The Statistician (1988) 37, pp. 129–140

Holt-Winters forecasting: some practical issues

CHRIS CHATFIELD & MOHAMMAD YAR

School of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom

Abstract. The Holt-Winters forecasting procedure is a variant of exponential smoothing which is simple, yet generally works well in practice, and is particularly suitable for producing short-term forecasts for sales or demand time-series data. Some practical problems in implementing the method are discussed, including the normalisation of seasonal indices, the choice of starting values and the choice of smoothing parameters. There is an important distinction between an automatic and a non-automatic approach to forecasting and detailed suggestions are made for implementing Holt-Winters in both ways. The question as to what underlying model, if any, is assumed by the method is also addressed. Some possible areas for future research are then outlined.

1 Introduction

The calculation of reliable forecasts is an important task in many areas. There are many different types of forecasting procedure, but no simple answer as to which is 'best'. There is a wide variety of forecasting problems requiring different treatment and so the choice of method depends on a variety of considerations including the objectives, the properties of the data, the number of series to be forecast, and so on—see Chatfield (1988).

Suppose we have an observed time series, denoted by $X_1, X_2 \ldots, X_n$ and wish to forecast X_{n+k}. The forecast made at time n, k steps ahead, will be denoted by $\hat{X}_n(k)$. Short-term forecasts are often based on univariate (or projection) methods in which $\hat{X}_n(k)$ depends only on $X_n, X_{n-1} \ldots$. They contrast with multivariate (or causal) methods where $\hat{X}_n(k)$ depends also on past values of one or more predictor variables. Another important distinction is between an automatic forecasting method, which requires no human intervention, and a non-automatic approach which requires subjective input from the forecaster. Many univariate methods can be made fully automatic but can also be used in a non-automatic form.

This paper is concerned with a variant of exponential smoothing which is often known as the Holt-Winters method, in honour of the pioneering work of C. C. Holt and P. R. Winters. The method allows data to be modelled by a local mean, a local trend and a local seasonal factor which are all updated by exponential smoothing. The seasonal effect may be additive or multiplicative, while a non-seasonal version is also available.

The Holt-Winters method is a robust, easy-to-use projection procedure which has been around for over 20 years and generally works quite well in practice. Given its apparently well-established position, it is perhaps surprising that many practical and theoretical issues remain unresolved. The main objective of this paper is to investigate these issues. In any case the Holt-Winters method is not as well known as it might be (Fildes & Lusk, 1984, Table 1), so that a second objective is to publicise the method. In addition the method is perceived as being less accurate than other methods (Fildes & Lusk, 1984, Table 2) despite empirical evidence to the contrary, so that a third objective is to improve its 'image'. Some early studies (e.g. Newbold & Granger, 1974) did suggest that the Box-Jenkins method gives more accurate results, but Holt-Winters has performed consistently well in more recent forecasting compositions, when compared with other more complicated projection methods (e.g. Makridakis *et al.*, 1982, 1984; Armstrong & Lusk, 1983), and also in case studies, even when compared with multivariate methods (e.g. Thury, 1985). For example, Bodo & Signorini (1987) found that Holt-Winters predicted the Italian production index better than ARIMA models (adjusted for trading-day variations) but that somewhat better forecasts could be found using much more complicated multivariate forecasts. These authors conclude that the univariate forecasts are "satisfactory given their low cost, and represent a benchmark against which to evaluate more sophisticated procedures".

When interpreting the results of forecasting competitions, it should also be borne in mind that the Holt-Winters method has usually been applied in a completely automatic way, in contrast to the Box-Jenkins approach which usually requires considerable subjective skill on the part of the analyst (although an automatic version is now available). We believe there is a need for simple projection methods, like Holt-Winters, which can be applied either in an automatic or non-automatic way, depending on the context, and that this fundamental difference in approach deserves more emphasis. An automatic approach is appropriate if for example the analyst's skill is limited or if there are a large number of series to handle as in inventory control. On the other hand, a non-automatic approach is appropriate with a small number of series where the forecaster can use his skill and background knowledge. For example Chatfield (1978) shows how automatic Holt-Winters forecasts can easily be improved using subjective judgement.

Of course, much research only leads to small improvements in accuracy and it may be thought that the law of diminishing returns applies to the work considered here. In fact small improvements in accuracy can be very important and may for example lead to substantial savings in stock control. Thus this research *is* worth pursuing.

Finally, it is worth reminding ourselves that forecasts are conditional statements about the future of the form: 'If such-and-such behaviour continues, then . . . '. Thus forecasts are not sacred and we should be prepared to modify them in the light of any external knowledge. It is worth reading Chapter 15 of Schumacher (1974), and noting in particular his view that a refined technique for short-term forecasting "rarely produces significantly different results from those of a crude technique", and that "Long-term forecasts are presumptuous. However long-term feasibility studies, based on clearly stated assumptioms, are well worth doing".

2 Preliminaries

In simple exponential smoothing, the one-step-ahead predictor can be written in the *recurrence* form

$$\hat{X}_t(1)=\alpha X_t+(1-\alpha)\hat{X}_{t-1}(1) \tag{1}$$

where α is the smoothing parameter, usually constrained so that $0<\alpha<1$, or in the equivalent *error-correction* form

$$\hat{X}_t(1)=\hat{X}_{t-1}(1)+\alpha e_t \tag{2}$$

where $e_t=X_t-\hat{X}_{t-1}(1)$ is the one-step-ahead forecast error at time t.

Holt-Winters generalises this approach to deal with trend and seasonality. Trend is difficult to define but we take it to be 'long-term change in the mean level per unit time'. If trend is thought to be linear, it is important to distinguish between a *global* linear trend of the form

$$\text{mean level at time } t=\mu_t=a+bt \tag{3}$$

where a, b are constant, and a *local* linear trend of the form

$$\mu_t = a_t + b_t t \tag{4}$$

where a_t, b_t change slowly through time in a random way. It is the quantity b (or b_t) which is the trend.

Modern thinking (e.g. Newbold, 1988) favours local linearity rather than global linear regression on time, and local linearity is included explicitly in many models such as model (13) in Section 6 and Harvey's (1984) structural models. Local linearity is also implicit in ARIMA modelling (e.g. Box, Pierce & Newbold, 1987).

Regarding seasonality, the main distinction is between additive and multiplicative seasonal factors, the latter being appropriate when the magnitude of the seasonal variation is proportional to the local mean.

3 The Holt-Winters method

A description of the method is given for example by Gardner (1985). The literature is confused by many different notations, particularly that S_t is sometimes used to denote the local mean level and sometimes the local seasonality. We therefore avoid the use of S_t and denote the local mean Level, Trend, and seasonal Index at time t by L_t, T_t, I_t respectively. Let α, γ, δ denote the smoothing parameters for updating the mean level, trend and seasonal index respectively and let p denote the number of observations per seasonal cycle. Then, in the multiplicative case, the formulae for updating L_t, T_t and I_t, when a new observation X_t becomes available, are

$$L_t = \alpha(X_t/I_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{5}$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \tag{6}$$

$$I_t = \delta(X_t/L_t) + (1-\delta)I_{t-p} \tag{7}$$

and the new forecast made at time t of the value k periods ahead is given by

$$\hat{X}_t(k) = (L_t + kT_t)I_{t-p+k} \tag{8}$$

for $k = 1, 2 \ldots, p$. There are analogous formulae for the additive case.

Another source of confusion is that equations (5)–(7) are often presented in the equivalent error-correction form such as

$$T_t = T_{t-1} + \alpha\gamma e_t/I_{t-p} \tag{9}$$

where the multiplier '$\alpha\gamma$' may also be called a smoothing parameter (e.g. Coutie *et al.*, 1964).

Simple exponential smoothing is the 1-parameter case where L_t alone is updated. In the 2-parameter case, L_t and T_t are updated and this is sometimes called Holt's linear trend model. It is suitable for non-seasonal data. The full 3-parameter case is sometimes called Winter's seasonal model.

In order to implement the method, the user must

(a) provide starting values for L_t, T_t, I_t at the beginning of the series,
(b) provide values for α, γ, δ,
(c) decide whether to normalise the seasonal factors, and
(d) choose between an automatic or non-automatic approach.

Despite much reference to the Holt-Winters method, the literature is generally unhelpful on these important practical issues. Partly as a result, software packages adopt widely different approaches which can produce substantially different forecasts for what is ostensibly the same method. Thus, this paper attempts to make some practical recommendations.

4 Some practical issues

4.1 Normalising the seasonal indices

The seasonal factors can be normalised to sum to zero in the additive case, or average to one in the multiplicative case. Winters (1960) originally recommended that the seasonal factors should be normalised at the start of the series but not thereafter, so that they would soon "contain some of the trend effect". Later authors suggested that they should be normalised once a year or even at every observation in order to be "consistent with any reasonable definition of seasonality" (Roberts, 1982). Mckenzie (1986, p. 378) says that the choice is "not clear" but that equivalent forecasts can be obtained in the additive case with both normalised and non-normalised systems by choosing smoothing parameters which correspond in an appropriate way.

In view of Mckenzie's results, the normalisation question is probably not important. Even so it is a pity that software packages rarely allow normalisation/non-normalisation as an option or even say which one is being used. The AUTOCAST package (Gardner, 1986) is a welcome exception and our empirical experience with it suggests that the normalisation question only matters when the seasonal pattern changes considerably and that a slightly better fit can be obtained on average with normalisation. In view of these empirical results and to ensure that the seasonal factors have a clear interpretation, not contaminated by trend, we recommend that the seasonal indices should be normalised at least once a year.

4.2 Choosing starting values

The user must provide starting values for the level, trend and seasonal indices at the beginning of the series in order to initiate the updating procedure. Many different ways of doing this have been suggested, but there is little advice in the literature as to which to choose. Gardner (1985, Section 2.4) is rather unhelpful.

One simple method (Granger & Newbold, 1986, Chapter 5) is to set L_0 equal to the average observation in the first year, namely

$$\bar{X}_1 = \sum_{t=1}^{p} X_t/p,$$

to set T_0 equal to zero, and to calculate $I_0, I_{-1}, \ldots, I_{-p+1}$ by comparing the appropriate observation in the first year with $\bar{X}_1$. For example $I_0 = (X_p - \bar{X}_1)$ in the additive case and $X_p/\bar{X}_1$ in the multiplicative case.

A less simplistic method calculates the initial trend from the average difference per time period between the first and second year averages and then calculates the initial level and seasonal indices after allowing for a trend adjustment. For example we could take

$$I_0 = \begin{cases} X_p - (\bar{X}_1 + (p-1)T_0/2) & \text{—additive case} \\ [X_p - (p-1)T_0/2]/\bar{X}_1 & \text{—multiplicative case} \end{cases} \tag{10}$$

Some limited empirical findings on these first two approaches are given by Chatfield (1978, Section 7). Other authors suggest calculating the starting values from the first 2 or 3 years data.

A different type of approach, used for example in the M-competition (Makridakis *et al.*, 1982), is *backcasting*, wherein the time-order of the data is reversed and the most recent data are used to start up the procedure in the reverse direction. The values at the 'end' (i.e. the beginning) of the series are then used as the starting values. This approach is analogous to backcasting in ARIMA-model-estimation (e.g. Box & Jenkins, 1970, p. 109) and has theoretical arguments (e.g. Ledolter & Abraham, 1984; Mckenzie, 1984) in its favour.

A third completely different type of approach is to calculate starting values from *all* the data in the fitting period. For example, Winters (1960) originally suggested calculating the initial trend from the average change per period between the first and last years data, namely

$$T_0=(\bar{X}_m-\bar{X}_1)/(m-1)p \tag{11}$$

where m years data are available. Winters (1960) and Montgomery & Johnson (1976) also suggest initialising the seasonal factors from some sort of average of the seasonal effects over the entire period of fit, and the program listed by Montgomery & Johnson (1976, Appendix C.2) uses this approach. In somewhat similar fashion one option for the AUTOCAST package (Gardner, 1986) fits a linear regression on time to all the (deseasonalised) data to get starting values.

These three types of starting value are quite different in character and may give rise to substantially different smoothing parameters when the latter are optimised (see Section 4.3). They may also give rise to differing forecasts, especially for series which are short or changing in structure. It is therefore surprising that there is little comparative discussion of starting values in the literature.

We will refer to starting values calculated from *all* the data assuming a global model as *global values*, and values calculated from the first part of the data as *initial values.* Backcasted values can be regarded as sophisticated initial values, and we suggest the basic choice is between global and initial values rather than, for example, between one type of initial value and another. The AUTOCAST manual (Gardner, 1986) says effectively that if the structure of the time series is changing, then initial values are more accurate, but if the structure is stable then global values are better. There is some empirical evidence for this in terms of one-step-ahead mean square error over the period of fit, but what about forecasts from the end of the fit period? For stable series, the forecasts may be rather similar so that the choice of starting values is not critical, but for changing series the forecasts using initial values may be better able to 'follow' the changes. In other words there may be little to lose and everything to gain by using initial values.

Global values may appear sensible when an underlying global trend is suspected as in equation (3), but there are then other more efficient ways of estimating the trend parameters. Holt-Winters is really designed to cope with a local linear trend, as in equation (4) and this suggests calculating starting values from observations near the start of the series only. If there is a global linear trend, then the updating procedure will soon follow it anyway, whereas, if there is not, then the use of global starting values will be silly as well as being a contradiction in terms.

For simple exponential smoothing, Clark *et al.* (1987) compared a global value (the mean of the series), a naive initial value (the first observation) and backcasting. They found that for most series the selection of a starting value was not important, although backcasting tended to provide slightly the more accurate forecasts. For 3-parameter Holt-Winters, the selection of starting values is more important, particularly for seasonal factors. The latter are only updated once a year and therefore have less chance to adjust. One commercial package we came across appears to set all initial seasonal factors equal to one and this gives very poor results.

We recommend that initial, rather than global, values be adopted. We further suggest that the Newbold-Granger values are too simplistic and recommend calculating values by backcasting or from the first 2 or 3 years data. We further suggest that package manuals should say clearly what starting values are used and if possible give more than one option.

4.3 Choosing the smoothing parameters

The user must also provide values for the three smoothing parameters, α, γ and δ. Some general advice on selecting sensible values is given for example by Gardner (1985, Section 2.3) and Armstrong (1985, p. 165). The parameters are usually constrained to the range (0, 1), although the stability region is smaller than this for $p>4$ (Sweet, 1985).

There are two general ways of selecting the parameters. The first is to estimate them by minimising some function of the forecast errors of historical data, while the second is to simply 'guestimate' them.

For the first approach, the usual procedure is to minimise the sum of squared one-step-ahead forecast errors over a suitable fitting period for which historical data are available. Given particular values of α, γ and δ, the updating procedure yields a sequence of one-step-ahead errors, $\{e_t\}$, so that Σe_t^2 may be calculated. A 'running-in' period of perhaps 2 years may be advisable, so that, with m years data, the sum is taken from $t=2p+1$ to $t=mp$. The sum of squared errors (or alternatively the mean square error—MSE—obtained by dividing by the relevant number of observations) is calculated over a suitable grid of values of (α, γ, δ) so that optimum values may be found. Unfortunately a grid search can be rather cumbersome, although it does enable the analyst to get a 'feel' for the shape of the sum-of-squares surface. An alternative approach is to use some sort of hill-climbing optimisation procedure but, while quicker, this is not always safe as the sum-of-squares surface is not necessarily convex. This means that a local, rather than global, optimum may be found. With a sensible starting value (e.g. $\alpha=\delta=0{\cdot}3$, $\gamma=0{\cdot}1$), this is usually not a problem, although some combination of the two approaches may be desirable. The FORECASTMASTER package (Goodrich & Stellwagen, 1986) uses a hill-climbing procedure which does not always converge even after 30 iterations, particularly when the poor default value of $\delta=0{\cdot}001$ is used. The AUTOCAST package uses a modified grid-search method which appears to work better. Unfortunately, many packages do not provide an option to optimise smoothing parameters. Then the analyst has to use trial-and-error or use the alternative 'guestimation' approach. The latter is generally unsatisfactory except with a very large number of series or with very short series where it may be unavoidable.

What general advice can be given on selecting (or guessing?) the smoothing parameters? It has often been suggested that values should be chosen in the range (0·1, 0·3), but this choice is arbitrary. In our experience (see also Chatfield, 1978), the values of α and δ often exceed 0·3, but the value of γ is often less than 0·1 and may even be zero, in which case the starting trend value is left unchanged. However, if the series exhibits exponential growth, then high values of α and γ are found which may even be as large as unity and this would indicate that the Holt-Winters method is unsuitable, as would the occurrence of any optimised smoothing parameters outside the stability region. Other properties of series which help to determine the smoothing parameters are the presence of outliers and the presence of high random variation, both of which tend to lower the optimal smoothing parameters, and the presence or absence of stability in the trend and seasonal variation. A changing structure generally leads to higher parameter values.

It is not generally realised that the optimum smoothing parameters are determined, not only by the properties of the series, but also by the choice of starting values. For example, the FORECASTMASTER package, which appears to use backcasting, gave optimum smoothing parameters for the famous airline data (Box & Jenkins, 1970, Series G) equal to (0·29, 0·03, 0·83) while AUTOCAST using global values gave (0·84, 0·0, 0·0). Both packages were checked as being 'correct' despite the enormous difference in parameter values. It appears that global starting values generally lead to smaller parameters, particularly if the series structure is stable. For stable series the sum of squares surface is usually very flat over a wide region so that the choice of smoothing parameters is not critical, although it should be realised that the estimates are often highly negatively correlated.

The choice of smoothing parameters may also be affected by the length of the series and of the selected running-in period. We suggest that the parameters should only be updated every year or two and that large adjustments should be avoided. The possibility of updating the smoothing parameters in a 'Trigg-and-Leach' sense is considered by Williams (1987) who shows that only α may be varied in this way if instability is to be avoided. Of course, the values which give the best fit do not necessarily give the best forecasts and a crude pseudo-Bayesian approach can be recommended in which optimum values near the extremes of zero or one are adjusted towards the more usual values.

It is also worth considering alternatives to minimising one-step-ahead MSE. For example, mean absolute error may correspond to a more sensible loss function and give a response surface which is less affected by outliers and the non-convexity problem, while if forecasts are required more than one step ahead (as they usually are!), then it may be more sensible (but more difficult) to minimise the k-steps-ahead MSE.

Despite the above guidelines, it is clearly difficult to guestimate the smoothing parameters and poor results may ensue if they are constrained to the interval (0·1, 0·3) despite the flatness of the response surface. If a single set of values has to be chosen (e.g. in automatic forecasting), then we recommend $\alpha=\delta=0{\cdot}4$ and $\gamma=0{\cdot}1$, but wherever possible the properties of the data and the choice of starting values should be taken into account.

5 Implementing the Holt-Winters method

5.1 An automatic approach

This section makes suggestions for implementing Holt-Winters when a completely automatic approach is required. Seasonality is more often multiplicative than additive, and so the multiplicative seasonal model is normally used for all series, although a non-seasonal model should be used for any groups of series which are known *a priori* to be non-seasonal. We recommend that starting values be calculated from the first year or two's data, or by using backcasting, and that smoothing parameters be optimised automatically over a suitable length of historical data. If the latter is not possible, then we recommend $\alpha=\delta=0{\cdot}4$ and $\gamma=0{\cdot}1$ as suitable guestimates. Implemented in this way, the Holt-Winters method will generally give reliable, robust forecasts but will of course, like other automatic projection methods, have difficulty in coping with unexpected features such as sudden changes in structure.

5.2 A non-automatic approach

Suppose now that a non-automatic approach is indicated. Chatfield (1988) discusses a variety of approaches including multivariate and judgemental forecasting as well as more complicated univariate procedures, including Box-Jenkins. However, forecasters sometimes overlook a simpler alternative, namely the 'thoughtful' use of procedures which are often regarded as automatic. A general strategy for implementing an appropriate univariate procedure is described by Chatfield (1988) and we concentrate here on matters relating particularly to the use of Holt-Winters. It is helpful to list the following steps:

(1) Get as much background information as necessary and carefully define the objectives. Find out how the forecasts are actually going to be used, and whether point or interval forecasts are required. The statistician must be prepared to ask lots of questions.

(2) Plot the series and examine its structure. This is arguably the most important step in time-series analysis and is an essential preliminary to perhaps adjusting and then modelling the data (e.g. Chatfield, 1985). Although plotting is apparently easy to do, the choice of scales and plotting symbol may not be as straightforward as expected (e.g. see Chatfield 1977, Fig. 1), and the production of 'good' graphs (and tables) deserves more attention. Look for trend, seasonal variation, outliers and changes in structure which may be slow or sudden. Holt-Winters is primarily suited for reasonably regular series whose variation is dominated by trend and seasonal variation. It is not suitable for series showing exponential growth, for series dominated by short-term correlation (where Box-Jenkins is worth considering) or for series showing major discontinuities. In the latter case consider the possibility of discarding the data observed before the discontinuity, or of fitting a Box-Jenkins intervention model or of accepting that no projection forecast is likely to be beneficial.

(3) Examine potential outliers and consider adjusting suspect observations preferably after taking account of external information. An additive error will affect not only the forecast error in the time period where it occurs, but also the forecasts in subsequent periods. Hillmer (1984) describes a simple method of adjusting observations for the effect of additive outliers, but there is the usual problem that outliers arise for a variety of reasons and a large forecast error may for example indicate a permanent intervention. Adjustments for calendar variations may also need to be considered (e.g. Cleveland, 1983).

(4) Assess the form of the seasonal variation, if any, by examining the time-plot, and choose the most appropriate exponential smoothing method. Consider the possibility of transforming the data (e.g. to make the seasonal effect additive) but bear in mind that a model for the raw data is normally more helpful and that a forecast for a transformed variable has to be 'transformed back' with a possible biasing effect.

(5) Calculate initial starting values for the mean, trend and seasonal indices and estimate the smoothing parameters as described in Sections 4.2 and 4.3.

(6) Check the adequacy of the method by examining the one-step-ahead forecast errors, particularly their autocorrelation function. If the errors are autocorrelated, then Holt-Winters is not optimal so that it may be wise to try a completely different forecasting method or a different variant of Holt-Winters. Alternatively, if the first-order autocorrelation coefficient, r_1, is 'large' (either positive or negative), then Chatfield (1978, Section 5.3) suggests equation (10) may be modified to

$$\hat{X}_t(k)=(L_t+kT_t)I_{t-p+k}+r_1^k e_t$$

A related topic is the automatic monitoring of forcasts using techniques such as cusum or tracking-signal charts (e.g. Montgomery & Johnson, 1976, Chapter 7; Gardner, 1983).

(7) Compute forecasts for as far ahead as required. Decide if the forecasts need to be adjusted subjectively for any reason.

The analyst who adopts the above strategy will generally get reasonable results, although there will of course be some situations where alternative non-automatic approaches are indicated.

6 Relationship with other methods

This section considers the relationship between the Holt-Winters method and other projection procedures, notably the Box-Jenkins approach (e.g. Box & Jenkins, 1970; Vandaele, 1983) and the use of state-space or structural models (e.g. Harvey, 1984). Alternative discussions are given by Abraham & Ledolter (1986) and Newbold (1988).

It is well-known that simple exponential smoothing is optimal for an ARIMA (0, 1, 1) model, while two-parameter (non-seasonal) smoothing is optimal for an ARIMA (0, 2, 2) model. It has sometimes been suggested (e.g. Jenkins, 1974) that all exponential smoothing methods should be regarded as special cases of (Box-Jenkins) ARIMA-modelling, but this view has been discredited in recent years. In particular the 3-parameter Holt-Winters method, with additive seasonality, is optimal for an ARIMA-model (see Mckenzie, 1976; Abraham & Ledolter, 1986) which is so complicated that it would never be identified in practice, while multiplicative Holt-Winters does not have an ARIMA equivalent at all. In any case the methods are implemented in different ways even when there is an equivalent ARIMA model in theory (e.g. see the discussion of Chatfield's (1978) Series A). More fundamentally, the use of differencing to remove trend and seasonality is increasingly being challenged (e.g. Harvey & Durbin, 1986), although it can be argued that Holt-Winters and structural modelling have forecast functions which imply differencing which is no less real for not being explicit.

As regards state-space or structural models, it is well-known that simple exponential smoothing is optimal for the no-trend model given by

$$\left.\begin{aligned} X_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + a_t \end{aligned}\right\} \tag{12}$$

where μ_t denotes the local level and ε_t, a_t are independent error terms. Holt-Winters two-parameter smoothing is well-known (e.g. Abraham & Ledolter, 1986, p. 59) to be optimal for the state-space linear growth model given by

$$\left.\begin{aligned} X_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + a_{1,\,t} \\ \beta_t &= \beta_{t-1} + a_{2,\,t} \end{aligned}\right\} \tag{13}$$

where β_t denotes the local trend. However, it is not clear what state-space model, if any, is the optimal model for Holt-Winters three-parameter smoothing. The basic structural model (abbreviated BSM) of Harvey (1984) models trend and seasonality explicitly in a somewhat similar way to the (additive) Holt-Winters method. The BSM is optimally updated using the Kalman filter, and, after suitable differencing, can be shown to be equivalent to a complicated MA($p+1$) process. If all factors, including the seasonal effect, are multiplicative, then logarithms may be taken so as to apply the BSM, but a distinctive feature of Holt-Winters is that it can have additive trend and multiplicative seasonality. Although the BSM is technically more complicated to implement (it requires a much longer computer program), there are many advantages to forecasting procedures based on a proper statistical model. In addition the BSM can handle data irregularities more easily and can be extended to incorporate explanatory variables. In view of these potential advantages, it is to be hoped that the BSM will be thoroughly tested in some future forecasting competition, although Holt-Winters seems likely to continue in use in view of its simplicity and known accuracy.

Method or model?

In view of the above discussion, it is instructive to ask what underlying model, if any, is assumed by Holt-Winters three-parameter smoothing. Writers often talk about the 'Holt-Winters model' when they really mean the 'Holt-Winters method'. Some authors (e.g. Montgomery & Johnson, 1976, Section 5.2; Bowerman & O'Connell, 1979;

Mckenzie, 1986) do explicitly assume an underlying model for the additive seasonal case which is:

$$X_t = a + bt + i_t + \varepsilon_t \tag{14}$$

where $\{\varepsilon_t\}$ are independent and identically distributed with zero mean, and the seasonal indices are such that $i_t = i_{t-p}$. However the assumptions of global trend, constant seasonality and independent errors are not necessary to apply the Holt-Winters method. Indeed since least-squares estimates will be optimal for model (14), there seems little point using non-optimal Holt-Winters if this model really is thought to be appropriate. Mckenzie (1986) derives prediction intervals for the additive Holt-Winters method assuming (14) to be true but we suggest these results are unhelpful and potentially misleading since Holt-Winters is really suitable for use with a local linear trend. Sweet's (1985) results for what he calls the "additive and multiplicative Holt-Winters seasonal *model*" (our italics) also need to be treated with caution.

7 Further developments

Despite the simplicity and widepsread use of Holt-Winters, much work remains to be done. One promising line of research is to investigate the damping of the trend term. In the non-seasonal case, Gardner & Mckenzie (1985, Equation (6) with the misprint corrected) suggest the forecast

$$\hat{X}_t(k) = L_t + \sum_{i=1}^{k} \varphi^i T_t \qquad (0 \le \varphi \le 1)$$

rather than $(L_t + kT_t)$. Initial empirical results (e.g. Schnaars, 1986) look promising and the theory has now been extended to the seasonal case (Gardner & Mckenzie, 1987).

Another topic under active investigation concerns ways of finding prediction intervals for k-step-ahead forecasts. Mckenzie (1986) and Sweet (1985) have derived results assuming an underlying global trend model, but, as noted in Section 6, this is not an appealing (or optimal) model. Johnston & Harrison (1986) have derived results for the optimal model corresponding to 2-parameter Holt-Winters while Yar & Chatfield (1988) have extended these results to the seasonal case.

As to theoretical questions, further research is needed to clarify the relationship with the basic structural model (see Section 6) and more generally with the use of state-space models, such as those involved in state-space forecasting, although our initial experience with the latter using the FORECASTMASTER package has been rather disappointing (but see Koehler & Murphree, 1988).

This leads us naturally to the final important development considered here, namely the increasing availability of many different forecasting packages. They should enable the Holt-Winters method to be more widely and easily used, but many users will become dependent upon packages for which the details are unclear or of dubious caliber. The criteria for good software need to be carefully defined and more reviews and information need to be disseminated. For example, some Holt-Winters packages do not allow the user to optimise smoothing parameters or have other restrictions. The authors welcome comments from readers on packages they have tried (good or bad!). The best Holt-Winters package we have tried so far is AUTOCAST (Gardner, 1986). FORECASTMASTER (Goodrich & Stellwagen, 1986) covers a wider range of methods but the Holt-Winters seasonal method is restricted to the multiplicative case.

Acknowledgement

This paper is based on research funded by the Economic and Social Research Council (ESRC).

References

ABRAHAM, B. & LEDOLTER, J. (1986) Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures, *International Statistical Review*, 54, pp. 51–66.

ARMSTRONG, J.S. (1985) *Long-Range Forecasting*, 2nd edn (New York, Wiley).

ARMSTRONG, J.S. & LUSK, E.J. (Eds.) (1983) Commentary on the Makridakis time series competition (M-competition), *Journal of Forecasting*, 2, pp. 259–311.

BODO, G. & SIGNORINI, L.F. (1987) Short-term forecasting of the industrial production index, *International Journal of Forecasting*, 3, pp. 245–259.

BOWERMAN, B.L. & O'CONNELL, R.T. (1979) *Time Series and Forecasting: an applied approach* (North Scituate, Mass., Duxbury Press).

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

BOX, G.E.P., PIERCE, D.A. & NEWBOLD, P. (1987) Estimating trend and growth rates in seasonal time series, *Journal of the American Statistical Association*, 82, pp. 276–282.

CHATFIELD, C. (1977) Some recent developments in time-series analysis, *Journal of the Royal Statistical Society, Series A*, 140, pp. 492–510.

CHATFIELD, C. (1978) The Holt-Winters forecasting procedure, *Applied Statistics*, 27, pp. 264–279.

CHATFIELD, C. (1985) The initial examination of data (with discussion), *Journal of the Royal Statistical Society, Series A*, 148, pp. 214–253.

CHATFIELD, C. (1988) What is the best method of forecasting?, *Journal of Applied Statistics*, 15, pp. 19–38.

CLARK, S.D., CREMINS, S.J. & SCHNAARS, S.P. (1987) A comparison of starting values for simple exponential smoothing, *7th International Symposium on Forecasting*, Boston, Massachusetts.

CLEVELAND, W.S. (1983) Seasonal and calendar adjustment, in: BRILLINGER, D.R. & KRISHNAIAH, P.R. (eds) *Handbook of Statistics*, 3, pp. 39–72 (Amsterdam, North-Holland).

COUTIE, G.A. *et al.* (1964) *Short-term Forecasting* (Edinburgh, Oliver & Boyd).

FILDES, R. & LUSK, E.J. (1984) The choice of a forecasting model, *Omega*, 12, pp. 427–435.

GARDNER, E.S. JR. (1983) Automatic monitoring of forecast errors, *Journal of Forecasting*, 2, pp. 1–21.

GARDNER, E.S. JR. (1985) Exponential smoothing: the state of the art, *Journal of Forecasting*, 4, pp. 1–28.

GARDNER, E.S. JR. (1986) *AUTOCAST User's Manual* (Bridgewater, NJ, Core Analytic).

GARDNER, E.S. JR. & MCKENZIE, E. (1985) Forecasting trends in time series, *Management Science*, 31, pp. 1237–1246.

GARDNER, E.S. JR. & MCKENZIE E. (1987) Seasonal forecasting systems with damped trends (in preparation).

GOODRICH, R.L. & STELLWAGEN, E.A. (1986) *FORECASTMASTER User's Manual* (Cambridge, Mass., Scientific Systems).

GRANGER, C.W.J. & NEWBOLD, P. (1986) *Forecasting Economic Time Series*, 2nd edn (New York, Academic Press).

HARVEY, A.C. (1984) A unified view of statistical forecasting procedures, *Journal of Forecasting*, 3, pp. 245–275.

HARVEY, A.C. & DURBIN, J. (1986) The effects of seat-belt legislation on British road casualties: a case study in structural modelling (with discussion), *Journal of the Royal Statistical Society, Series A*, 149, pp. 187–227.

HILLMER, S. (1984) Monitoring and adjusting forecasts in the presence of additive outliers, *Journal of Forecasting*, 3, pp. 205–215.

JENKINS, G.M. (1974) Contribution to the discussion of Newbold & Granger (1974), *Journal of the Royal Statistical Society, Series A*, 137, pp. 148–150.

JOHNSTON, F.R. & HARRISON, P.J. (1986) The variance of lead-time demand, *Journal of the Operational Research Society*, 37, pp. 303–308.

KOEHLER, A.B. & MURPHREE, E.S. (1988) A comparison of results from state-space forecasting with forecasts from the Makridakis Competition, *International Journal of Forecasting*, 4, pp. 45–55.

LEDOLTER, J. & ABRAHAM, B. (1984) Some comments on the initialization of exponential smoothing, *Journal of Forecasting*, 3, pp. 79–84.

McKenzie, E. (1976) A comparison of some standard seasonal forecasting systems, *The Statistician*, 25, pp. 3–14.

McKenzie, E. (1984) General exponential smoothing and the equivalent ARMA process, *Journal of Forecasting*, 3, pp. 333–344.

McKenzie, E. (1986) Error analysis for Winter's additive seasonal forecasting system, *International Journal of Forecasting*, 2, pp. 373–382.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting*, 1, pp. 111–153.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1984) *The Forecasting Accuracy of Major Time Series Methods* (New York, Wiley).

Montgomery, D.C. & Johnson, L.A. (1976) *Forecasting and Time Series Analysis* (New York, McGraw-Hill).

Newbold, P. (1988) Predictors projecting linear trend plus seasonal dummies, *The Statistician*, 37, pp. 247–263.

Newbold, P. & Granger, C.W.J. (1974) Experience with forecasting univariate time-series and the combination of forecasts (with discussion), *Journal of the Royal Statistical Society series A*, 137, 131–165.

Roberts, S.A. (1982) A general class of Holt-Winters type forecasting models, *Management Science*, 28, pp. 808–820.

Schnaars, S.P. (1986) A comparison of extrapolation models on yearly sales forecasts, *International Journal of Forecasting*, 2, pp. 71–85.

Schumacher, E.F. (1974) *Small is Beautiful* (London, Sphere Books).

Sweet, A.L. (1985) Computing the variance of the forecast error for the Holt-Winters seasonal models, *Journal of Forecasting*, 4, pp. 235–243.

Thury, G. (1985) Macroeconomic forecasting in Austria, *International Journal of Forecasting*, 1, pp. 111–121.

Vandaele, W. (1983) *Applied Time Series and Box-Jenkins models* (New York, Academic Press).

Williams, T.M. (1987) Adaptive Holt-Winters forecasting, *Journal of the Operational Research Society*, 38, pp. 553–560.

Winters, P.R. (1960) Forecasting sales by exponentially weighted moving averages, *Management Science*, 6, pp. 324–342.

Yar, M. & Chatfield, C. (1988) Prediction intervals for Holt-Winters seasonal forecasting (in preparation).