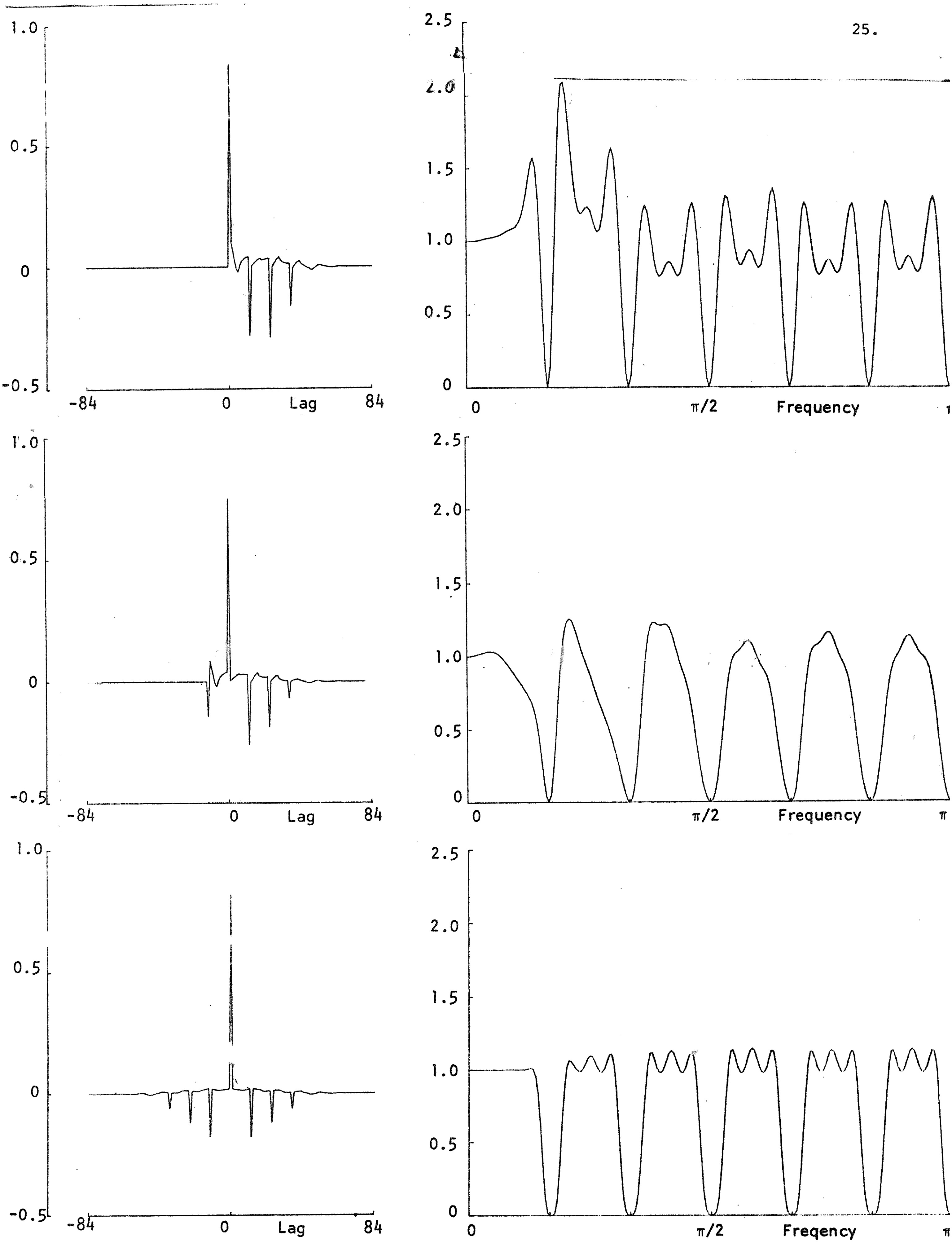

Figure 1 : Weights and transfer function of $a_o(L)$, $a_{12}(L)$ and $a_{84}(L)$