WORKING PAPER TSRA-87-012E

TIME SERIES RESEARCH & ANALYSIS DIVISION
METHODOLOGY BRANCH

1759

SEASONAL ADJUSTMENT IN THE 80'S
Some Problems and Solutions

by
Estela Bee Dagum, G. Huot and M. Morry

This is a preliminary version., Do not quote without authors' permission.
Comments are welcome.

SEASONAL ADJUSTMENT IN THE 80's:
Some Problems and Solutions

by

Estela Bee Dagum(1), Guy Huot(1) and Marietta Morry(1)

(1) Time Series Research and Analysis Division, Statistics Canada, Ottawa, K1A-OT6.

ABSTRACT

Most statistical bureaus do not apply uniform practices for the seasonal adjustment of related series and, consequently, the presence of inconsistencies may occur. This lack of uniformity results mainly from the fact that: (1) the seasonal adjustment method, the X-11-ARIMA, can be used in four different modes (each mode producing different current seasonally adjusted values of the same series); and (2) that most key economic and social indicators are large aggregates and the results differ whether they are seasonally adjusted directly or indirectly (through the aggregation of each of the seasonally adjusted components). Other problems faced currently by Statistics Canada and other statistical bureaus which will also be discussed in this paper are: the smoothing of highly irregular seasonally adjusted series, the estimation of trading day variation and the estimation of Easter effect.

KEYWORDS: current seasonal adjustment, smoothed seasonally adjusted estimates, trading day variation, Easter adjustment, aggregation

1. INTRODUCTION

During this decade Statistics Canada and other statistical bureaus focussed their attention on several important issues concerning the seasonal adjustment of time series, namely, (1) the seasonal adjustment of current observations; (2) the smoothing of highly irregular seasonally adjusted series; (3) the seasonal adjustment of aggregates; (4) the estimation of trading day variation; and (5) the estimation of Easter effects.

The main purpose of this paper is to discuss each of the above problems in the context of the X-11-ARIMA seasonal adjustment method developed by Dagum (1980) and applied by Statistics Canada and other statistical agencies throughout the world.

Section 2 introduces the four modes in which the X-11-ARIMA computer package can be used to produce a current seasonally adjusted value and the consequences of applying different modes to related series. Section 3 deals with the nature and characteristics of the smoothing (trend-cycle) filters available in X-11-ARIMA to help reveal better the short term trend present in highly irregular series. Section 4 discusses the problem of how to handle the seasonal adjustment of aggregate series and its components. Section 5 deals with the estimation of trading day variations and finally, Section 6 addresses the problems associated with estimation of Easter variations.

2. SEASONAL ADJUSTMENT OF CURRENT VALUES

A current seasonally adjusted value can be obtained using either a concurrent or a forward seasonal estimate. A "concurrent" seasonal estimate (factor or effect depending on whether a multiplicative or additive model is assumed) is obtained by seasonally adjusting each time a new observation is available, all the data available up to and including that observation. On the other hand, a forward seasonal estimate is obtained from a series that ended in the previous year. A common practise is to generate these forward seasonal estimates say, for year t+1, from data that ended in December of the previous year t.

There are four modes in which the X-11-ARIMA computer program can be applied to produce a current seasonally adjusted value. These four modes are: (I) using ARIMA extrapolations and concurrent seasonal factors, (II) using ARIMA extrapolations and forward seasonal factors; (III) without the use of ARIMA extrapolations and applying concurrent seasonal factors and (IV) without the use of ARIMA extrapolations and applying forward seasonal factors.

Statistical bureaus use the four modes to obtain current seasonally adjusted values but differ in their frequency. Thus, for example, the dominant mode in Statistics Canada is (I) followed by mode (III) whereas in the U.S. Bureau of Labor, the dominant mode is (II) followed by mode (IV). The current seasonally adjusted value produced by each type of seasonal adjustment varies and is subject to different degrees of error. Even within the same statistical agency there are frequent occurrences of different modes being applied to related series. This practice is not advisable since it can result in giving different signals on the short-term trend of the same type of series. Therefore, it is very important to decide the model by which similar or related series will be adjusted to avoid inconsistencies.

Under the assumption of an additive decomposition model, the seasonal adjustment of a current value X_t can be obtained by,

$$\hat{X}_t(\ell)=X_t-\hat{S}_t(\ell); \qquad (2.1)$$

where $\hat{S}_t(\ell)$ denotes a forward seasonal estimate.

or

$$\hat{X}_t(0)=X_t-\hat{S}_t(0) \qquad (2.2)$$

where $\hat{S}_t(0)$ denotes a conconcurrent seasonal estimate.

The current seasonally adjusted value will become "final" in the sense that it will no longer be revised after 42 more observations are added. Thus,

$$\hat{X}_t(m)=X_t-\hat{S}_t(m) \qquad (2.3)$$

where $\hat{S}_t(m)$ denotes a final seasonal estimate.

Therefore, the total revision of a concurrent and of a forward seasonal estimate can be written as,

$$r_t(0,m)=\hat{S}_t(0)-\hat{S}_t(m); \; m>0 \qquad (2.4)$$

$$r_t(\ell,m)=\hat{S}_t(\ell)-S_t(m); \; m>0>\ell \qquad (2.5)$$

Under the assumption of an additive decomposition and no replacement of extreme values, $S_t(m)$ can be expressed by

$$\hat{S}_t(m)= \sum_{j \; -m}^{m} h_{m,j} \, X_{t-j}=h^{(m)}(B)X_t; \qquad (2.6)$$

where $\hat{S}_t(m)$ is the final seasonal estimate from a series $X_{t-m},\ldots,X_t,\ldots,X_{t+m}$; and $h_{m,j}=h_{m,-j}$ are the symmetric moving average weights to be applied to the series. $h^{(m)}(B)$ denotes the corresponding linear filter using the backshift operator B, such that $B^n=X_{t-n}$. Young (1968) showed that the length of this symmetric filter $h^{(m)}(B)$, for monthly series, is 145 but that it can be well approximated by 85 weights because the values of the weights attached to distant observations are very small and, thus, m=42.

Following equation (2.6) we can express a concurrent seasonal estimate $S_t(0)$ and a forward seasonal estimate $\hat{S}_t^{(\ell)}$ by

$$\hat{S}_t(0) = \sum_{j=-2m}^{0} h_{0,j} X_{t-j} = h^{(0)}(B)X_t ; m=42 \quad (2.7)$$

where $h^{(0)}(B)$ denotes the asymmetric <u>concurrent</u> seasonal filter.

$$\hat{S}_t^{(\ell)} = \sum_{j=-2m}^{\ell} h_{\ell,j} X_{t-j} = h^{(\ell)}(B)X_t ; m=42 \quad (2.8)$$

where $h^{(\ell)}(B)$ denotes the asymmetric forward seasonal filter and ℓ = 1,2,...,12 for a monthly series.

The revision of a concurrent seasonal estimate $\hat{S}_t(0)$ depends on the distance between the concurrent and the final filter; that is, $d[h^{(0)}(B), h^{(m)}(B)]$ and the innovations of the new observations $X_{t+1}, X_{t+2}, ..., X_{t+m}$.

Similarly, the revision of a forward seasonal estimate $\hat{S}_t^{(\ell)}$ depends on $d[h^{(\ell)}(B), h^{(m)}(B)]$ and on the new innovations introduced by $X_{t-\ell+1}, ..., X_t, X_{t+1}, ..., X_{t+m}$

Theoretical studies by Dagum (1982a and 1982.b) have shown that,

$$d[h^{(0)}B, h^{(m)}(B)] < d[h^{(\ell)}(B), h^{(m)}(B)] \quad (2.9)$$

for ℓ=1,2,...,12.

The distance between the two filters is defined as the mean squared difference between the frequency response function of the filters over all the seasonal frequencies.

Relation 2.9 is true whether ARIMA extrapolations are used or not. Furthermore, the two studies also showed that if ARIMA extrapolations are used then

$d[h^{(0)}(B), h^{(m)}(B)]$ using ARIMA extrapolations (2.10)
$< d[h^{(0)}(B), h^{(m)}(B)]$ without ARIMA extrapolations.

and similarly,

$d[h^{(\ell)}(B), h^{(m)}(B)]$ using ARIMA extrapolations (2.11)
$< d[h^{(\ell)}(B), h^{(m)}(B)]$ without ARIMA extrapolations
for $\ell = 1,2,...,12$

Several studies by Dagum (1978), Bayer and Wilcox (1981); Kenny and Durbin (1982), McKenzie (1984), Dagum and Morry (1984); Pierce (1980) and Pierce and McKenzie (1985) have shown that

$$r^{(0,m)} < r^{(\ell,m)} \qquad (2.12)$$

except in a few cases where

$$r^{(0,m)} > r^{(\ell,m)} \qquad (2.13)$$

The relationship (2.13) can be observed when the current observations of the latest year are strongly revised since X_t gets the largest weight in the estimations of $\hat{S}_t^{(0)}$.

From the viewpoint of the total revisions of the seasonal estimates, the results of the above empirical studies permit to rank the four modes as follows: Mode I (ARIMA extrapolations with concurrent seasonal estimates) gives the smallest total revision; Mode III (no ARIMA extrapolations with concurrent seasonal estimates) ranks second; Mode II (ARIMA extrapolations with forward seasonal estimates) ranks third and Mode IV (ARIMA extrapolations with forward seasonal estimates) ranks fourth.

It is important that users are aware of the relative quality of each adjustment to produce the most reliable (minimum revision) seasonally adjusted estimates. However, it is sometimes possible that operational and cost considerations only allow for a suboptimal solution such as the application of forward seasonal estimates to produce seasonally adjusted current values. If this is the case, it is imperative that the same approach be followed by all users of related series to ensure consistency.

3. SMOOTHING OF HIGHLY IRREGULAR SEASONALLY ADJUSTED SERIES

One of the main purposes of the seasonal adjustment of economic time series is to provide information on current economic conditions, particularly to determine the stage of the cycle at which the economy stands. Since seasonal adjustment means removing only seasonal variations, thus leaving trend cycle estimates together with irregular fluctuations, it is often difficult to detect the short term trend or cyclical turning points for series strongly affected by irregulars. In such cases, it may be preferable to smooth the seasonally adjusted series using trend-cycle estimators which suppress as much as possible the irregulars without affecting the cyclical component.

The use of trend-cycle estimates or of smoothed seasonally adjusted data instead of a highly irregular seasonally adjusted series have been discussed by several writers and recently by Moore et als (1981), Kenny and Durbin (1982), Maravall (1986), and Dagum and Laniel (1987). Although not yet practised widely, some statistical agencies such as Statistics Canada and the Australian Bureau of Statistics smooth some of their series.

The statistical properties of the trend-cycle estimators (filters) available in X-11-ARIMA can be studied looking at the corresponding frequency response functions. The frequency response functions of a filter is defined by,

$$\Gamma^{(m)}(\omega) = \sum_{j=-m}^{n} y_{m,j} \exp(-i2\pi\omega j) \qquad 0 \leq \omega \leq 1/2 \qquad (3.1)$$

where $y_{m,j}$ are the weights of the filter and ω is the frequency in cycle per time unit. In general, the frequency response function may be expressed in polar form by

$$\Gamma(\omega) = A(\omega) + iB(\omega) = G(\omega)\exp[i\phi(\omega)] \qquad (3.2)$$

where $G(\omega) = [A^2(\omega) + B^2(\omega)]^{1/2}$ is called the gain of the filter and $\phi(\omega) = \arctan[B(\omega)/A(\omega)]$ is called the phase shift of the filter and is expressed in radians. The expression (3.2) for $\Gamma(\omega)$ shows that if the input function is a sinusoidal variation of unit amplitude and a phase shift $\psi(\omega)$, where $\psi(\omega)$ is linear, the output function will also be sinusoidal but of amplitude $G(\omega)$ and phase shift $\psi(\omega)+\phi(\omega)$. The gain and phase shift vary with the frequency ω. For symmetric filters, the phase shift is 0 or $\pm\Pi$ and for asymmetric filters it can take any value between $\pm\Pi$ being undefined at those frequencies where the gain is 0.

The combined linear filters applied to the original series to generate a central (symmetric) trend-cycle estimate have been calculated by Young (1968) for Census Method II-X-11 variant. This filter is similar to that of X-11-ARIMA with and without ARIMA extrapolations. Dagum and Laniel (1987) extended Young's (1968) results to include the estimation of the asymmetric trend cycle filters of X-11-ARIMA with and without the ARIMA extrapolations.

Figure 1 shows the gain functions of the central (symmetric) seasonal adjustment filters and smoothed seasonally adjusted data (trend-cycle) filters. It is apparent that the trend-cycle filters suppress all the noise present in the series, where the noise is defined as the power present in all frequencies $\omega \leq .166$. This frequency corresponds to the first harmonic of the fundamental seasonal frequency of a monthly series. This pattern results from the convolution of the seasonal adjustment filters with the 13-term Henderson trend-cycle filter.

(Please place Figure 1 about here)

Figure 2.a shows the gain functions of the concurrent and first-month revised trend-cycle filters of X-11-ARIMA without ARIMA extrapolations. Figure 2.b shows their corresponding phase-shift functions expressed in months instead of radians. We can observe that the gain for all $\omega<.167$ is much larger for these two asymmetric filters as compared with the central filter. Furthermore, there are large amplifications for frequencies near the fundamental seasonal. All this means that the concurrent and first revised smoothed seasonally adjusted values will have more noise than the final estimates. On the other hand, it is apparent that the phase shifts are very small, less than one month for the most important cyclical frequencies $0\leq\omega<.055$ (i.e. cycles of periodicities equal to and longer than 18 months).

(Please place Figures 2.a and 2.b about here)

Figure 3.a and 3.b show the gain and phase-shift functions of the concurrent and first-month revised trend-cycle filters of X-11-ARIMA with ARIMA extrapolations. The extrapolations are obtained from an IMA model $(0,1,1)(0,1,1)_{12}$ with $\theta=.40$ and $\Theta=.60$. The gain functions are closer to the symmetric (central) filter than those of X-11-ARIMA without the ARIMA extrapolations. There are no amplifications around the fundamental seasonal frequency and a similar attenuation of power at higher frequencies. On the other hand, there is more phase-shift (being near to one month) for low frequencies and less phase shift for all high frequencies.

(Please place Figures 3.a and 3.b about here)

All this suggests that the smoothed seasonally adjusted estimates of highly irregular series will give more reliable signals of the short term trend than the seasonally adjusted values and even in the smoothed series gains in quality are apparent if ARIMA extrapolations are used.

4. THE SEASONAL ADJUSTMENT OF AGGREGATED TIME SERIES

Many time series are aggregates of a certain number of component series. Macro-economic indicators, for instance, can often be broken down by industry, region, commodity, etc. or as for the unemployment rates by age, sex, region, etc. This poses the crucial problem of seasonal adjustment of aggregated or composite time series and raises the following questions:

Should the original data of the composite series be seasonally adjusted (the direct method) or should the original data corresponding to each of the component series be seasonally adjusted before aggregation (the indirect method)?
Do the direct and indirect methods coincide?
What is the criterion for choosing between these two procedures?
When the indirect method is used, should there be a partial aggregation of the component series before seasonal adjustment?
What are the criteria to identify the optimal partial aggregation?

These questions leave no doubt about the complexity of the seasonal adjustment of most key indicators. It is also apparent that inconsistencies can occur when different approaches are followed for similar or related series.

4.1 The direct versus the indirect method

Which of the two methods should be used to seasonally adjust aggregated series is a question with no straightforward answer but is fundamental since the two methods give in general different results. This is because of the non-linearities in the X-11-ARIMA program which can only be eliminated if the following conditions are satisfied: (1) all the component series as well as the aggregate are seasonally adjusted with the additive model; (2) no replacement of extreme values is made; and (3) the variable trend-cycle curve is kept constant during all the iterations. From a seasonal point of view, postulating in all cases an additive decomposition model is inappropriate, and condition 2 is not reasonable in most cases.

From a statistical viewpoint, the indirect method is generally more efficient than the direct one unless the seasonal pattern of the aggregate series is easily detectable whereas the seasonal pattern of some of the important component series is not (due, for instance, to a strong irregularity and the presence of outliers) (Pfeffermann, Salama and Ben Tuvia, 1984). The indirect method would also be generally preferable for most indices and percentage rates such as the unemployment rates (Dagum 1979), or when some of the component series have an additive structure and the others have a multiplicative one. The indirect procedure then ensures that the correct decomposition model is used for each series. On the other hand, the direct method is advantageous for example when the component series have a similar seasonal pattern. The seasonal factors obtained from the composite series can be used to adjust the component series. Then the adjusted composite series coincide with the aggregated adjusted series which is often desirable from theuser's point of view. The indirect method can also achieve this goal.

In most practical cases, however, a selection criterion is applied by running the X-11-ARIMA program which tests the smoothness of both the direct and the indirect aggregated series

(Lothian and Morry, 1977 and Dagum, 1979). Since the main purpose of seasonal adjustment is to determine, for instance, the stage of the cycle at which the economy stands, the goal is then to have a seasonally adjusted aggregated series that approximate as much as possible the trend-cycle curve. The trend and cyclical variations are assumed to be "smooth functions of time". Thus the selection of the better method is based on the degree of smoothness of both the adjusted composite series and the aggregated adjusted series. Spectral analysis is another way of finding out which one of the two methods is better at removing seasonality and smoothing the series.

4.2 The basic elementary unit of an aggregate, i.e., the partial aggregation problem

The widely used indirect procedure poses a problem. Consider the unemployment rate, the components of which are grouped, for instance, by age and sex. The partial aggregation problem is to collapse together the component series which will enable a more accurate identification of seasonal variations. In this example, male and female teenage employment can be combined, if they show similar seasonal patterns, to form what we shall call a basic elementary unit of the aggregate. On the other hand, it might be inappropriate to combine adult males and females unemployment if their seasonal patterns are non similar, each one of the two series is then by itself a basic elementary unit.

Any possible combination of basic elementary units will give the same raw unemployment rate but different seasonally adjusted unemployment rates. A large discrepancy between the adjusted series is a serious problem. Therefore, it is essential to identify the better set of basic elementary units.

Dagum (1979) has presented a set of four properties that an economic time series must have to be considered as a basic elementary unit.

Property 1:

The seasonal component of a basic elementary unit must be identifiable, i.e., the seasonal variable must be large relative to the irregular and/or the cycle. The X-11-ARIMA program tests series for the presence of seasonality. A similar analysis can be made by looking at the correlogram or the spectrum of the basic elementary unit series.

Property 2:

The seasonal pattern of a basic elementary unit must be reduced to its simplest shape and/or result from seasonal causes that belong to a homogeneous class. This does not preclude the partial aggregation of all series that have two troughs and peaks generated by two seasonal causes. For example, adult females of different age groups might have more unemployment in winter due to slowdowns in economic activity and more unemployment at the end of the summer due to other factors such as children education.

Property 3:

The seasonal pattern of basic elementary units with seasonals having similar amplitudes and the "same" timing in their peaks and troughs must be aggregated (confer to the example given in property 2).

Property 4:

A basic elementary unit should represent an observable economic phenomenon and not a mathematical abstraction. This implies that most of indices and percentage rates such as the unemplooyment rates cannot be considered as basic elementary units for seasonal adjustment. The direct procedure is thus inappropriate in these cases.

Note that the component series of a macro-economic indicator are selected for their economic content and are related to a given observable economic phenomenon. On the one hand, the four properties prevent any loss of information that may result from partial aggregation. On the other hand, the organization of basic units into subsets according to the four properties is legitimate and does not place statistics before economics since the component series were initially selected because of their intrinsic economic relevance.

5. THE ESTIMATION OF TRADING DAY VARIATION

Business and economic time series which are flows, in the sense that they result from the accumulation of daily values over the calendar month, are affected by calendar variations. In addition to seasonal influences these calendar variations also need to be removed from the series to obtain a clear signal of the short-term trend. There are two types of calendar variations; one which is due to the changing number of times each day of the week occurs in a month, normally referred to as trading day variation, the other is caused by the variable timing of certain holidays with respect to the calendar of which Easter has the largest impact.

Trading day variations arise mainly because the activity varies with the days of the week. Other sources are associated with accounting and reporting practices for example, stores that do their bookkeeping activities on Friday tend to report higher sales in months with five Fridays than in months with four Fridays.

The trading day effects are estimated in the X-11-ARIMA program using ordinary least squares on a simple deterministic regression model developed by Young (1965) and first introduced into the Census Method II X-11 variant (Shiskin, Young and Musgrave, 1967).

In Young's approach, the following model is fitted to the series I_t (this is the series from which the trend-cycle and seasonal fluctuations have already been removed):

$$I_t = TD_t + \varepsilon_t \qquad (5.1)$$

where it is assumed: $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$.

The trading day variation TD_t can be expressed as:

$$TD_t = \sum_{i=1}^{6} \delta_i \ T_{it} \qquad (5.2)$$

where

$$\delta_i = \xi_i - \xi \text{ and } T_{it} = Y_{it} - Y_{7t}$$

the δ_i 's represent the difference between the Monday, Tuesday,...,Saturday effects ξ_i and the average daily effect ξ . Y_{it} denotes the number of times a given day i of the week occur in month t and Y_{7t} denotes the number of times that Sunday occurs in the month t. Here the δ_i's are fixed and estimated using ordinary least squares. As long as the relative weight of daily activities is stable throughout the length of the time series, this approach produces reasonable estimates. However, this is not always a realistic assumption. It is well known that, for example, in retail trade, store opening hours and consequently, consumer shopping patterns have changed over the last decade; there are much more retail outlets open on Sundays and even during the week stores keep longer hours than previously.

This problem can be partially overcome by estimating the regression coefficients based on the last six or seven years of data only rather than on the full span of the series to reflect more recent weight patterns. However, if the daily weight structure changes rapidly the estimated weights, which describe the average weight structure over the last six or seven years, will statistically distort the actual trading day variation present in the current observations.

To overcome this limitation of the method, we are presently investigating the possibility of introducing two other trading day models in the X-11-ARIMA program which would allow for a stochastic behaviour of the daily coefficients instead of the presently assumed deterministic one.

One model, already proposed by Monsell (1983) assumes that the generating process of the 's is a random walk. That is,

$$TD_t = \sum_{i=1} \delta^1{}_{it}\, T_{it} \tag{5.3}$$

where

$$\underset{\sim}{\delta}^1_t = \underset{\sim}{\delta}^1_{t-1} + \underset{\sim}{\chi}_t$$

and

$$\underset{\sim}{\delta}^1_t = (\delta^1_{1t}, \delta^1_{2t}, \ldots \delta^1_{6t}); \quad \underset{\sim}{\chi}_t \sim NID\,(0, \sigma^2_\chi I_6) \tag{5.4}$$

The second stochastic model assumes that the vector of daily coefficients follows a random walk model with a random drift. That is,

$$TD_t = \sum_{i=1}^{6} \delta^2{}_{it}\, T_{it} \tag{5.5}$$

$$\underset{\sim}{\delta}^2 = \underset{\sim}{\delta}^2_{t-1} + \underset{\sim}{\rho}_t + \underset{\sim}{\chi}_t \tag{5.6}$$

$$\underset{\sim}{\rho}_t = \underset{\sim}{\rho}_{t-1} + \underset{\sim}{\psi}_t \tag{5.7}$$

Equations (5.5), (5.6) and (5.7) provide a local approximation to a linear trend in the daily coefficients. The level and slope of the trend are assumed to be generated by stochastic processes.

Error vectors $\underset{\sim}{\chi}_t$ and $\underset{\sim}{\psi}_t$ are assumed to be mutually independent and $NID(0,\sigma^2_\chi\ I_6)$ and $NID(0,\sigma^2_\psi\ I_6)$ respectively. These two stochastic models are written in state-space form and the estimates $\hat{\underset{\sim}{\delta}}{}^i_t$, i=1,2 of $\underset{\sim}{\delta}^i_t$, together with their mean square error matrices, are estimated with the Kalman filter and fixed interval smoother. Maximum likelihood estimators are used to estimate the remaining hyper-parameters σ^2 $\sigma_\chi{}^2/\sigma^2$ and σ^2_ψ/σ^2

If these models will be made available to users it is crucial to provide guidelines (probably based on the signal to noise ratio present in the data), concerning the most suitable model for a given series.

Finally whenever there is trading day variation present in a series, whether of a deterministic or stochastic nature, it is important that this variation be removed before fitting an ARIMA model to the series in the extrapolating phase of the X-11-ARIMA seasonal adjustment procedure, since trading day movements would not be picked up adequately by the ARIMA model.

6. ISSUES CONCERNING THE ESTIMATION OF EASTER EFFECTS

As mentioned previously, there is another type of calendar variation associated with moving holidays whose effect needs to be removed from the time series to avoid distortion in the month-to-month movement. Liu (1980) describes how the varying placement of the Chinese New Year plays a role in modelling time series. Pfefferman and Fisher (1980) discuss the impact of religious festivals in Israel, based on the Hebrew calendar, in economic time series compiled according to the Gregorian calendar. In the western world the most important example of holiday variation is associated with the Easter holiday which may occur as early as March 22 or as late as April 25.

There has been a renewed interest in Easter adjustment recently due to the impact of the relatively early arrival of Easter (March 28) in 1986. Since this holiday has not fallen in March since 1978, in time series affected by Easter variation, the March to April movement of 1986 was not in line with those observed in previous years for a while. In international trade statistics, especially in the value of imports, a drop was noted in March followed by a rise in April caused by processing the end of March forms only in April due to the Easter closing of customs offices. New car dealers reported relatively low sales in volumes in March of 1986 because of not being open for business during Easter. These two are examples of series in which the presence of Easter is accompanied by a decrease in level. The opposite effect can be observed in some other series; for example, the number of marriages shows a marked increase during Easter. In these cases whether the effect is positive or negative, the impact is immediate, i.e. only the holiday period displays changed activity.

There is another category of Easter variation which manifests itself in an increased or decreased activity not only during the holiday but days and sometimes even weeks before it. This type of gradual impact occurs in the sales of chocolates, flowers, women's clothing etc. It follows that in cases like this the Easter effect will not only depend on whether the holiday fell in March or in April but it also depends on the date in April. An early April date of Easter will influence the March figures to some extent (the earlier the date the higher the effect) if the build-up period in activity covers also the end of March. Thus we distinguish between two kinds of Easter variation one with immediate impact and one with gradual impact. Accordingly, the estimation of Easter effect will depend on which model applies to a given series.

In the context of X-11-ARIMA, there are two basic approaches to estimate Easter effects, those which use the original data (adjusted for trading day variation) for building a regression model and those which base the regression on the irregulars obtained in a previous seasonal adjustment. The holiday adjustment proposed by Baron (1973) falls into the first category while the OECD method and the technique described by Pfefferman and Fisher (1980) belong to the second category.

When the estimation is based on the original data normally the proportion of the March values to the sum of March and April values is regressed against the date of Easter calculated as the number of days between the date and the earliest possible Easter date. The limitation of this approach is that it assumes a constant seasonal factor for each month throughout the entire period and it also ignores possible large trend movements that can invalidate the estimates.

This drawback in the method is basically eliminated through the second approach, namely by using the estimated irregulars rather than the original series in the regression model. Even here some of the Easter effect could theoretically contaminate the estimation of the seasonal and trend-cycle component in the first stage but experience with real and simulated data shows that this is not a serious problem. Thus, the second approach is considered more appropriate for the majority of time series.

The estimation procedure depends on the nature of Easter variation present for it implies different Easter effect models. Since the estimation of Easter effects is currently not available in the X-11-ARIMA program, we describe in detail the two types of models that will be incorporated in the near future.

6.1 Immediate Impact Model

Let I_{ij} be the irregular component estimated in a first run of the X-11-ARIMA seasonal adjustment. Here i denotes the year and j refers to the month of March. (Correspondingly j+1 refers to the month of April.)

Using the additive decomposition model the irregular $I_{i,j}$ and $I_{i,j+1}$ can be further decomposed as:

$$I_{i,j} = -E_i + \varepsilon_{i,j} \qquad (6.1)$$

$$I_{i,j+1} = E_i + \varepsilon_{i,j+1}$$

where E_i denotes the Easter effect in year i and ε_{ij} is assumed to be an i.i.d random variable.

Let Z_i be defined as the number of days between Easter Sunday in year i and March 22, the earliest possible Easter date.

Let $f(Z_i)$ be the indicator defined as

$$f(Z_i) = 1 \quad \text{if } Z_i \leq 9 \text{ (Easter falls in March)} \qquad (6.2)$$

$$f(Z_i) = 0 \quad \text{if } Z_i \geq 9 \text{ (Easter falls in April)}$$

then

$$E_i = 1/2f(Z_i) \left[\frac{\sum_{i \in M} (I_{i,j+1} - I_{i,j})}{n_M} - \frac{\sum_{i \in A} (I_{i,j+1} - I_{i,j})}{n_A} \right] \qquad (6.3)$$

M - is the subset of years when Easter fell in March

n_M - is the number of years when Easter fell in March

A - is the subset of years when Easter fell in April

n_A - is the number of years when Easter fell in April.

Thus when the impact is immediate, the Easter effect can be described by a step function.

6.2 Gradual Impact Model

Here we will introduce the concept of early April Easter by which we mean an Easter date which is before the k-th of April, where k is the number of days before Easter whose activity is affected by Easter. Correspondingly, an Easter date on or after the k-th of April will be referred to as a late April Easter. In practice, k varies between 3 and 10 depending on the series.

It is assumed that the effect increases (or decreases) linearly during the k days before Easter. According to this model, equation (6.1) still holds but equations (6.2) and (6.3) need to be modified as follows:

$$
\begin{aligned}
f(Z_i) &= 1 && \text{if } Z_i \leq 9 \\
f(Z_i) &= k+9-Z_i && \text{if } 9 < Z_i < k \\
f(Z_i) &= 0 && \text{if } Z_i \geq k
\end{aligned}
\qquad (6.4)
$$

and

$$E_i = 1/2\, f(Z_i) \left[\frac{\sum_{i \in M} (I_{i,j+1} - I_{i,j})}{n_M} - \frac{\sum_{i \in LA} (I_{i,j+1} - I_{i,j})}{n_{LA}} \right] \qquad (6.5)$$

where LA is the subset of late April Easter years (on or after the k-th of April.

n_{LA} is the number of late April Easter years.

Here the effect follows a step function with a linear segment in the middle, of the form _/‾ or ‾_, the value depending on the Easter date Z_i. In the estimation of the Easter effect the sloping linear segment is defined as the line joining the horizontal line fitted to the March Easter values and the horizontal line fitted to the late April Easter values. This formulation was chosen over fitting a separate regression line to the early April values because normally there would be too few points available in this interval to produce a reliable estimate of the slope of the line.

The build-up period k will usually be provided by subject-matter experts. For example, Easter induced sales activity would not start earlier than 3 or 4 days before the holiday in the case of perishable items such as flowers but for women and children clothes a longer build-up period would apply.

If there is no prior information available about the length of the build-up period k users can take advantage of an option to be introduced into the X-11-ARIMA program to estimate this length from the data. The estimation will be based on the difference between the April and March irregulars during April Easter years plotted against the Easter date. A sloping linear segment and a horizontal segment will be fitted to these differences in accordance with model (6.5). The length k of the sloping linear segment will vary from 0 to 10 and the sum of squared differences between the fitted model and the observed irregular differences will be calculated for each k. The optimal build-up period will be determined by that value of k for which these sum of squares is the minimum. The formulas for this calculation can be found in Appendix A.

If the decomposition of the series is multiplicative, the estimation procedure for E_i is identical for both models to the additive case the only difference is that the irregulars $I_{i,j}$ and $I_{i,j+1}$ are assumed to be of the form:

$$I_{i,j} = (1-E_i) \times \varepsilon_{i,j} \tag{6.6}$$

$$I_{i,j+1} = (1+E_i) \times \varepsilon_{i,j+1}$$

Similarly to the treatment of trading day variation, the estimated Easter effect also needs to be removed from the series before the ARIMA modelling phase of the X-11-ARIMA procedure to improve the overall seasonal adjustment quality.

APPENDIX A

Determining the optimal build-up period k_{opt}

Notation:

M - subset of March Easter years

EA(k) - subset of early April Easter years, i.e. years in which Easter occurred between the 1st and kth of April

LA(k) - subset of late April Easter years, i.e. years in which Easter occurred on or after the k-th of April

$n_{EA(k)}$ - number of early April Easters

$n_{LA(k)}$ - number of late April Easters

$I_{i,j}$ - irregular in year i month of March

$I_{i,j+1}$ - irregular in year i month of April

Z_i - the number of days between Easter in year i and March 22 (the earliest possible Easter date)

Let

$$D_i = I_{i,j+1} - I_{i,j}$$

and

$$\bar{D} = \frac{\sum_{i=1}^{n} (I_{i,j+1} - I_{i,j})}{n}$$

For k = 0,1,...,10 calculate the following sum of squares:

$$error2_{LA(k)} = \sum_{i=1}^{n_{LA(k)}} (D_{LA(k)i} - \bar{D}_{LA(k)})^2$$

$$error2_{EA(k)} = \sum_{i=1}^{n_{EA(k)}} \left\{ D_{EA(k)_i} - \left[\bar{D}_M + \frac{Z_i - 9}{k} (\bar{D}_{LA(k)} - \bar{D}_M) \right] \right\}^2$$

Then the optimal build-up period k_{opt} is defined as:

$$k_{opt} = k \ \min_{k}(\text{error}^2_{EA(k)} + \text{error}^2_{LA(k)})$$

Thus the optimal buildup period is the one for which the total error sum of squares is minimum.

REFERENCES

Baron, R. (1973): "Analysis of Trend and Seasonality in Statistical Series, Methodology and Applications in Israel", Israel Central Bureau of Statistics, Technical Paper No. 39.

Bayer, A. and Wilcox D. (1981): "An Evaluation of Concurrent Seasonal Adjustment" Technical Report, Washington, D.C. Board of Governors of the Federal Reserve System, Special Studies Section.

Dagum, E.B. (1978): Comparison and Assessment of Seasonal Adjustment Methods for Labour Force Series, Washington, D.C.: U.S. Government Printing Office, Stock No. 052-003-00603-1.

Dagum,E.B. (1979): On the Seasonal Adjustment of Economic Time Series Aggregates: A Case Study of the Unemployment Rate. National Commission on Employment and Unemployment Statistics, Background paper No. 31, Washington, D.C.

Dagum, E.B. (1980): The X-11-ARIMA Seasonal Adjustment Method, Ottawa: Statistics Canada Catalogue No. 12-564E.

Dagum, E.B. (1982.a): "Revisions of Time Varying Seasonal Filters" Journal of Forecasting, 1, pp. 173-187.

Dagum, E.B. (1982.b): "The Effects of Asymmetric Filters on Seasonal Factor Revisions" Journal of the American Statistical Association, Vol. 77, No. 380, pp. 732-738.

Dagum, E.B. (1982.c): "Revisions of Seasonally Adjusted Data Due to Filter Changes" Proceedings of the Business and Economic Section, American Statistical Association pp. 39-45.

Dagum, E.B. and Morry, (1984): "Basic Issues on the Seasonal Adjustment of the Canadian Consumer Price Index" Journal of Business and Economic Statistics, vol. 2, No. 3, pp. 250-259.

Dagum, E.B. and N. Laniel (1987): "Revisions of Trend-Cycle Estimators of Moving Average Seasonal Adjustment Method" Journal of Business and Economic Statistics.

Kenny, P. and Durbin, J. (1982): "Local Trend Estimation and Seasonal Adjustment of Economic Time Series" Journal of the Royal Statistical Society, Ser. A vol. 145, Part 1, pp 1-41.

Liu, L.M. (1980): "Analysis of Time Series with Calendar Effects", Management Science, 26, pp. 106-112.

Lothian, J. and Morry, M. (1977): "The Problem of Aggregation; Direct or Indirect", Research Paper, Time Series Research and Analysis Division, Statistics Canada.

Maravall, A. (1986): "An Application of Model-Based Estimation of Unobserved Components" International Journal of Forecasting, 2, N.3, pp. 305-318.

McKenzie, S. (1984): "Concurrent Seasonal Adjustment with Census X-11" Journal of Business and Economic Statisdtics, Vol. 2, No. 3, pp 235-249.

Moore, G.H., Box, G.E.P., Kaitz, H.B., Stephenson, J.A. and Zellner, A. (1981): "Seasonal Adjustment of the Monetary Aggregates" Report of the Committee of Experts on Seasonal Adjustment Techniques, Washington, D.C. Board of Governors of the Federal Reserve System.

Monsell, B.C. (1983): "Using the Kalman Smoother to adjust for Moving Trading Day" Research Report SED/RR-83/04 U.S. Department of Census.

Pfeffermann, D., and Fisher, J. (1980): "Festival and Working Days Prior Adjustments in Economic Time Series," in Time Series, ed. O.D. Anderson, New York; North-Holland.

Pfeffermann, D., Salama, E. and Ben-Tuvia, S. (1984): "On the Aggregation of Series: A New Look at an Old Problem", Research Paper, Central Bureau of Statistics, Jerusalem, Israel.

Pierce, D.A. (1980): "Data Revisions with Moving Average Seasonal Adjustment Procedures" Journal of Econometrics, 14, pp. 95-114.

Pierce, D. and McKenzie, S. (1985): "On Concurrent Seasonal Adjustment" Special Studies Paper, No. 164, Federal Reserve Board, Washington, D.C.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967): "The X-11-Variant of Census Method II Seasonal Adjustment" Technical Paper No. 15, Bureau of Census, U.S. Department of Commerce.

Young, A.H. (1965): "Estimating Trading Day Variation in Monthly Economic Time Series", Technical Paper No. 12, Bureau of Census, U.S. Department of Commerce.

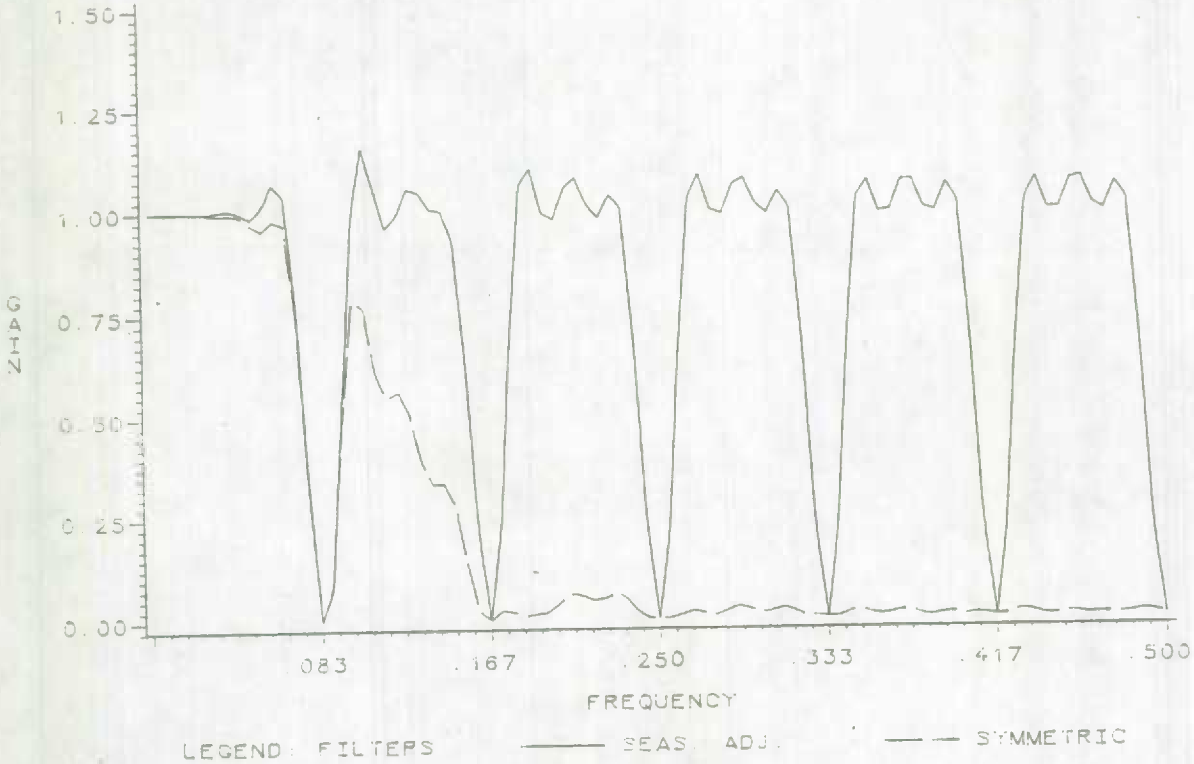

FIGURE 1: Gain Functions of the Central (Symetric) Trend–Cycle and Seasonal Adjustment Filters OF X–11–ARIMA

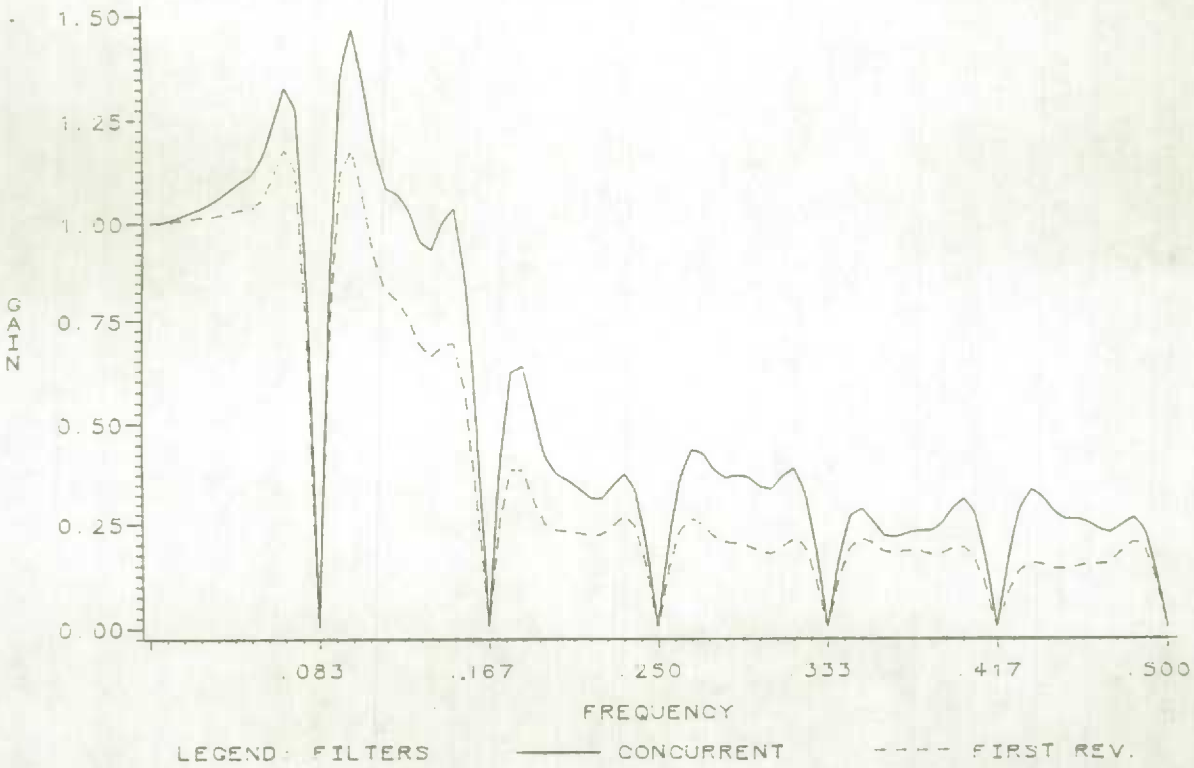

FIGURE 2.a: Gain Functions of the Concurrent and First-Month Revised Filters of X-11-ARIMA without ARIMA Extrapolations.

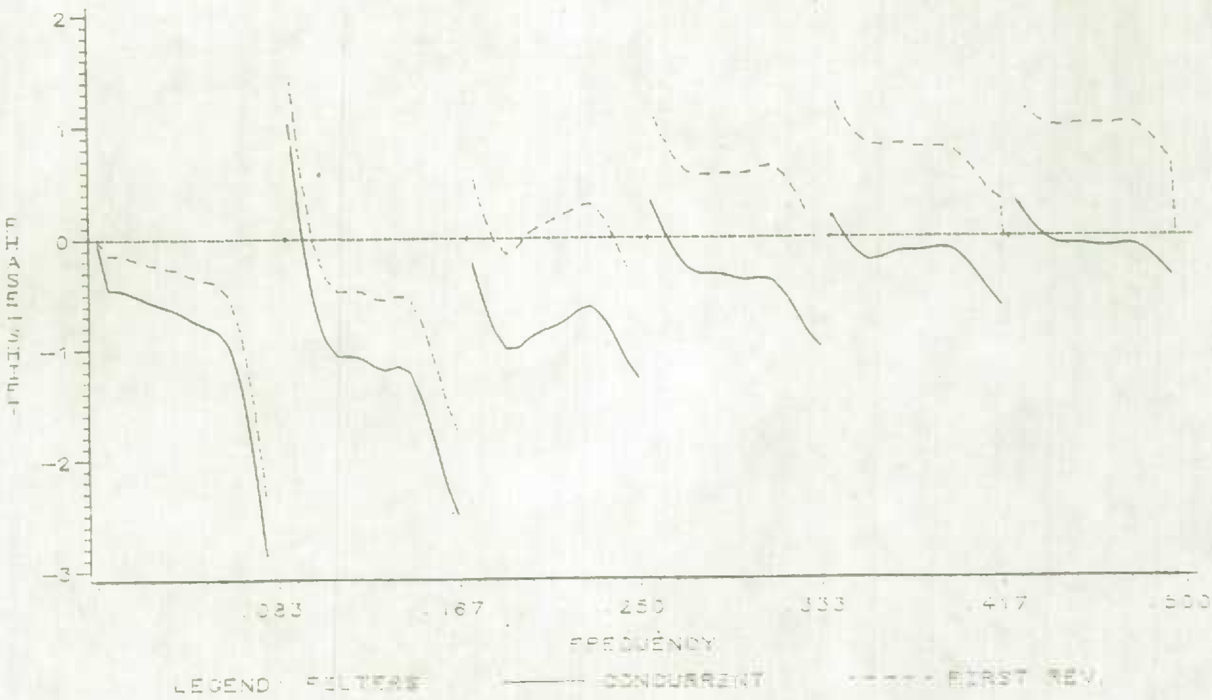

FIGURE 2.b: Phase-Shift Functions of the Concurrent and First-Month Revised Filters of X-11-ARIMA without ARIMA Extrapolations

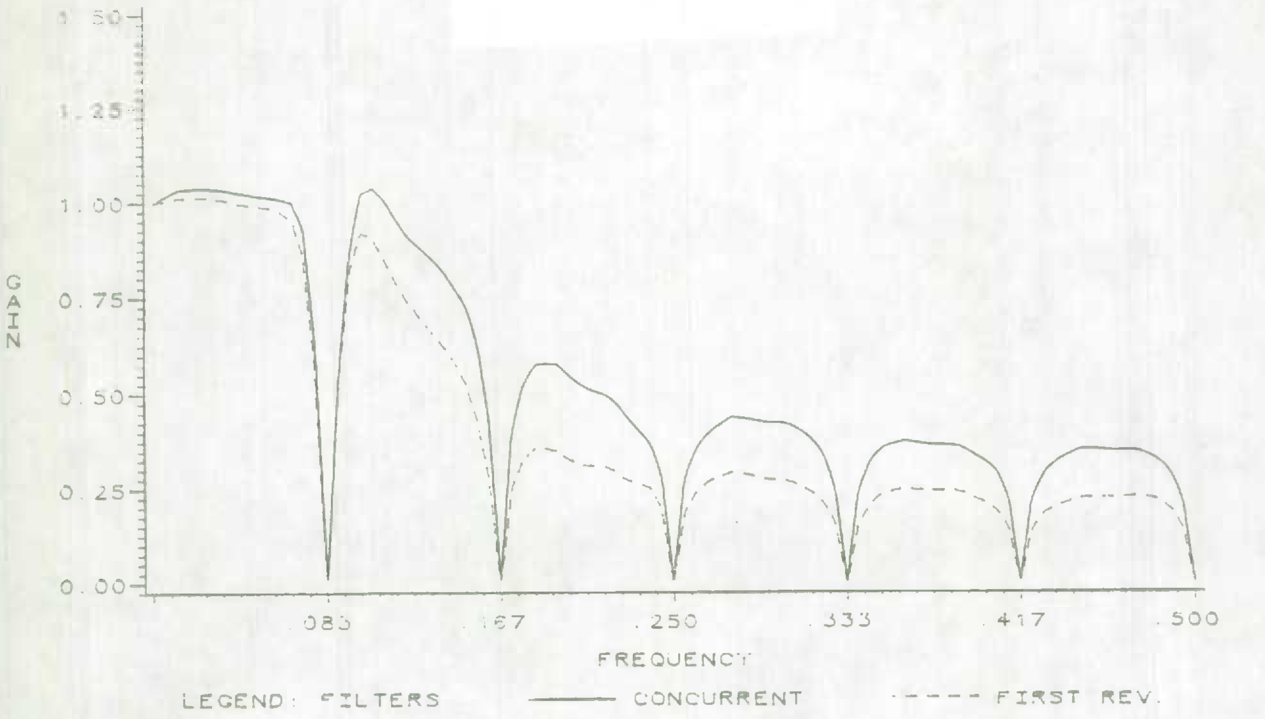

FIGURE 3.a: Gain Functions of the Concurrent and First-Month Revised Filters of X-11-ARIMA with ARIMA extrapolations. (θ=.40 Θ=.60)

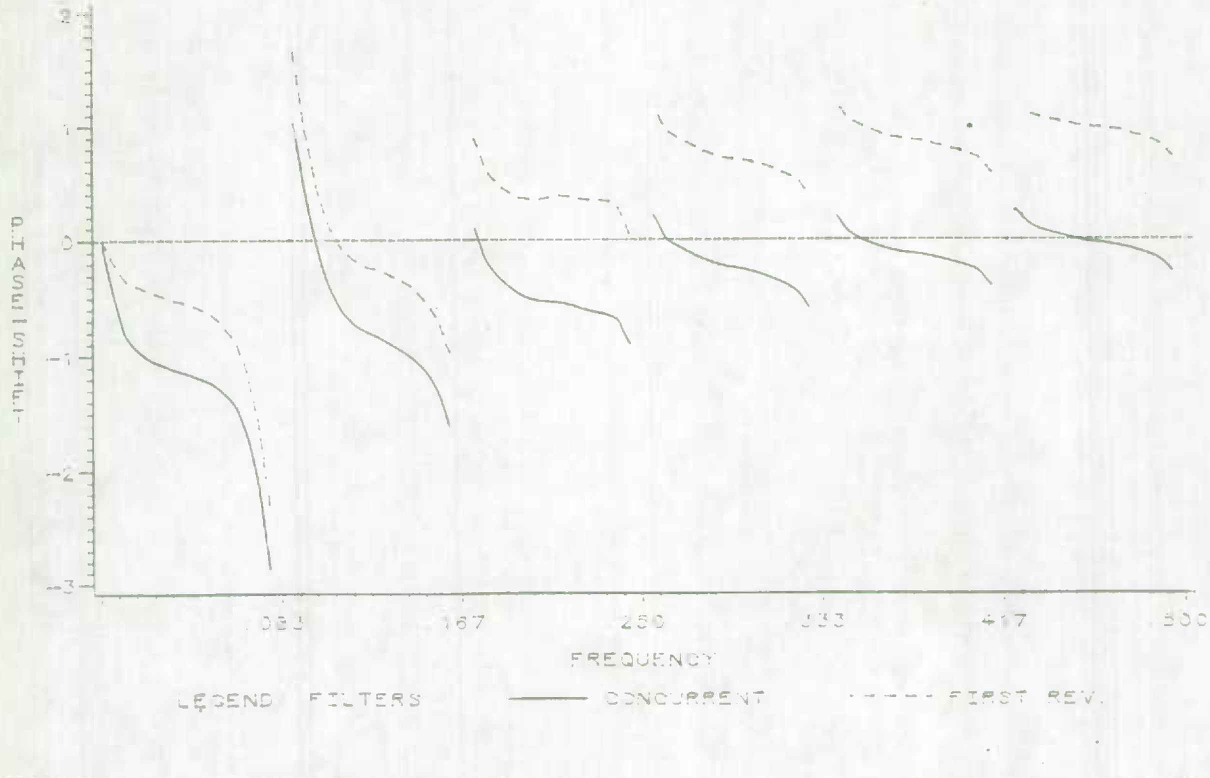

FIGURE 3.b: Phase-Shift Functions of the Concurrent and First-Month Revised Filters of X-11-ARIMA with ARIMA Extrapolations. (θ=.40 Θ=.60)