Appl. Statist. (1978),
27, *No.* 3, p. 264–279

The Holt–Winters Forecasting Procedure

By C. Chatfield

University of Bath, Britain

[Received July 1977. Final revision May 1978]

Summary

The Holt–Winters forecasting procedure is a simple widely used projection method which can cope with trend and seasonal variation. However, empirical studies have tended to show that the method is not as accurate on average as the more complicated Box–Jenkins procedure. This paper points out that these empirical studies have used the automatic version of the method, whereas a non-automatic version is also possible in which subjective judgement is employed, for example, to choose the correct model for seasonality. The paper re-analyses seven series from the Newbold–Granger study for which Box–Jenkins forecasts were reported to be much superior to the (automatic) Holt–Winters forecasts. The series do not appear to have any common properties, but it is shown that the automatic Holt–Winters forecasts can often be improved by subjective modifications. It is argued that a fairer comparison would be that between Box–Jenkins and a non-automatic version of Holt–Winters. Some general recommendations are made concerning the choice of a univariate forecasting procedure. The paper also makes suggestions regarding the implementation of the Holt–Winters procedure, including a choice of starting values.

Keywords: HOLT–WINTERS METHOD; BOX–JENKINS METHOD; FORECASTING; EXPONENTIAL SMOOTHING

1. Introduction

An important class of forecasting procedures is that of *univariate* or projection methods, where forecasts of a given variable are based only on the current and past values of this variable. One simple widely used method of this type is the Holt–Winters procedure (e.g. Winters, 1960; Chatfield, 1975, p. 87; Montgomery and Johnson, 1976, Chapter 5; Granger and Newbold, 1977, p. 164). This generalizes simple exponential smoothing so as to cope with trend and seasonal variation.

A technical description of the method is given in the appendix. There are two types of seasonal model: an *additive* version which assumes that the seasonal effects are of constant size and a *multiplicative* version which assumes that the seasonal effects are proportional in size to the local deseasonalized mean level. Both seasonal models assume that the local deseasonalized mean level may be modified by an additive trend term and also that there is an additive error term of constant variance.

The user must decide whether to use the additive or multiplicative seasonal model or a non-seasonal model. He must then select starting values for the seasonal factors (if a seasonal model is used) and also for the local mean and trend. As each new observation becomes available, the local mean, the seasonal factors and the trend are all updated by exponential smoothing using three smoothing constants which we will denote by α, β, γ respectively. Forecasts can then be produced for any number of steps ahead.

Because it is so straightforward, the Holt–Winters method is particularly suitable for production planning and stock control when forecasts are required for a large number of variables. Then a fully *automatic* version of the method is usually used so that a computer-based system can be set up to make routine forecasts without human intervention. In this case, the multiplicative seasonal model is usually used for *every* series. Alternatively, a non-automatic version of the method is possible where, for example, the user chooses the most appropriate model, takes any outliers or discontinuities into consideration and keeps a careful check on the forecast errors.

This paper presents a new look at the Holt–Winters procedure, describes the analysis of seven economic series, makes a number of practical suggestions regarding the implementation of the procedure (both in its automatic and non-automatic form) and finally makes recommendations regarding the choice of a univariate forecasting procedure, particularly in regard to deciding between the non-automatic Holt–Winters procedure and the Box–Jenkins approach.

2. A Comparison with Other Procedures

The choice of an appropriate forecasting procedure depends on a variety of considerations such as the objective of the forecasting exercise, the number of observations available and the number of variables to be forecast. Multivariate methods are advocated in econometrics and there have been some inconclusive attempts to compare econometric models with univariate methods (e.g. Cooper and Nelson, 1975). The bivariate version of the method proposed by Box and Jenkins (1970) is also of current interest, but there are as yet few case studies available. So in this paper we restrict attention to univariate methods which have been of prime interest to statisticians.

With so many methods available, it is difficult for the user to choose the one most suitable for his situation. Some guidance on the relative accuracy of the different procedures is provided by the large-scale empirical comparisons of Newbold and Granger (1974) and Reid (1975) using, in each case, over 100 series of macroeconomic data. Their results suggest that the Holt–Winters (abbreviated HW) procedure has comparable accuracy to the best of the other simple automatic procedures such as Brown's method and Harrison's sea-trend system. However, the univariate Box–Jenkins method gave more accurate forecasts than the HW procedure for about two-thirds of the series analysed. The Box–Jenkins (abbreviated BJ) procedure requires the user to identify an appropriate model in a general class of stochastic processes called autoregressive-integrated-moving-average models (or ARIMA models). Another empirical comparison by Groff (1973) came to different conclusions regarding the accuracy of the BJ procedure, but these results appear to be invalidated by the fact that the BJ identification procedure was not used for each individual series.

Now accuracy, as measured by mean-square forecast error, is only one consideration when selecting a forecasting procedure. Simple projection methods are often preferred to the BJ procedure for a variety of practical considerations such as cost, available expertise and the fact that the BJ procedure requires at least 50 observations to have a good chance of success.

Of course, it is sometimes argued (e.g. Jenkins, 1974) that the HW method is optimal for a special case of the ARIMA class of models, the implication being that one might as well use the more general BJ procedure. But in fact the multiplicative HW model does *not* have an ARIMA equivalent, while the ARIMA model corresponding to the additive HW model is so complicated that it would never be identified in practice (see Chatfield, 1977). Thus, for all practical purposes, the HW method is not a special case of the BJ procedure. Indeed, what is remarkable about the Newbold–Granger and Reid comparisons is that HW actually performs better than BJ for about one-third of the series. Possible reasons for this phenomenon are discussed by Chatfield (1977) and two particular analyses where it happens are those described by Chatfield and Prothero (1973) and Montgomery and Contreras (1977).

Finally, in defence of the HW procedure against the empirical results of the Newbold–Granger and Reid studies, it must be said that these comparisons were not altogether fair in one important respect. While the results give important guidance on the relative accuracy of completely automatic methods, they say little about the comparison with the BJ procedure because, if one is thinking of using a complicated method like BJ, it would seem fairer to compare it with *non-automatic* versions of the simpler procedures. It would after all be a miracle if the completely automatic HW procedure outperformed the BJ procedure on average, though as we have seen it does actually do this for about one-third of the series. This highlights a general problem with empirical comparisons, as noted by Box (1970) in a different context, that they are often made in circumstances which greatly favour one contender and the result is as expected. For example, when a research worker proposes a new forecasting method he naturally understands it better than other methods and may well get better results with it than other methods. With regard to the HW method, empirical comparisons have hitherto ignored the possibility of using subjective judgement to improve forecasts and so, in this paper, I attempt to compare the automatic and non-automatic HW procedures.

One of the difficulties of attempting empirical comparisons of forecasting procedures using real data is that there is no such thing as a "random sample of time series". For my study I used a specially selected non-random sample for the following reason. One interesting feature of the results reported by Newbold and Granger (1974) was that BJ sometimes gave *much* better forecasts than HW, whereas HW was rarely superior to BJ by more than a small margin. This was perhaps to be expected given that the BJ procedure allows a choice from a much wider class of models. By using series for which BJ gave much better forecasts than automatic HW, I would endeavour to answer two queries. Firstly, to see if series of this type have any general features which would yield guidelines for indicating when BJ is likely to be preferable to automatic HW and, secondly, to see if the HW results can be improved by using a non-automatic version of the method. I contacted Dr P. Newbold and he kindly sent me seven series from those examined in the Newbold–Granger comparison for which BJ gave much better forecasts than HW. These series are listed in Table 1.

TABLE 1

The series analysed

		Number of monthly observations in	
Code	*Description*	*Fitting period*	*Forecasting period*
A	Yield on short term Government securities	204	48
B	Number of TV licences	168	24
C	National Coal Board Juvenile recruitment	72	27
D	Employment in Manufacturing (Canada)	108	30
E	Industrial Production: Manufacturing (Canada)	108	30
F	Unemployment (Belgium)	108	30
G	Hourly earnings in manufacturing (Canada)	108	30

Newbold and Granger (1974) divided each series into two parts; a fitting period, which was used to fit an appropriate model, and a forecast period when forecasts were compared with actual observations. The lengths of each period are shown in Table 1.

Unfortunately, records no longer exist of the Box–Jenkins models fitted in the Newbold–Granger comparison or of the mean-square forecast errors (over the forecast period) which resulted from using the BJ or automatic HW methods. However, records of the ratio (BJ m.s.e./HW m.s.e.) do exist. For series *B* to *G* which all show seasonal variation, this ratio lies between 0·4 and 0·7. Series *C* gives the lowest ratio. There were few non-seasonal series in the Newbold–Granger set, the "worst" cases having a higher ratio between 0·8 and 0·9. Series *A* was selected at random from those cases.

The absence of detailed records on the Newbold–Granger comparison is annoying and raises a general query about empirical studies which are not available for re-examination, but this should not prevent us achieving our objectives. Admittedly, any improvements that we make on the automatic HW procedures will be improvements on our findings only, but there is good reason to suppose that similar improvements will be possible for the Newbold–Granger automatic results.

Throughout the paper, I not only estimate parameters by least squares but also compare accuracy by means of mean-square forecast errors, thus effectively assuming a quadratic cost-of-error function. This assumption is often queried and it has to be admitted that it is usually made for practical convenience and because there is often no clearly superior alternative (e.g. see Granger and Newbold, 1973). However, Granger's (1969) results do give some comfort that the choice of a quadratic cost function is not too critical, particularly if the error distribution and the cost function are approximately symmetric. To check on this, I re-estimated the HW smoothing parameters, which are evaluated in Section 4, by minimizing mean *absolute* one-step-ahead forecast errors. This made virtually no difference for all seven series.

3. Graphical Analysis of the Data

I began by plotting the data in Figs 1–7 to show up the main qualitative features of the series. Points to look for include trend, seasonality, outliers and any evidence of a change in structure. We note the following features.

Seasonal pattern

Series A is non-seasonal, while Series B loses its small seasonal component after the first four or five years. The rest of the series are seasonal, but in markedly different ways. For example, the seasonal variation is relatively small for Series G, but large for F. The size of the variation increases with the mean for Series E, but decreases for G as the mean increases. For Series C, the seasonal variation appears to decrease sharply in size after the first two or three years.

Trend

There is increasing trend for Series B, D, E and G, while the mean level for Series A increases and decreases sharply many times in a manner reminiscent of a random walk. Series C shows little trend while Series F shows two turning points, first decreasing, then increasing, and then decreasing again.

Random component

Some of the series show a high random component (e.g. Series A and C) while others show a low random component (e.g. Series B).

To summarize the previous remarks, it appears that the seven series exhibit very different properties. Indeed, some of the series exhibit non-stationary changes in trend or seasonality which may be thought unusual or exceptional. But, as Harrison and Stevens (1975) have pointed out, such data series are in no way exceptional or pathological.

Although I have analysed a relatively small sample of time series, I am forced to surmise that time series, for which BJ performs much better than (automatic) HW, do *not* appear to have any general common features.

The reader may now ask if the fitted BJ or HW models have any features in common. In Section 4 we will see that the HW smoothing parameters vary considerably for the seven series. As for the BJ models, the ones fitted by Newbold and Granger are no longer available. This author could refit BJ models (as he has for Series A—see Section 6), but the models might well be different to those fitted by Newbold and Granger. Given the different properties of the series, it seems highly unlikely that the BJ models do have any features in common and, even if they do, it would be unlikely to help in deciding *beforehand* if a BJ analysis is worth doing.

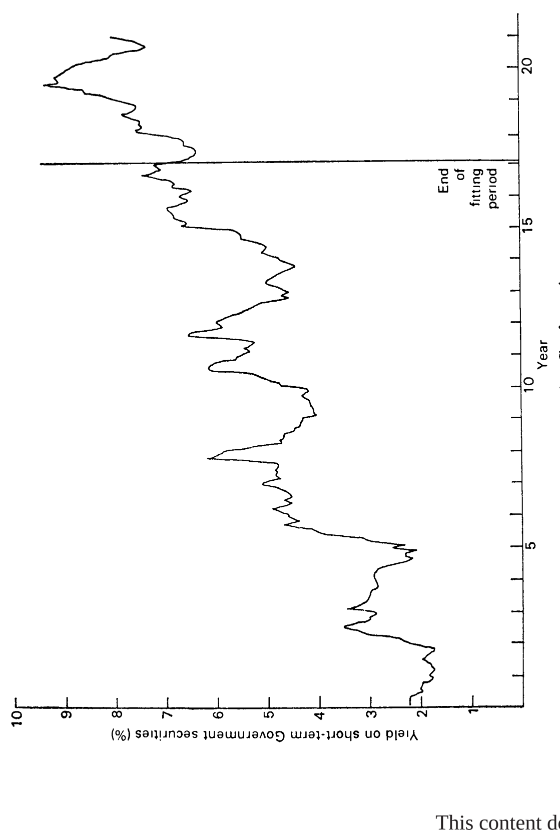

FIG. 1. Series *A*.

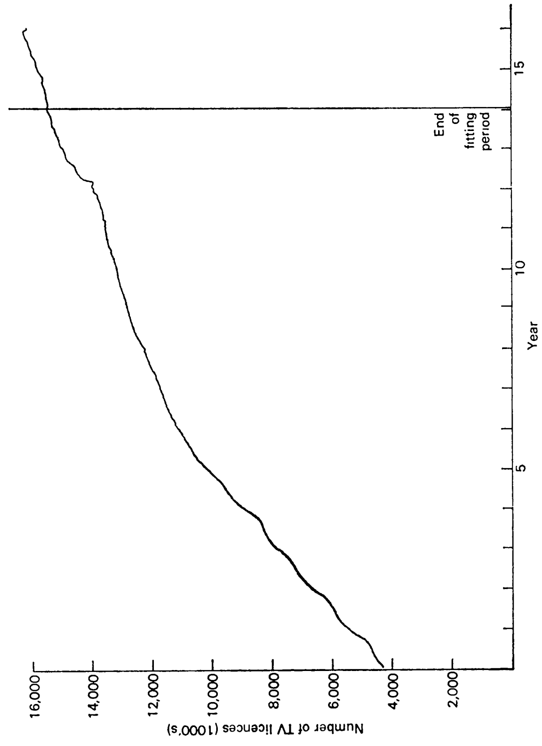

FIG. 2. Series *B*.

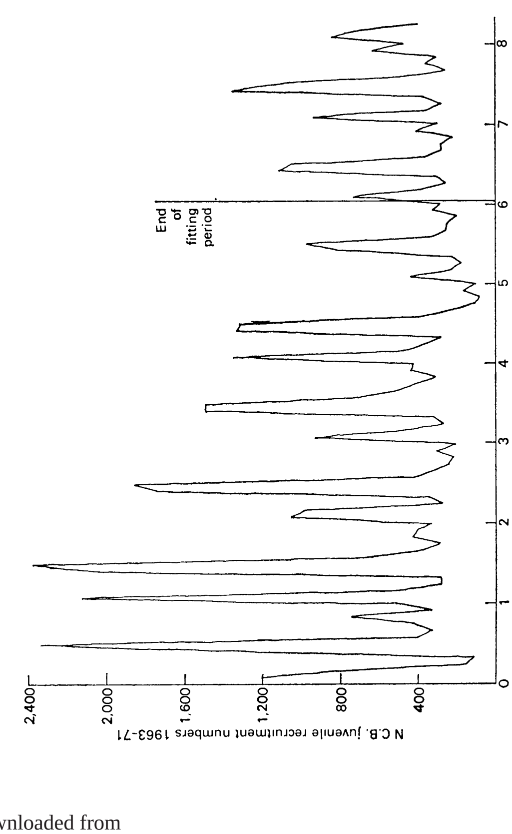

FIG. 3. Series *C*.

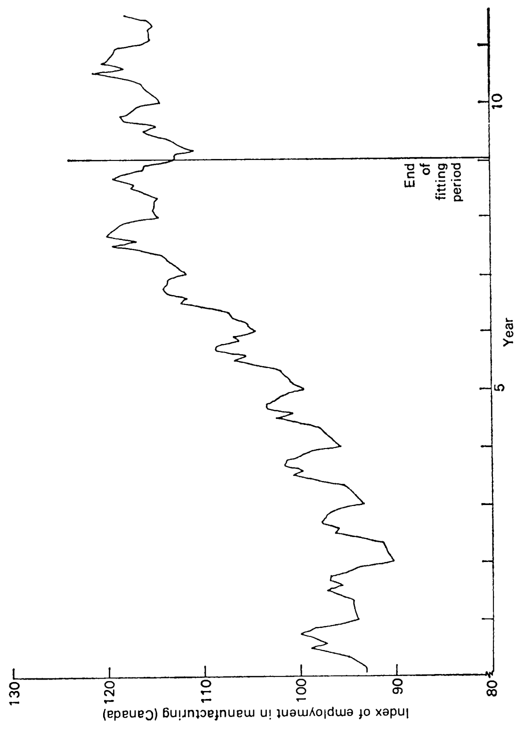

FIG. 4. Series *D*.

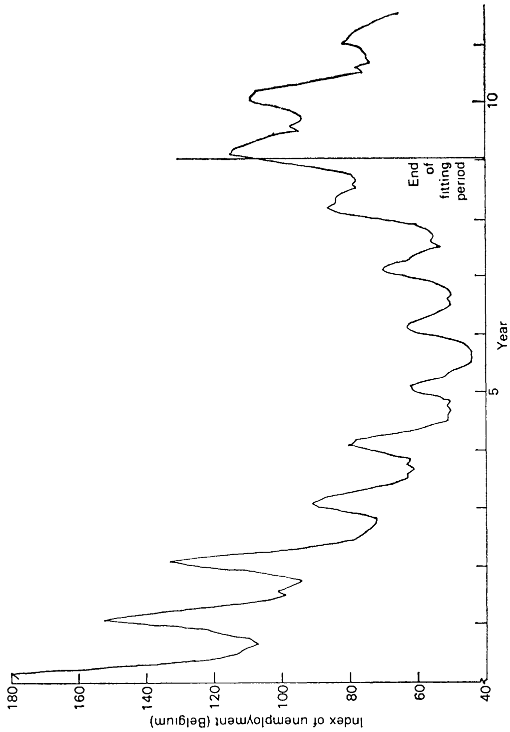

FIG. 6. Series *F*.

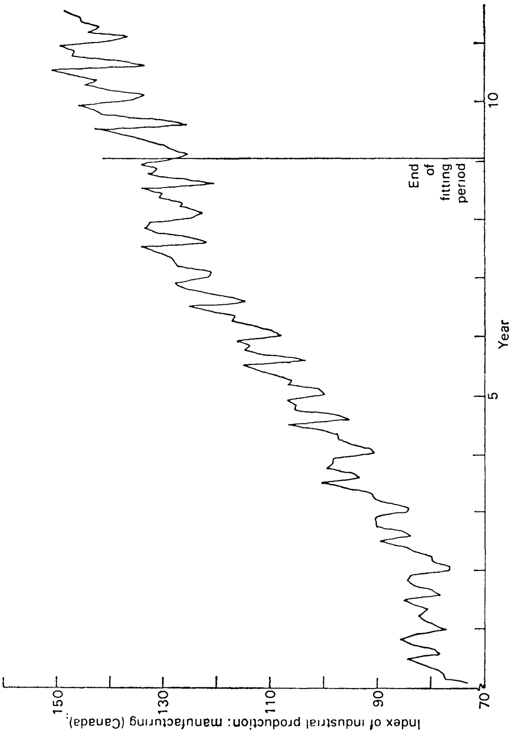

FIG. 5. Series *E*.

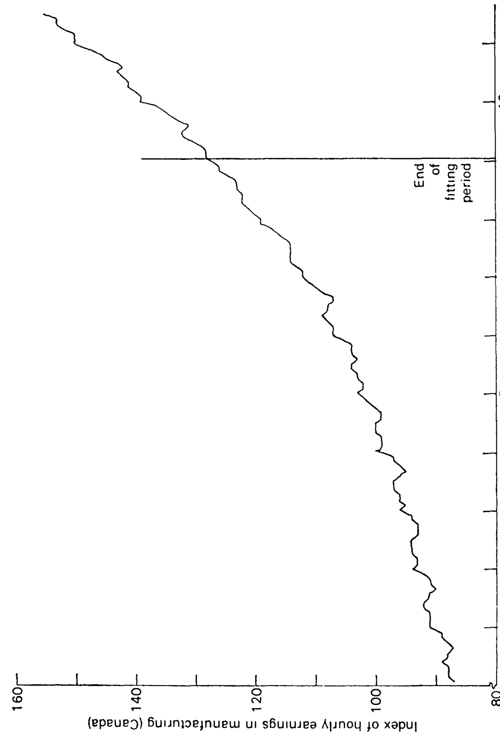

FIG. 7. Series *G*.

4. The Automatic Holt–Winters Procedure

The second main objective of this study was to find general ways of improving the automatic HW procedure. As the Newbold–Granger results were no longer available, the next step was to carry out the automatic HW procedure on each series to produce results for comparison purposes. There is no guarantee that my results will be exactly the same as those produced by Newbold and Granger but they should be adequate for the purpose of examining the non-automatic HW procedure.

Two computer programs were written. The first estimates the best values of the three smoothing constants by minimizing the sum of squared one-step-ahead forecast errors over the period of fit excluding the first two or three years' data which were used as a run-in period. The second program computes one-step-ahead forecasts over the forecast period using the optimal smoothing constants. The programs allow seasonal and non-seasonal data to be analysed and also allow the seasonal variation to be additive or multiplicative. The procedure is as described by Winters (1960) except that starting values for trend and seasonal terms were estimated from the first two years' data rather than from the first and last years of the fitting period.

For the seasonal Series (B to G), Newbold and Granger used a multiplicative seasonal model in every case (as also did Reid, 1975), while for Series A a non-seasonal model was used. I did the same. Using the same fitting periods as Newbold and Granger, I found the estimated values of the three smoothing constants to be as in Table 2. These estimates deserve comment

Table 2

Estimated smoothing constants

Series	α (*mean*)	β (*seasonal*)	γ (*trend*)
A	0·8	—	0·20
B	0·3	1·0	0·15
C	0·4	1·0	0·03
D	0·6	1·0	0·10
E	0·6	0·8	0·05
F	0·6	1·0	0·14
G	0·2	0·4	0·20

as some of the values are much higher than those customarily quoted in the literature. For example, Coutie *et al.* (1964) say that typical values for sales data are $0{\cdot}1<\alpha<0{\cdot}3$ and $0<\gamma<0{\cdot}03$, while Reid (1975) says typical values might be $\alpha = 0{\cdot}2$, $\beta = 0{\cdot}3$ and $\gamma = 0{\cdot}1$, in my notation. Now changes of 0·1 or 0·2 in the values of α and β will not make much difference, but the above "typical" values are considerably different from some of my estimates and would do rather badly for some of the series.

On the basis of my experience with these and other series, I suggest that it is rather dangerous to try to give "typical" values for the smoothing constants. Thus, the HW smoothing parameters should not be guessed, but rather estimated from the data.

Using the optimal smoothing constants, I next computed mean-square one-step-ahead forecast errors over the respective forecasting periods. From these I calculated the respective coefficients of determination, namely the proportions of the total corrected sums of squares which were "explained" by the one-step-ahead forecasts. Although HW was expected to have inferior accuracy to BJ for these series, the coefficients of determination were all found to be reasonably high, ranging from 77% for Series C to over 99% for Series G. In the latter case, the HW method would almost certainly be judged sufficiently accurate even though the BJ forecasts may be even better.

5. The Non-automatic Holt–Winters Procedure

This section considers a variety of ways in which the automatic HW procedure can be modified so as to produce improved forecasts. I do not claim that these modifications are in any sense "new", as they will already be used by good statisticians where appropriate. But one important aim of this paper is to stress that they were not used in the empirical comparisons of Newbold–Granger and Reid as these authors used the automatic version of the method.

5.1. *Correct Choice of Model*

Series B is seasonal for the first four or five years, but non-seasonal thereafter. As Newbold and Granger used a seasonal model, I naturally did the same in the "automatic" run, but it is clear that a non-seasonal model may be appropriate for making forecasts in the forecast period. This is done by setting the multiplicative seasonal factors equal to one, and setting β to be zero. The two other smoothing constants were then re-estimated to be $\alpha = 0{\cdot}8$ and $\gamma = 0{\cdot}2$. Note particularly that the value of α is completely different to the value in Table 2. The mean-square forecast error in the forecast period was then calculated and found to be *less than half* the value for the seasonal model, thereby increasing the coefficient of determination in the forecast period from 0·915 to 0·964. This astonishing improvement may be compared with the Newbold–Granger finding that the BJ mean-square forecast error was between 0·6 and 0·7 times their automatic HW mean-square forecast error.

At first sight the improvement noted above may be somewhat surprising, as many statisticians think the non-seasonal model is a special case of the seasonal model obtained by setting $\beta = 0$. But this ignores the effect of the initial seasonal factors. If a seasonal model is applied to non-seasonal data, the initial seasonal factors will reflect random variation and they will always be present if $\beta = 0$. If a positive value of β is taken, the effect of the initial values will gradually die out only to be replaced by further random variation. If, for example, we try the seasonal model on Series A, which is clearly non-seasonal, the mean-square forecast error more than doubles even with an optimum choice for β.

Now, in most model-fitting problems, the addition of unnecessary parameters leads to a small improvement in mean-square error, even though it may not be significantly large. But here we have a situation where the addition of unnecessary parameters makes the mean-square error much worse, because the starting values for the non-existent seasonal factors are not set equal to their "best" values, namely unity. Thus, it is clearly important to use a non-seasonal model for non-seasonal data.

But what happens when the seasonal variation is small compared with the random variation? Is it better to use a seasonal or non-seasonal model? For Series C to F, which have large seasonal components, we found as expected that the mean-square error for the seasonal model is much smaller than for the non-seasonal model. But for Series G, where the seasonal variation is small compared with other sources of variation, we found that a slightly smaller mean-square error could be obtained by setting the initial seasonal factors to be unity, but allowing β to be non-zero. This modification may be justified as follows. If the random variation is bigger than the seasonal variation, then unity may be closer on average to the true seasonal values than the values calculated from the first year or two's data only, but subsequent smoothing may give better seasonal estimates.

Apart from the choice between seasonal and non-seasonal models, the choice between multiplicative and additive seasonality may also be important. This choice is usually made by subjectively assessing the graph of the data and is difficult to make part of an automatic procedure.

In our group of series, the seasonal components of Series C, D, E and F do appear to be approximately multiplicative, but that of Series G does not since it actually decreases in size as the mean level increases. This means that the seasonal component of Series G is neither additive nor multiplicative, but it seems worth seeing if the additive model is a better approximation than the multiplicative model. In the event, the additive model gave virtually no improvement of fit, presumably because the size of the seasonal variation is small compared with the trend. No doubt there will be other series where the choice of the correct seasonal model will be crucial. Indeed, the author was once asked by a commercial organization to find out why their HW forecasts were "blowing up". It transpired that the data had not been plotted, that a multiplicative model was being used when the data were clearly additive, and that their computer program "went haywire" as a result because the seasonal factors were not normalized after each year's cycle.

5.2. *Adjustment of the Data*

Another general point to consider is whether any observations need to be adjusted or deleted. Freak values or outliers may arise from some known phenomenon as, for example, when sales are affected by a strike. Such values need to be "adjusted" before the analysis commences, in whatever way seems appropriate. A second point to consider is whether it is reasonable to regard the data as having been generated by a single model. Sometimes the properties of a time series may change substantially during the period under study so that the first part of the data may have little or no relevance for the purpose of constructing a model and making forecasts from the end of the series. In this sort of situation it may be better to discard the first part of the data before attempting to fit a model.

From a look at our seven series, there do not appear to be any freak values. The only adjustments which appear to be worth considering are for Series *C* and *F*, where the seasonal variation in the first two years or so is larger than in later years, and for Series *B*, where the seasonal variation disappears after about four to five years. These three series were reanalysed omitting the first two years' data in the case of Series *C* and *F*, and omitting the first six years' data in the case of Series *B*. However, these modifications did not give any improvement on average to the HW forecasts in the forecasting period. It appears that the effects of the early observations have by then more or less died out. Thus the HW forecasts for our seven series cannot be improved by modifying the data in a common-sense way, though no doubt subjective data-adjustment may be useful for some other series.

5.3. *The Treatment of Autocorrelated Errors*

The final modification we consider here resulted from noting that the one-step-ahead forecast errors produced by the HW procedure tended to go in "runs" having the same sign. In other words, the first-order autocorrelation coefficient, r_1, of the forecast errors was sometimes "large" and positive indicating that the forecasts were not optimal. The values of r_1 in both the fitting and forecasting period were calculated for all seven series and are given in Table 3.

TABLE 3

First-order autocorrelation coefficients of one-step-ahead HW forecast errors

Series	r_1 *in fitting period*	r_1 *in forecasting period*
A	0·38	0·49
B	0·24	−0·02
C	0·10	−0·19
D	0·33	0·36
E	0·01	0·13
F	0·42	0·70
G	0·24	0·34

One general procedure for improving HW forecasts suggested by Newbold and Granger (1974) is to take a linear combination of the forecasts from the HW procedure and from a technique called stepwise autoregression (Granger and Newbold, 1977, Section 5.4). This combination of two sets of "automatic" forecasts gives results comparable in accuracy to those of Box–Jenkins. Intuitively, the stepwise autoregression forecasts may be able to take account of autocorrelation in the time series which cannot be described by a HW model. Here we try an even simpler method of taking autocorrelation into account which has been suggested by D. J. Reid for improving forecasts from general exponential smoothing (see Granger and Newbold, 1977, p. 169). It is called a two-stage forecasting procedure by Gilchrist (1976, p. 262).

The modification consists of fitting a first-order autoregressive model to the HW one-step-ahead forecast errors, which we denote by $\{e_t\}$. The quantity $\{\lambda e_t\}$ is added to the one-step-ahead forecast made at time t, where λ is a new parameter which we will call the autoregressive parameter. An intuitively sensible estimate of λ is the value of r_1 obtained for the forecast errors in the fitting period using the unmodified HW procedure.

Now, when testing the null hypothesis that a series of N observations come from a purely random process, it can be shown (see Chatfield, 1975, p. 62) that values of $|r_1|$ exceeding $2/\sqrt{N}$ are significantly different from zero at the 5 per cent level. Box and Pierce (1970) have shown that, when calculating r_1 for the residuals from a fitted ARMA process, $2/\sqrt{N}$ will generally provide an *over-estimate* of the critical value. It seems a reasonable conjecture that this may also apply to the residuals from a fitted HW seasonal model although the updating procedure for estimating parameters may introduce additional autocorrelation. I have not attempted a theoretical analysis as I do not "believe" any of the models to be completely accurate. Rather I see the method described above as a robust way of treating any departures from the HW model and suggest $2/\sqrt{N}$ as an approximate critical value for testing whether values of r_1 are significantly different from zero.

In Table 3 we see that the values of r_1 for Series A, B, D, F and G are significantly different from zero in the fitting period, indicating that it should be possible to improve the HW forecasts for these series. The modified HW procedure was tried over the forecasting period for all these five series. For series B the modification gave slightly worse results—a 2 per cent increase in mean-square error. But for Series A, D, F and G substantial improvements were made, averaging a 23 per cent reduction in mean-square error. Averaged over all seven series, the improvement is 12 per cent. This compares with an average improvement of 20 per cent reported by Newbold and Granger (1974) when combining HW with stepwise autoregression. Given the simplicity of the modified HW procedure and Reid's earlier success with a similar modification, the method seems worthy of further consideration. Indeed I suggest that the value of r_1 should be routinely calculated in the fitting period to see if the modification is worth trying.

As the modified HW procedure depends on four parameters, namely α, β, γ and λ, it might appear advantageous to estimate the parameters simultaneously in the fitting period. However, this gave virtually the same values for our seven series as were obtained by estimating α, β and γ by the original method and taking $\hat{\lambda} = r_1$. As the latter procedure involves minimizing a sum of squares in only three dimensions, it will be preferred.

For a k-steps-ahead predictor, the modification resulting from the above approach will become $\lambda^k e_t$, which quickly becomes small as k increases, so that the modification will not help in forecasting several steps ahead. In this connection, we note that Newbold and Granger found that the advantage enjoyed by BJ over automatic HW also decreased with k.

6. Series A

Although the HW procedure has been applied to all seven series in our set, it is not at all clear that it is an appropriate method for Series A which is non-seasonal and shows no regular trend. Rather the mean level moves up and down in a "random" way which experienced time-series analysts will recognize as looking something like a random walk.

For a random walk, successive first differences form a purely random process (or discrete white noise) and the optimal predictor of the observation at time $(t+1)$ is simply the previous observation, namely x_t. Over the forecasting period, this gives a mean square error which is better than the standard HW procedure though not quite as good as the modified HW procedure.

TABLE 4

The autocorrelation function of the first differences of Series A

Lag	*Autocorrelation coefficient*	*Lag*	*Autocorrelation coefficient*
0	1	7	−0·10
1	0·32	8	−0·14
2	0·03	9	−0·03
3	0·09	10	−0·01
4	0·00	11	0·01
5	0·01	12	0·09
6	−0·01	13	0·00

The random walk forecast can be further improved as there tend to be several increases or several decreases in a row. The autocorrelation function of the differenced series is shown in Table 4, and we see that the coefficient at lag one is the only one significantly different from zero. This indicates that the first differences form a first-order MA process or, equivalently, that x_t follows an ARIMA model of order (0, 1, 1) namely

$$x_t - x_{t-1} = \nabla x_t = \varepsilon_t + \theta \varepsilon_{t-1}, \tag{6.1}$$

where $\{\varepsilon_t\}$ denotes a discrete-time (unobservable) white-noise process. The value of θ was estimated by least squares over the fitting period to be 0·36. The resulting one-step-ahead prediction made at time t is

$$\hat{x}(t, 1) = x_t + 0{\cdot}36 e_t,$$

where $e_t = x_t - \hat{x}(t-1, 1)$, and the mean-square error over the forecasting period turned out to be 31 per cent better than the standard HW procedure and 8 per cent better than the modified HW procedure.

Here the Box–Jenkins identification procedure is surprisingly easy: partly because the series is non-seasonal, partly because first differencing is adequate and partly because only a one-parameter model is indicated. The necessary computer programs were written in less than a day.

Now, simple exponential smoothing is actually optimal for an ARIMA (0, 1, 1) process (e.g. Granger and Newbold, 1977, p. 172), and simple exponential smoothing is widely thought to be a special case of the HW procedure, so why was I getting different mean-square errors from my HW and BJ computer programs? The answer was simple yet far-reaching. Simple exponential smoothing is a special case of the non-seasonal HW model only if the initial trend value is set equal to zero. When this was done, the best estimate of γ for Series A turned out to be zero and identical forecasts from the HW and BJ methods were obtained with the smoothing constant α estimated to be 1·36.

Now the HW smoothing constants α, β and γ are usually constrained to the range (0, 1). But simple exponential smoothing, namely

$$\hat{x}(t, 1) = \alpha x_t + (1-\alpha)\,\hat{x}(t-1, 1) \tag{6.2}$$

is optimal for

$$\nabla x_t = \varepsilon_t - (1-\alpha)\,\varepsilon_{t-1}. \tag{6.3}$$

Comparing (6.3) with (6.1) we see that $\theta = \alpha - 1$. Now (6.1) is invertible for $|\theta| < 1$, or equivalently for $0 < \alpha < 2$ (see Box and Jenkins, 1970, p. 107). When α lies in the range $(1, 2)$ the simple exponential smoothing predictor given by (6.2) can still be expressed in the form

$$\hat{x}(t, 1) = \alpha x_t + \alpha(1-\alpha)\, x_{t-1} + \alpha(1-\alpha)^2\, x_{t-2} + \dots .$$

The sum of the coefficients still converges but the weights now oscillate in sign. Here the BJ procedure provides valuable insight into what is essentially a generalization of simple exponential smoothing. Newbold and Granger restricted the HW parameter α to the range $(0, 1)$ and this explains why their BJ forecasts were better.

7. Starting Values

The analysis of the Series A data highlighted the need to take a careful look at starting-up values. I subsequently found that the Newbold–Granger starting values (see Granger and Newbold, 1977, p. 165) were as follows:

(a) Mean: the average observation in the first year, as also suggested by Winters (1960), and used in my first run.

(b) Trend: this was set to zero, unlike the recommendations of Winters (1960) and Chatfield (1975). In my first run I used the average monthly difference between the first and second years' averages.

(c) Seasonal factors: these were calculated from the first year's data only, by comparing each observation with the overall average in the first year. No adjustments are made for trend as the initial trend value is set equal to zero. Winters (1960) and Montgomery and Johnson (1976) suggest averaging over the whole of the fitting period with a trend adjustment, while my first run used averages over the first two years with a trend adjustment.

I tried the Newbold–Granger starting values on all seven series using the automatic procedure. Over the fitting period there was little difference in mean-square error on average, though there were sizeable differences for individual series. For example, Series F was 55 per cent worse using the Newbold–Granger values, while Series A and D were 26 per cent and 16 per cent better respectively. Over the forecast period, the Newbold–Granger values did somewhat better on average. Series F was 41 per cent better while Series A was 14 per cent better if α is restricted to the range $(0, 1)$ but 23 per cent better if α is allowed to take the value 1·36. The estimated smoothing constants were much the same except for the value of γ. For Series A we have already seen that the value was reduced from 0·2 to 0·0, while for Series D the value increased from 0·1 to 0·2.

The choice of starting values clearly requires more investigation and we need more experience with other sets of data. The choice will depend to some extent on the properties of the series. For data like Series A, which contain no steady trend, it is better to set the initial trend value to zero. The advantage of the Newbold–Granger starting values is that they can also adapt to cope with the situation where trend is present. Thus, on the grounds of simplicity and accuracy, I am inclined to recommend the use of the Newbold–Granger starting values, though other starting values may sometimes be appropriate.

I repeated the comparison of the automatic and non-automatic HW procedures using the Newbold–Granger starting values. Once again, it was easy to improve the automatic forecasts using subjective judgement. In particular, when a non-seasonal model was used for Series B, an even larger reduction in mean-square forecast error was obtained than when using the previous starting values. It was also found that the one-step-ahead forecast errors tended to be autocorrelated so that further reductions in mean-square error could be made by the method described in Section 5.3. Thus the non-automatic HW procedure is superior to the automatic version for both sets of starting values.

8. Discussion and Conclusions

The analyses reported in this paper suggest the following.

(a) Series for which BJ forecasts are much better than forecasts from the automatic HW procedure do not appear to have any common properties.

(b) It is often easy to improve automatic HW forecasts by simple subjective modifications.

Now, earlier empirical studies have concentrated on comparing the BJ procedure with automatic procedures, although they are at opposite extremes of complexity and really have different purposes. In this paper we have concentrated on a non-automatic version of the HW procedure whose complexity is in-between these two extremes. The differences from the automatic procedure are as follows.

(a) A careful subjective choice is necessary to choose the correct seasonal model, either additive, or multiplicative or non-seasonal. It is particularly important not to use a seasonal model for non-seasonal data.

(b) Subjectively adjust any outliers due to known effects. Also consider omitting the observations in the early part of the series if their properties are different to those of the later part of the series.

(c) If the forecast errors are correlated, fit a first-order autoregressive model to them. It is suggested that the first-order autocorrelation coefficient of errors in the fitting period should be calculated routinely.

Although these subjective modifications require some effort, it is clearly much less than that required by the full BJ procedure.

Apart from the above suggestions, this paper also makes the following practical recommendations for the HW procedure, whether used in its automatic or non-automatic form.

(a) Starting values for the mean, trend and seasonal factors may be calculated from the first year's data only, with the initial trend-value set equal to zero.

(b) The "typical" values for the HW smoothing parameters which are quoted in the literature are often a long way from the estimated values, and it is recommended that the smoothing parameters should always be estimated rather than guessed.

In conclusion I would like to make some general recommendations regarding the choice of a univariate forecasting procedure. These are based, not only on my experience with the series analysed in this paper, but also on my experience with other series and on the many analyses reported in the literature by other authors.

(a) Automatic procedures are appropriate when there are a large number of items to forecast. For example, in production planning and stock control there may be several hundred (or even several thousand) series to consider. The automatic version of the HW procedure seems as good as any.

(b) Non-automatic procedures are appropriate in most other situations, as few series are so standard that it is safe to treat them automatically. There are many such procedures to choose from, but my experience is mainly with the non-automatic version of the HW procedure and the BJ procedure. I have seen little evidence to suggest that any other method is definitely superior to either or both of these methods, although several other methods are comparable in accuracy and the reader should use one he feels happy with. Restricting attention to non-automatic HW and BJ, we should recognize that both have a place in the forecaster's toolbag and that the choice between them is not easy. Sometimes practical considerations will rule out the BJ procedure as, for example, if there are insufficient observations or insufficient expertise available. If this is not the case then I suggest the following.

(i) For series in which the variation is dominated by trend and seasonal variation (e.g. those in Figs 2–7), I recommend the use of the non-automatic HW procedure. The BJ procedure, which requires considerably more effort, will rarely give much improvement in fit. This is because the effectiveness of the BJ procedure will here depend primarily on the intial differencing procedure rather than on the ARMA model-fitting stage (see also the remarks of Akaike, 1973, and Parzen, 1974, p. 725).

(ii) For series in which the random variation is "large" compared with trend and seasonal variation, the BJ procedure is worth trying. This applies particularly to series, like that in Fig. 1, which show a random-walk type of behaviour.

(iii) For series showing discontinuities, like that in Fig. 8, there is little point in trying time-series projection methods. The author was recently asked to produce univariate forecasts for this series, but refused. The sudden jump corresponds to a sales drive. If forecasts are required, then informed guesswork is likely to be better than statistical projections. A useful rule-of-thumb is that if you think you can produce good forecasts for a series "by eye", then statistical projections will probably work well. But if you can't, then they won't!

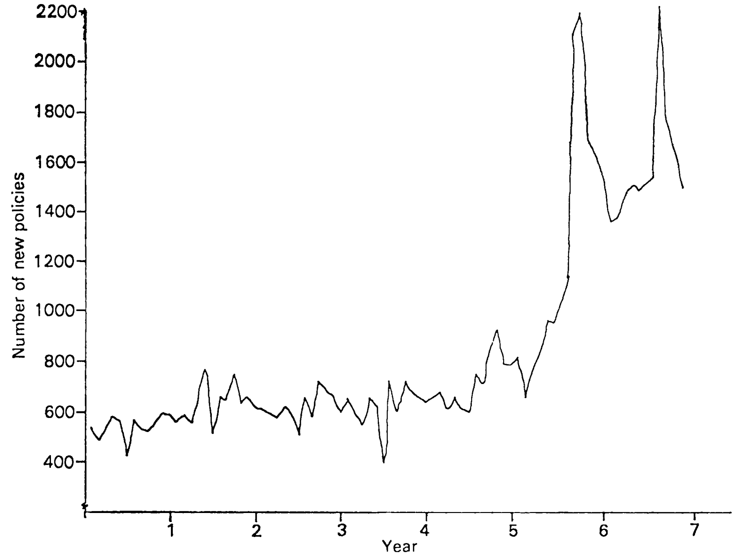

FIG. 8. Numbers of new insurance policies issued by a particular life office.

ACKNOWLEDGEMENTS

I am grateful for constructive comments on earlier versions of this paper by a referee, A. S. C. Ehrenberg, E. McKenzie, P. Newbold, G. J. A Stern and K. D. C. Stoodley, the latter having reminded me of the connection between exponential smoothing and an ARIMA (0, 1, 1) model.

REFERENCES

AKAIKE, H. (1973). Contribution to the discussion of the paper by Chatfield and Prothero *J. R. Statist. Soc*, A, **136**, 330.

BOX, G. E. P. (1970). Book review of De Bruyn's book on cusum charts. *Rev. Int. Statist. Inst.*, **38**, 305.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time-series Analysis, Forecasting and Control.* San Francisco: Holden-Day (rev. edn publ. 1976).

BOX, G. E. P. and PIERCE, D. A. (1970). Distribution of residual autocorrelations in ARIMA time-series models. *J. Amer. Statist. Ass.*, **65**, 1509–1526.

CHATFIELD, C. (1975). *The Analysis of Time Series: Theory and Practice*. London: Chapman and Hall.
—— (1977). Some recent developments in time-series analysis. *J. R. Statist. Soc.* A, **140**, 492–510.
CHATFIELD, C. and PROTHERO, D. L. (1973). Box–Jenkins seasonal forecasting: Problems in a case-study. *J. R. Statist. Soc.* A, **136**, 295–336.
COOPER, J. P. and NELSON, C. R. (1975). The ex-ante prediction performance of the St Louis and FRB–MIT–PENN econometric models, and some results on composite predictions. *J. of Money, Credit and Banking*, **7**, 1–31.
COUTIE, G. A. *et al.* (1964). *Short-term Forecasting*. I.C.I. monograph No. 2. Edinburgh: Oliver and Boyd.
DURBIN, J. and MURPHY, M. J. (1975). Seasonal adjustment based on a mixed additive-multiplicative model. *J. R. Statist. Soc.* A, **138**, 385–410.
GILCHRIST, W. (1976). *Statistical Forecasting*. London: Wiley.
GRANGER, C. W. J. (1969). Prediction with a generalized cost of error function. *Op. Res. Quart.*, **20**, 199–207.
GRANGER, C. W. J. and NEWBOLD, P. (1973). Some comments on the evaluation of economic forecasts. *Appl. Econ.*, **5**, 35–47.
—— (1977). *Forecasting Economic Time Series*. New York: Academic Press.
GROFF, G. K. (1973). Empirical comparison of models for short range forecasting. *Man. Sci.*, **20**, 22–31.
HARRISON, P. J. and STEVENS, C. F. (1975). Bayes forecasting in action: case studies. Warwick: Statistics Research Report No. 14.
JENKINS, G. M. (1974). Contribution to the discussion of the paper by Newbold and Granger. *J. R. Statist. Soc.* A, **137**, 148–150.
MCKENZIE, E. (1976). A comparison of some standard seasonal forecasting systems. *The Statistician*, **25**, 3–14.
MONTGOMERY, D. C. and CONTRERAS, L. E. (1977). A note on forecasting with adaptive filtering. *Op. Res. Quart.*, **28**, 87–91.
MONTGOMERY, D. C. and JOHNSON, L. A. (1976). *Forecasting and Time-series Analysis*. New York: McGraw-Hill.
NEWBOLD, P. and GRANGER, C. W. J. (1974). Experience with forecasting univariate time-series and the combination of forecasts. *J. R. Statist. Soc.* A, **137**, 131–165.
PARZEN, E. (1974). Some recent advances in time-series modelling. *Trans. I.E.E.E. on Automatic Control*, **AC-19**, 723–730.
REID, D. J. (1975). A review of short-term projection techniques. In *Practical Aspects of Forecasting* (H. A. Gordon, ed.), pp. 8–25, London: Operational Research Society.
WINTERS, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Man. Sci.*, **6**, 324–342.

APPENDIX

The additive seasonal HW model assumes that the observation at time t, x_t, is given by

$$x_t = (\text{local mean}) + (\text{seasonal factor}) + \text{error},$$

while the multiplicative model assumes

$$x_t = (\text{local mean}) \times (\text{seasonal factor}) + \text{error}.$$

Both models assume an additive trend term such that

$$(\text{local mean at time } t) = \{\text{local mean at } (t-1)\} + \text{local trend}.$$

In practice the seasonal effect may be intermediate between additive and multiplicative, as in the models described by Durbin and Murphy (1975), but the Holt–Winters procedure requires one to choose either an additive or multiplicative model. Both models also assume an additive error of constant variance, but in practice the error variance may increase with the mean. If both the seasonal and error terms are thought to be multiplicative, then the logarithms of the data may be analysed using the additive model, although it should be noted that this effectively assumes that the trend term is also multiplicative.

In order to describe the updating and forecasting procedures, we introduce the following notation:

m_t = estimate of the deseasonalized mean level at time t;

F_t = estimated seasonal factor for period t;

r_t = estimated trend term for period t (i.e. expected increase or decrease in the deseasonalized mean level in one time period);

s = number of observations in seasonal cycle (so that, for example, $s = 12$ for monthly data).

For the updating procedure, let us assume that, from the data up to period $(t-1)$, we have estimates of the mean and trend in period $(t-1)$ and also of the seasonal factors up to period $(t-1)$. Earlier estimates of the mean and trend are not required, but note that the last s seasonal factors need to be stored. When a new observation, x_t, becomes available, the mean, seasonal and trend terms are all updated using three smoothing constants which we denote by α, β and γ. In the multiplicative case (which is the most commonly used), the updating formulae are

$$m_t = \alpha x_t/F_{t-s} + (1-\alpha)(m_{t-1}+r_{t-1}), \tag{A.1}$$

$$F_t = \beta x_t/m_t + (1-\beta) F_{t-s}, \tag{A.2}$$

$$r_t = \gamma(m_t - m_{t-1}) + (1-\gamma) r_{t-1}. \tag{A.3}$$

Then forecasts from time t can be made using the formula

$$\begin{aligned} \hat{x}(t,h) &= \text{forecast of } x_{t+h} \text{ made at time } t \\ &= (m_t + hr_t) F_{t-s+h} \quad (h = 1, 2, \ldots). \end{aligned} \tag{A.4}$$

Formulae (A.1)–(A.4) can readily be adapted for the additive case. (A.3) stays the same, but (A.1), (A.2) and (A.4) become

$$m_t = \alpha(x_t - F_{t-s}) + (1-\alpha)(m_{t-1}+r_{t-1}), \tag{A.5}$$

$$F_t = \beta(x_t - m_t) + (1-\beta) F_{t-s} \tag{A.6}$$

and

$$\hat{x}(t,h) = m_t + hr_t + F_{t-s+h}. \tag{A.7}$$

The seasonal factors need to be normalized after each year's cycle by making their average one in the multiplicative case and by making them sum to zero in the additive case.

The reader should be warned that the above notation is by no means standard. Nearly every author seems to use a different set of symbols and so great care is needed when comparing expressions given by different authors. For example, the smoothing constants are variously given as (A, B, C) by Newbold and Granger (1974), (A, D, C) by Granger and Newbold (1977), $(\gamma_1, \gamma_3, \gamma_2)$ by McKenzie (1976), while Reid (1975) used α, β, γ but in each case "his parameter" = one − "my parameter". Even more confusing is the fact that s_t is variously used to denote the mean (Winters, 1960), trend (Reid, 1975) and seasonal variation (Chatfield, 1975).

Starting-up values are needed for the mean, trend and seasonal terms and these are discussed in Section 7. Estimates of the smoothing constants are also required as described in Section 4.