THE SURROGATE HENDERSON FILTERS IN X-11

MIKE DOHERTY

Statistics New Zealand

Summary

This paper explains the surrogate Henderson filters that are used in the X-11 variant of the Census Method II seasonal adjustment program to obtain trends at the ends of time series. It describes a prediction interpretation for these surrogate filters, justifies an approximation to the filters, proposed by Kenny & Durbin (1982), and proposes a further interpretation of the results. The starting point for the paper is unpublished work by Musgrave (1964a, 1964b). His work has continuing relevance to current seasonal adjustment practice. This paper makes that work generally available for the first time, and reviews and extends it.

Key words: asymmetric filters; de Forest extension; moving average; seasonal adjustment; trend.

1. Introduction

Problems arise in applying moving average filters to time series near the ends of the series, where the amount of data available is shorter than the length of the filter we want to use. The authors of the X-11 seasonal adjustment program faced and solved this problem for Henderson moving averages. (X-11 uses Henderson moving averages to obtain trend estimates.) However, to quote Wallis (1982a p. 30):

The technical manual for the X-11 program gives a set of asymmetric filters, again to be used when not enough data points are available for application of the symmetric filter, but the source of these is somewhat mysterious. Certainly, no reference is given by the authors of the technical manual. Likewise, it is not possible to reproduce the figures in the technical manual by adopting Henderson’s criteria — namely, designing a filter of the appropriate length to pass a cubic unaltered and minimise the variance of the third difference of the series.

Until the preprint version of the present paper was circulated (Doherty, 1992), the origin and explanation of these asymmetric filters remained a mystery to many working on seasonal adjustment. The filters, in fact, derive from unpublished work done at the Bureau of the Census in the 1960s by Musgrave (1964a, b).

In this paper I make Musgrave’s work generally available for the first time. I review and synthesize his work, relating it to subsequent work on the predictor interpretation of filter extension, and to an interpretation suggested by Kenny & Durbin (1982) and I offer suggestions to explain why Musgrave’s work has been successful. I have tried to write the paper I would have liked to have had available to me when I first started to puzzle over the tables of asymmetric Henderson filters in the X-11 manual. Some of what I say has been anticipated in other unpublished work by Laniel (1985), as explained in Section 2.

X-11 remains a benchmark against which to test other seasonal adjustment packages. Musgrave’s work is therefore still very relevant to current research, and Doherty (1992) has...
led to further research in the field. For example, it was one of the motivating influences on the work of Gray & Thomson (1996).

In the present paper, Section 2 gives further background and the history of the problem. Section 3 explains the Musgrave criteria which were implemented in X-11. I set these criteria in a more general context in Section 4. Section 5 describes the calculation of a constant in the formulae in terms of the criterion used for choice of filters, and Section 6 explains the Kenny–Durbin interpretation. Section 7 offers another interpretation by way of concluding remarks. Appendix A briefly reviews two needed results from the theory of affine prediction. Appendix B gives a derivation of the extension formulae used in Section 4.

2. History and background

The main reference on X-11 is Shiskin, Young & Musgrave (1967). For Henderson filters in general, the most accessible reference is the appendix to Kenny & Durbin (1982).

X-11 uses 5-term Henderson filters to produce trends for quarterly series, and 9-term, 13-term or 23-term Henderson filters to produce trends for monthly series.

The 5-term Henderson filters have apparently been extended by using the average of the last two actual values to stand in for the missing values. This sort of idea is used in X-11 for extending most of the seasonal filters (Shiskin & Eisenpress, 1957). The result is also essentially obtained by Wallis (1982b), although he makes the result more elaborate. From now on I exclude the 5-term Henderson filters from this account.

The 9-term, 13-term and 23-term surrogate Henderson filters in X-11 are derived in a very different manner, from criteria of John C. Musgrave in unpublished US Bureau of the Census papers (Musgrave, 1964a, b). Prior to this paper, the only place known to me in the published literature which stated this was Salzman (1968 p. 92 footnote 7).

Normand Laniel of Statistics Canada discovered, apparently without knowledge of Musgrave’s work, a criterion yielding the X-11 surrogate Henderson filters, and described this in the unpublished paper Laniel (1985). The criterion he found (which is the one originated by Musgrave) was quoted in 1988 in the X-11 ARIMA/88 seasonal adjustment program manual (Dagum, 1988). It was used in the X-11 ARIMA/88 program to obtain further significant digits for these filters.

Laniel’s work became known in some other statistical offices. For example, the Australian Bureau of Statistics used it to obtain surrogate 7-term, 15-term and 17-term Henderson filters. Their only published account, Castles (1987), gives no details or references, perhaps because it was aimed at a wide audience.

The present paper was first circulated in 1992 and is the work usually cited as Doherty (1992). I apologize to the seasonal adjustment community for the delay in publication.

The closed form I derive for the solution has been incorporated in the Bureau of the Census X-12-ARIMA program. The present paper was cited, and its results summarized, in Findley et al. (1998 Appendix B) which announced the new capabilities of X-12-ARIMA.

3. Musgrave’s work

Musgrave (1964a, b) notes that his work applies equally to extension of seasonal filters or trend filters. He wrote originally for the seasonal case. Translating to the trend case, his assumptions are:

(i) we have a set of filter weights \( w_1, \ldots, w_n \) with \( \sum_{i=1}^{n} w_i = 1 \);
(ii) we suppose our time series $X_i$ has the form

$$X_i = \alpha + \beta i + \epsilon_i,$$

where $\alpha$ and $\beta$ are constants, and the $\epsilon_i$ are independent identically distributed errors with mean 0 and constant variance $\sigma^2$.

For $m < n$, we want a set of weights $u_1, \ldots, u_m$ with $\sum_{i=1}^m u_i = 1$, that minimizes the mean square revision between the initial and the final trend estimates; i.e. such that $E((\sum_{i=1}^m u_i X_i - \sum_{i=1}^n w_i X_i)^2)$ is a minimum, where $E$ denotes expectation. To avoid misunderstanding, note that I am not taking expectations conditional on the initial $X_1, \ldots, X_m$.

The resulting surrogate weights are

$$u_r = w_r + \frac{1}{m} \sum_{i=m+1}^n w_i + \left(\frac{r - \frac{m+1}{2}}{\frac{m(m-1)(m+1)}{12} \beta^2 \sigma^2} \right) \sum_{i=m+1}^n \left(\frac{1}{2}\right) w_i.$$  \hfill (1)

With choices for the parameters $\beta^2 / \sigma^2$ (see Section 5) this is the formula implemented in X-11 for surrogate Henderson 9-term, 13-term and 23-term filters. (To be strictly accurate, the formula is not in Musgrave (1964a, b), where the equations were solved numerically. I do not know if the results in the X-11 manual came from numerical solution of the equations or a form of the above formula.)

To understand the formula better, and relate it to more recent work, it is helpful to go to a slightly more general context.

4. Surrogate filters and minimum revision

Suppose, as before, we are given a filter $w_1, \ldots, w_n$ with $\sum_{i=1}^n w_i = 1$, and that we seek a filter $u_1, \ldots, u_m$ with $\sum_{i=1}^m u_i = 1$, where again $m < n$. Suppose the series $X_i$ locally (i.e. for $i = 1, \ldots, n$) has finite first and second moments, and the $X_i$ are independent. Suppose the $X_i$ have constant variance $\sigma^2$.

We again choose the $u_i$ to minimize the mean square revision; i.e. the $u_i$ are such that $E((\sum_{i=1}^m u_i X_i - \sum_{i=1}^n w_i X_i)^2)$ is a minimum. Suppose $T$ is the affine subspace formed of all linear combinations $\sum_{i=1}^m d_i X_i$ for scalar $d_i$ with $\sum_{i=1}^m d_i = 1$. Then in seeking the $u_i$ above, we are equivalently asking for a $T$ in $T$ such that $E((T - \sum_{i=1}^n w_i X_i)^2)$ is a minimum. Results from the theory of affine prediction, recalled in Appendix A, show the $T$ we seek has the form

$$T = \sum_{i=1}^n w_i \hat{X}_i,$$ \hfill (2)

where $\hat{X}_i$ is the optimum predictor in $T$ of $X_i$, i.e. $\hat{X}_i$ is the $T_i \in T$ for which $E((X_i - T_i)^2)$ is a minimum.

First I establish some more notation. Define

$$\hat{X}_m = \frac{1}{m} \sum_{i=1}^m X_i \quad \text{and} \quad X_m^* = \sum_{i=1}^M \frac{\mu_i - \bar{\mu}}{\sigma^2} X_i,$$

where

$$\mu_i = E(X_i) \quad \text{and} \quad \bar{\mu} = \frac{1}{m} \sum_{i=1}^m \mu_i.$$

© Australian Statistical Publishing Association Inc. 2001
Appendix B shows that
\[ \hat{X}_j = \begin{cases} X_j & \text{for } j \leq m \\ \bar{X}_m + c_j X_m^* & \text{for } j > m, \end{cases} \]
where
\[ c_j = \frac{\mu_j - \bar{\mu}}{1 + E(X_m^*)}. \]
Note that \( 1 + E(X_m^*) \) is non-zero, since
\[ E(X_m^*) = \sum_{i=1}^{m} \frac{\mu_i - \bar{\mu}}{\sigma^2} \]
Substituting the expressions for \( \hat{X}_j \) back into (1) and collecting the coefficients of \( X_j \), we obtain
\[ u_i = w_i + \frac{1}{m} \sum_{j=m+1}^{n} w_j + \frac{c_j}{\sigma^2} \sum_{j=m+1}^{n} (\mu_j - \bar{\mu}) w_j. \]

For the Musgrave case,
\[ X_i = \alpha + \beta i + \epsilon_i, \]
we obtain
\[ \mu_j - \bar{\mu} = \beta \left( j - \frac{m+1}{2} \right) \quad \text{and} \quad E(X_m^*) = \frac{\beta^2}{\sigma^2} \frac{m(m+1)(m-1)}{12}. \]
Substituting these expressions in (3) gives the result (1).

5. Values adopted for the slope ratios

The formula for the surrogate filters in Section 3 contains the parameter \( \beta^2 / \sigma^2 \). In X-11 the choice of whether to adopt a 9-term, 13-term or 23-term Henderson moving average for trend estimation depends on \( \bar{I}/\bar{C} \). Here, for an additive adjustment, \( \bar{I} \) is the average of absolute month to month change in the estimated irregular, and \( \bar{C} \) is the average of the absolute month to month changes in an estimate of the trend.

For a multiplicative adjustment, the \( \bar{I}/\bar{C} \) ratio is also used. However, the numerator is the average of the absolute monthly percentage changes in an estimated irregular; the denominator is the average of the absolute monthly percentage changes in an estimated trend.

The use of this criterion goes back to earlier work, e.g. Bongard (1960) and Marris (1960). More recently, interesting work in the seasonal factor case has been done by Lothian (1984). Shiskin et al. (1967 p. 4) suggest use of a
- 9-term moving average if \( \bar{I}/\bar{C} \leq 0.99 \),
- 13-term moving average if \( 1 \leq \bar{I}/\bar{C} \leq 3.49 \),
- 23-term moving average if \( 3.5 \leq \bar{I}/\bar{C} \).

Suppose that the trend is the one we used earlier in Musgrave’s model, \( C_i = \alpha + \beta i \), and that the \( \epsilon_i \) are independent \( \text{N}(0, \sigma^2) \). For this series,
\[ \bar{I} = E(\mid \epsilon_{i+1} - \epsilon_i \mid). \]
As \( \epsilon_{i+1} - \epsilon_i \overset{d}{=} \text{N}(0, 2\sigma^2) \), it follows that \( \bar{I} = 2\sigma / \sqrt{\pi} \). And \( \bar{C} = \beta \), so
\[ \frac{\beta^2}{\sigma^2} = \frac{4}{\pi} \frac{1}{(\bar{I}/\bar{C})^2}. \]
This is apparently the formula used for the surrogate filters, with
\[
\bar{I}/\bar{C} = \begin{cases} 
1 & \text{for 9-term}, \\
3.5 & \text{for 13-term}, \\
4.5 & \text{for 23-term}.
\end{cases}
\]
Note that the first two values are the ends of the corresponding range quoted above (Shiskin et al., 1967).

In X-11-ARIMA/88 (Dagum, 1988) the tabulated coefficients for the 23-term Henderson surrogate uses \(\bar{I}/\bar{C} = 4.5\). The 7-term, 15-term and 17-term Henderson surrogates in the work of the Australian Bureau of Statistics (Castles, 1987) have \(\bar{I}/\bar{C} = 4.5\).

The above derivation of the \(\bar{I}/\bar{C}\) ratio is for the additive case. A referee asked how this is related to the multiplicative case, as most adjustments are multiplicative. Restating the question, why should the same values of the ratio be carried over for a multiplicative adjustment? Early work on seasonal adjustment often converted from the multiplicative case to the additive case by taking logs. I think a pioneer would have answered that the \(\bar{I}/\bar{C}\) ratio is defined in the multiplicative case to correspond approximately to the corresponding quantity in the associated additive case, so the use of the same value for both is appropriate.

A (conjectural) reconstruction of the reasoning is as follows. Let \(i_n\) be the irregular in the multiplicative case. Taking logs to convert to the associated additive case, a typical term in the numerator for the ‘corresponding’ additive \(\bar{I}/\bar{C}\) ratio is then
\[
|\log(i_{n+1}) - \log(i_n)| = \left| \log \left(1 + \frac{i_{n+1} - i_n}{i_n} \right) \right| \approx \left| \frac{i_{n+1} - i_n}{i_n} \right|.
\]
The final approximation uses \(\log(1+x) \approx x\) for \(x\) small, as we hope the percentage changes are. The term on the right-hand side of (4) is the corresponding term in the numerator of the ‘multiplicative’ form of the \(\bar{I}/\bar{C}\) ratio. A similar argument holds for the terms in the denominator.

6. Relationship to the work of Kenny and Durbin

Kenny & Durbin (1982) observed that the surrogate Henderson weights were approximately what you would get if the series to be smoothed were extended using a line fitted by least squares. To quote Kenny & Durbin (1982 p. 39):

We have found it possible to reproduce the X-11 Henderson weights almost exactly by using linear extrapolation, provided that in the most one-sided cases the extrapolated values are read off one or two steps before the point to which they relate.

This approximation is an interesting interpretation of the formula, so I show how it follows.

If values \(Y_1, \ldots, Y_m\) are observed at times \(t = 1, \ldots, m\), then it can be shown using the normal equations that the fitted least squares line corresponding to these values is
\[
\frac{1}{m} \sum_{i=1}^{m} Y_i + \left( t - \frac{m + 1}{2} \right) \frac{12}{(m-1)m(m+1)} \sum_{i=1}^{m} \left( i - \frac{m + 1}{2} \right) Y_i.
\]
If the extrapolated values corresponding to time \(i\) are read off from the line at time \(i - \delta_i\) then the surrogate weights are
\[
u^*_r = w_r + \frac{1}{m} \sum_{i=m+1}^{n} w_i + \frac{12 \left( r - \frac{m + 1}{2} \right)}{(m-1)m(m+1)} \sum_{i=m+1}^{n} \left( i - \delta_i - \frac{m + 1}{2} \right) w_i.
\]
In the Musgrave case this is the same as the surrogate filter formula we had earlier, provided
\[
\sum_{i=m+1}^{n} w_i \delta_i = \frac{\sum_{i=m+1}^{n} \left( i - \frac{m+1}{2} \right) w_i}{1 + \frac{m(m-1)(m+1)}{12} \beta^2 \sigma^2}.
\]
In particular, note the solution
\[
\delta_i = \frac{i - \frac{m+1}{2}}{1 + \frac{m(m-1)(m+1)}{12} \beta^2 \sigma^2} \quad \text{for } i = m+1, \ldots, n.
\]
This solution, which comes from equating the individual predictors, not just the resulting trend, is independent of the initial $w_i$. Tabulating it confirms the observations of Kenny & Durbin (1982).

7. Concluding remarks

The Kenny and Durbin interpretation has the following variant. Let $\gamma$ be the coefficient of variation of the slope of a line fitted by least squares to $X_1, \ldots, X_m$ and assuming the Musgrave model $X_i = \alpha + \beta i + \epsilon_i$. Then,
\[
\gamma^2 = \frac{12}{m(m-1)(m+1)} \frac{\sigma^2}{\beta^2}.
\]
Let the estimated coefficients for the least squares line be $\hat{\alpha}$ and $\hat{\beta}$, so that the predicted value from the least squares line for the (out-of-range) observation $i$ is $\text{LSQ}_i = \hat{\alpha} + \hat{\beta} i$. Then the (implicit) Musgrave predictor for out of range $X_i$ is
\[
\frac{\gamma^2}{1 + \gamma^2} \bar{X}_m + \frac{1}{1 + \gamma^2} \text{LSQ}_i.
\]
This goes some way towards explaining the success of the Musgrave surrogates. The extrapolated values are a compromise between the mean of the most recent values and a least squares line fitted to the most recent values.

Appendix A: Affine prediction

This paper has quoted facts about prediction in an affine subspace $\mathcal{T}$. Since this may not be as well known as the case where $\mathcal{T}$ is a genuine subspace, I repeat the necessary facts here.

Suppose $\mathcal{T}$ is a (closed) non-empty affine subspace of some space of random variables with finite first and second moments. Given a random variable $Z$, with finite first and second moments, there is a unique $\hat{Z} \in \mathcal{T}$ minimizing $\text{E}((\hat{Z} - Z)^2)$ and this $\hat{Z}$ is uniquely characterized (in the $L^2$ sense) by the properties:
\[
\hat{Z} \in \mathcal{T} \quad \text{and} \quad \text{E}((\hat{Z} - Z)T_1) = \text{E}((\hat{Z} - Z)T_2) \quad \text{for all} \quad T_1, T_2 \in \mathcal{T},
\]
where $\hat{Z}$ is the optimal predictor of $Z$ in $\mathcal{T}$. This can be deduced from the better known subspace case. Alternatively, see Drygas (1970 Section 2.31).

Further, given random variables $Z_1, \ldots, Z_n$ with finite first and second moments, and constants $w_1, \ldots, w_n$ such that $\sum_{i=1}^{n} w_i = 1$, the optimal predictor of $\sum_{i=1}^{n} w_i Z_i$ in $\mathcal{T}$ is $\sum_{i=1}^{n} w_i \hat{Z}_i$, the weighted sum of the optimal predictors of the individual variables. This is a consequence of the characterization above.

© Australian Statistical Publishing Association Inc. 2001
Previous authors (Pierce, 1980; Wallis, 1982a; Cleveland, 1983) have proved results showing that the mean square expected revision can be minimized by using prediction methods. Our emphasis in Section 4 differs from theirs in two respects.

1. The set $T$ of potential predictors in our case is only an affine subspace, not in general a subspace, i.e. in general $T$ does not contain 0. We are dealing with a restricted set of predictors, because we have imposed the condition $\sum_{i=1}^{m} d_i = 1$ on the members $\sum_{i=1}^{m} d_i X_i$ of $T$. This could be regarded as a robustness requirement — however well or badly the model for the series fits, we want the surrogate filters to pass a constant unaltered.

2. Only $X_1, \ldots, X_m$ are being used for prediction, not the semi-finite past or all the available data points. This is an advantage in the application, as we need only to trust the model for the series locally.

Appendix B: Derivation of the formulae

The assumptions and the notations of Section 4 remain in force.

For $i \leq m$, $X_i$ belongs to $S$, so the fact that $\hat{X}_j = X_j$ for $i \leq m$ is immediate.

To deal with the remaining cases, consider any random variable $Z$ with finite first and second moments, with $Z$ independent of $X_1, \ldots, X_m$. We obtain the optimal predictor $\hat{Z} \in T$ for $Z$ so that the result then applies immediately to $X_{m+1}, \ldots, X_n$.

Write $\hat{Z} = \sum_{j=1}^{m} d_j X_j$ with $\sum_{j=1}^{m} d_j = 1$. Define $G = \sum_{j=1}^{m} d_j X_j - Z$. For any $i = 1, \ldots, m$, since $\hat{X}_m$ and $X_i$ belong to $T$, we can apply the criterion in Appendix A with $T_1 = X_i$ and $T_2 = \hat{X}_m$. This gives

$$E(G X_i) = E(G \hat{X}_m).$$

Using the formula $E(Y_1 Y_2) = E(Y_1) E(Y_2) + \text{cov}(Y_1, Y_2)$ we obtain

$$E(G)\mu_i + \text{cov}(G, X_i) = E(G)\tilde{\mu} + \text{cov}(G, \hat{X}_m).$$

Using the independence of $Z, X_1, \ldots, X_m$, the definition of $G$ and the fact that $\sum d_j = 1$, this gives

$$E(G)\mu_i + d_i \sigma^2 = E(G)\tilde{\mu} + \frac{\sigma^2}{m}.$$ 

i.e.

$$d_i = \frac{1}{m} + \frac{E(G)}{\sigma^2} (\tilde{\mu} - \mu_i).$$

Multiplying by $X_i$ and summing from 1 to $m$, we get

$$\sum_{i=1}^{m} d_i X_i = \hat{X}_m - E(G)X_m^*. \quad (5)$$

Taking expectations

$$\sum_{i=1}^{m} d_i \mu_i = \tilde{\mu} - E(G)E(X_m^*); \quad (6)$$

but also, taking expectations in the definition of $G$,

$$E(G) = \sum_{i=1}^{m} d_i \mu_i - E(Z). \quad (7)$$
From (6) and (7) we obtain

\[ E(G) = \frac{\mu - E(Z)}{1 + E(X^*_m)} . \]

Substituting this back into (5) gives

\[ \hat{Z} = \bar{X}_m + \frac{E(Z) - \mu}{1 + E(X^*_m)} X^*_m . \]

Applying this with \( Z = X_{m+1}, \ldots, X_n \) produces the results we want.

References

Bongard, J. (1960). Notes on the characteristics of seasonal adjustment in Europe. In Seasonal Adjustment on Electronic Computers, pp. 179–204. OECD.


© Australian Statistical Publishing Association Inc. 2001