ISSN 1645-0647

Notes on Pragmatic Procedures and Exponential Smoothing

04-1998

António Ascensão Costa

Cemapre
Centro de Matemática Aplicada à Previsão e Decisão Económica

Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

NOTES ON
PRAGMATIC FORECASTING PROCEDURES AND EXPONENTIAL SMOOTHING

ANTÓNIO ASCENSÃO COSTA

ISEG - 1998

1 - INTRODUCTION

We use the term "pragmatic forecasting procedures" by opposition to model based approaches to forecasting (as ARIMA modelling and structural time series modelling). Pragmatic forecasting procedures were designed to solve the problem of generating good forecasts for certain general patterns of time series behaviour without clearly defining a data generating process' model nor the correspondent optimal forecast function.

Pragmatic forecasting procedures are extrapolative short term forecasting methods. Extrapolative are all univariate methods despite its statistical sofistication. All extrapolative methods have the implicit assumption that near future will be in the line of (at least) recent past. But, as things are changing it's better not to project actual conditions to far ahead and stay in short term horizons. What short term is is more difficult to define: practical needs are always pressing you to cover more periods ahead !

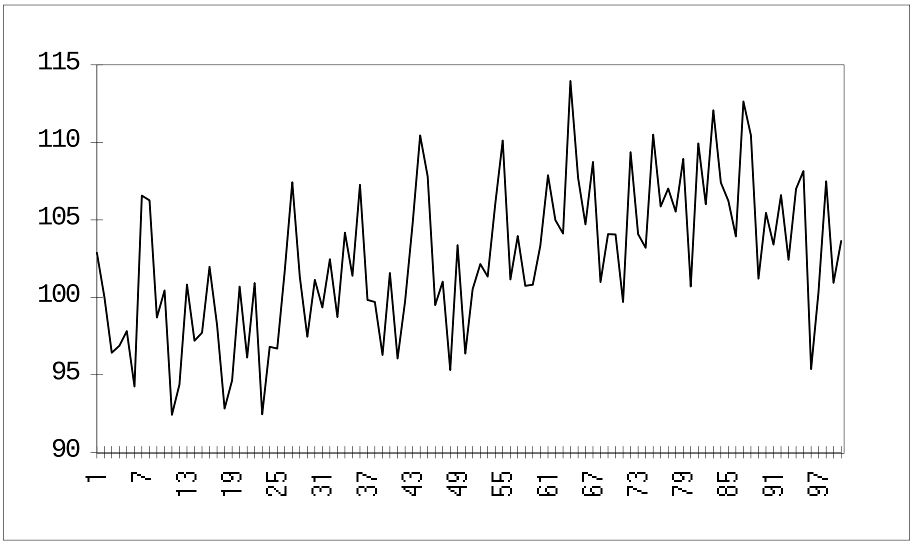
Pragmatic forecasting procedures all share the implicit (and also pragmatic) idea that a series has a set of components (trend, seasonals, and an error). If you conceive components as deterministic (an overall polinomial trend, a fixed seasonal pattern) a least squares regression on time and seasonal dummies is a good choice. But probably the error is red and forecasts black ! If you admit that components evolve according to a non specified pattern, then you need more flexible estimators. Estimators based on moving averages were a possible solution now in oblivion because estimators based in exponentially wheigted moving averages - exponential smoothing - were found superior and easy to use (and you don't even have to know that you are working with exponentially weighted moving averages).

Exponential smoothing - the set of forecasting methods that is known under the name of exponential smoothing (for a compreensive review see Gardner, 1985) - appeared by late fifties early sixties (Brown, 1959; Holt, 1960; Winters, 1960). Easy to use, understand and implement, even for large sistems, its popularity grew among users. Later, with the development of computational facilities, which also facilitates more flexible usage of exponential smoothing, more sofisticated methods entered into the competion but, among extrapolative time series methods, exponential smoothing was found hard to beat in forecasting perfomance. After all what came as a pragmatig forecasting procedure was found optimal for some data generating processes and very robust in general. So, it continues in use and it continues deserving atention both for application in large routine forecasting and for providing a perfomance benchmark for more sofisticated methods. Its behaviour shows that naïves are not as naïves as they used to be.

2 - "CONSTANT" LEVEL SERIES: SIMPLE SMOOTHING

Assume you want to forecast future values of a series that behave like the one shown in Graph 1:

Graph 1



The series has no trend and can be viewed as oscilations around a "locally constant level". The pragmatic solution to forecast future values is to estimate the level at the end of the series, let $\hat{a}_T$ be that estimate, and, since there is no trend, to project it into the future, i.e. the forecast function is:

$$\hat{Y}_{T+h} = \hat{a}_T \text{ , for } h = 1, 2, ...$$

Now, the problem is how to estimate the level. We can use means: an operator that filters erratic components. But since the level is changing it seems reasonable to think that recent values of the series are more important to estimate the actual level than long past ones. So, we want to discard some of the past information. This can be done by using ***moving averages***: only the last ***N*** observations enter to form the level estimate. And each time a new observation arrives the actual level estimate is revised: the most ancient observation is dropped and is replaced by the newcomer.

Simple exponential smoothing proposes another solution. Assume that at time ***t-1***, you have an estimate of the level, $\hat{a}_{t-1}$, and that at time ***t*** you observe y_t, and want to update the estimate of the level. Then, the new estimate may be formed via the recursion:

$$\hat{a}_t = \alpha y_t + (1-\alpha)\hat{a}_{t-1}, \text{ with } 0 < \alpha < 1 \; .$$

Actually, sistematic use of this estimator is equivalent to an *exponentially weighted moving average*:

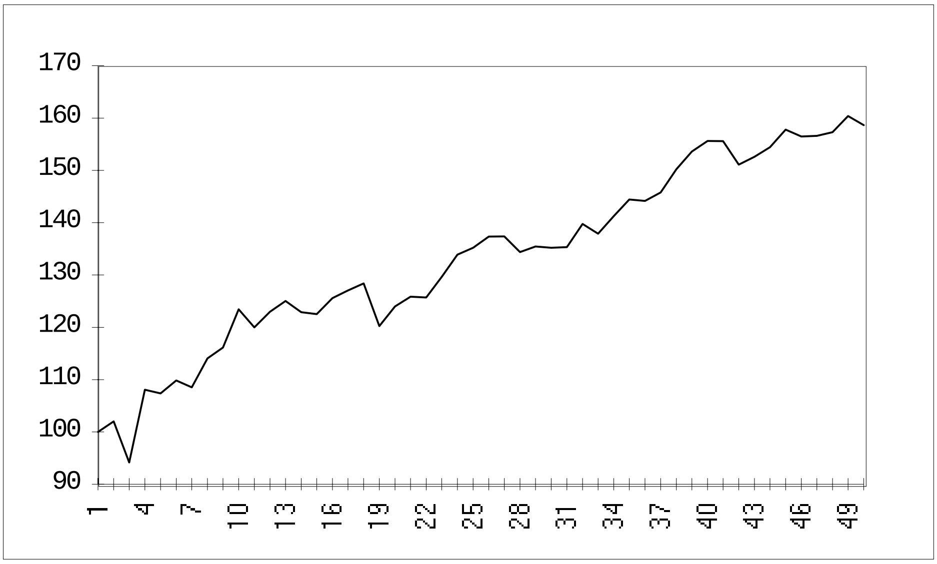
$$\hat{a}_t = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \cdots$$

High values of the smoothing constant discount past information very quickly but filter poorly erratic oscilations, low values of the smoothing constant have the opposite effect.

3 - "LOCAL" LINEAR SERIES - HOLT'S METHOD

Assume now that you want to forecast a series like the one in Graph 2:

Graph 2



The series is trended and this trend can be aproximated, at least locally, by a linear function, i. e. the trend is changing but linearity seems adequate for adjacent observations.

To forecast future values of a linear trend we need two estimates: an estimate of the ***level*** of trend ($\hat{a}_T$) and an estimate of the ***slope*** of trend ($\hat{b}_T$) at the end of the period. Then the forecast function is:

$$\hat{Y}_{T+h} = \hat{a}_T + \hat{b}_T.h \text{ , for } h = 1,2,...$$

How to get $\hat{a}_T$ and $\hat{b}_T$? Overall least squares is not adequate for a changing lienear trend. "Moving" least squares or weighted least squares are better. Pragmatic solutions use estimators based on moving averages, simple moving averages and double moving averages, simple exponential smoothing and double exponential smoothing, etc. Holt's solution uses the two following recursive estimators:

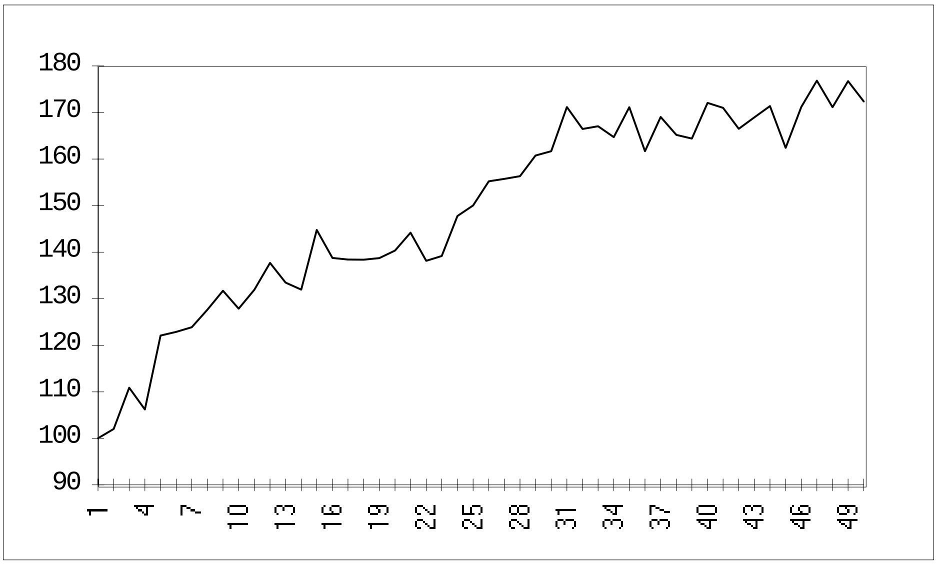
$$\hat{a}_t = \alpha y_t + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \text{ with } 0 < \alpha < 1$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1}, \text{ with } 0 < \beta < 1 \text{ .}$$

4 - DAMPED TRENDS

Empirical aplications shown that Holt's linear forecasts tend to overstate medium and long horizons. Gardner and McKenzie (1985) suggest the introduction of a parameter that moderates extrapolation as time horizon grows. This was found adequate for series like the one in Graph 3.

Graph 3



To damp extrapolations, the linear forecast function was modified by introducing a parameter ϕ, taking values between *0* and *1.* So the forecast function becames:

$$\hat{Y}_{T+h} = \hat{a}_T + \hat{b}_T \sum_{j=1}^{h} \phi^j \text{ , for } h = 1,2,...$$

In accordance, recursive estimators are now:

$$\hat{a}_t = \alpha y_t + (1-\alpha)(\hat{a}_{t-1} + \phi\hat{b}_{t-1}), \text{ with } 0 < \alpha < 1$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\phi\hat{b}_{t-1}, \text{ with } 0 < \beta < 1$$

The method can be viewed as a generalization of exponential smoothing forecasting patterns:

1) With $\phi = 0$ the method collapses in simple exponential smoothing.

2) With $0 < \phi < 1$ the trend is damping and $\lim \hat{Y}_{T+h} = \hat{a}_T + \hat{b}_T (\frac{\phi}{1-\phi})$. For low and moderate values of ϕ the eventual forecast function becomes practically constant a few steps ahead.

3) With $\phi = 1$ Holt's method emerges.

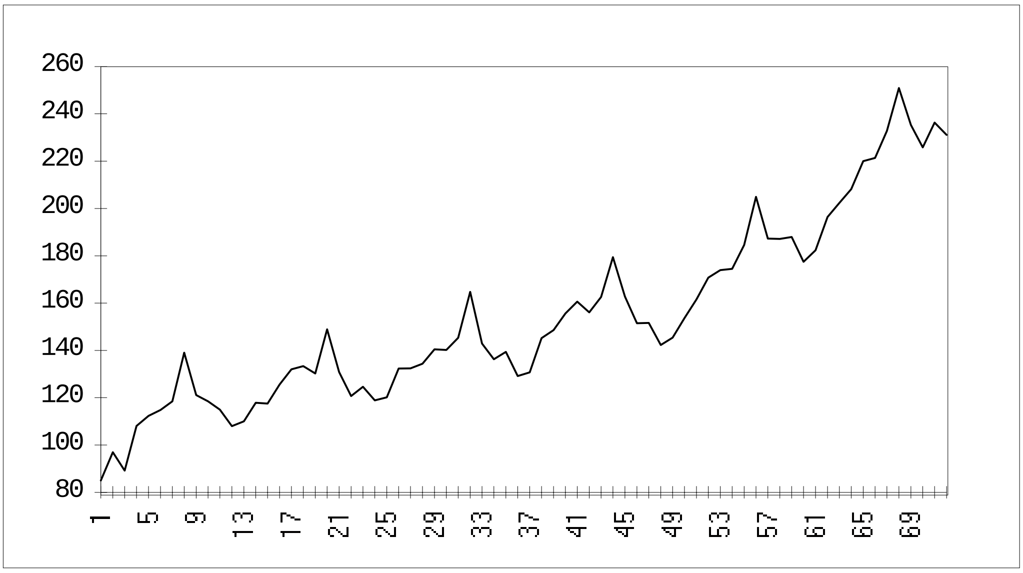
4) With $\phi > 1$ the forecast's function slope is recieving exponentially growing weight and forecasts become explosive.

5 - SEASONAL SERIES: HOLT-WINTERS METHODS

For series that exhibit seasonality, possibly non deterministic, the same principle of recursive estimators was extended to the estimation of seasonal factors (Winters, 1960). As seasonal patterns can be judge additive or multiplicative, two sets of recursions are available.

If additive seasonality is considered adequate (see Graph 4), the set of estimators and the forecast function are:

Graph 4



<u>Additive seasonality</u>

Estimators:

$$\hat{a}_t = \alpha(y_t - \hat{s}_{t-L}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \text{ with } 0 < \alpha < 1$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1}, \text{ with } 0 < \beta < 1$$

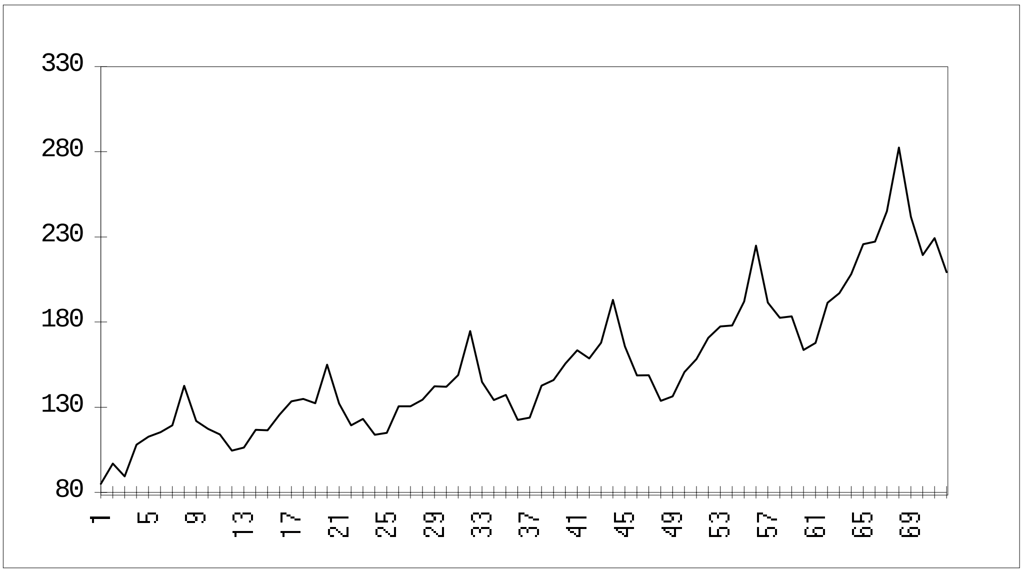
$$\hat{s}_t = \gamma(y_t - \hat{a}_t) + (1-\gamma)\hat{s}_{t-L}, \text{ with } 0 < \beta < 1$$

Forecast function: $\hat{Y}_{T+h} = \hat{a}_T + \hat{b}_T.h + \hat{s}_{T+h-kL}$,

for $h = 1, 2, \ldots$, and $k = 1$ for $h \le L$, $k = 2$ for $L < h \le 2L$ ...

For series exhibiting multiplicative seasonality (see Graph 5) the set of estimators and the forecast function are:

Graph 5



<u>Multiplicative seasonality</u>

Estimators:

$$\hat{a}_t = \alpha(y_t / \hat{s}_{t-L}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \text{ with } 0 < \alpha < 1$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1}, \text{ with } 0 < \beta < 1$$

$$\hat{s}_t = \gamma(y_t / \hat{a}_t) + (1-\gamma)\hat{s}_{t-L}, \text{ with } 0 < \beta < 1$$

Forecast function: $\hat{Y}_{T+h} = (\hat{a}_T + \hat{b}_T . h).\hat{s}_{T+h-kL}$,

for $h = 1, 2, ...$, and $k = 1$ for $h \leq L$, $k = 2$ for $L < h \leq 2L$ ...

If the trend is more locally constant than locally linear the equation of the slope estimator can be dropped. Damped trends can also be considered by introducing the ϕ parameter as above.

6 - PRACTICAL ISSUES

6.1 - IDENTIFICATION

The above presentation of exponential smoothing methods shows that exponential smoothing is more than a collection of methods. Exponential smoothing should be viewed as a forecasting methodology, with different methods for different time series patterns. So, selection of the appropriate exponential smoothing procedure is important in order to achieve more accuracy, a criticism that is often made to forecasting competions.

The simplest way to select the most appropriate method is by visual inspection, i.e., graphic analysis of the patterns of the series. A more objective selection procedure, that can be used in automatic forecasting, was proposed by Gardner and McKenzie (1988). We reproduce their identification rules in Table 1.

Table 1
Gardner and McKenzie identification rules

Case	Minimum variance series	Procedure selected
A	Original	Constant level
B	First difference	Damped trend
C	Second difference	Linear or exponential trend
D	First seasonal diference	Constant level, seasonal
E	First difference of D	Damped trend, seasonal
F	Second difference of D	Linear or exponential, seasonal

Tashman and Kurk (1996) report results on the use of three selection protocols (Gardner-McKenzie's variance analysis, a visual protocol and a selection procedure using Schwarz BIC).

6.2 - CHOOSING SMOOTHING PARAMETERS

When computational facilities were not as developped as they are now, work with exponential smoothing methods, tended to guesstimate the smoothing parameters. Values of the smoothing parameters between 0.1 and 0.3 were suggested. The most usual approach, now, is the estimation by minimization of some function of one-step-ahead forecast errors (usually the sum of squared errors). This approach has shown that the suggested range for guesstimates is often violated.

6.3 - STARTING VALUES

Any recursive estimator needs starting values. To start the simple exponential smoothing, we need an estimate of $\hat{a}_0$; for starting the Holt's method, we need $\hat{a}_0$ and $\hat{b}_0$; and for seasonal methods, we need $\hat{a}_0$, $\hat{b}_0$ and $\hat{s}_i$ for $i = 0, -1, \ldots, -L+1$.

Many heuristic solutions are in use. The current idea concerning starting values seems to be that, no matter how we start, starting values are no longer important after some recursions for moderate long series.

This line of reasoning can be suported by what happens with the simple exponential smoothing. In fact, after t periods, the estimator of the level of the series will be computed as:

$$\hat{a}_t = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1} y_1 + (1-\alpha)^t \hat{a}_0$$

and we see that the effect of the starting value is vanishing as t grows.

For seasonal factors, this will not happen so quickly, unless we have a very long series or use a very high smoothing parameter. And, more generally, even for simpler methods, empirical applications show that parameter optimization (and forecasts to some extent) is very sensitive to different starting values (Chatfield and Yar, 1988).

7 - OPTIMALITY OF EXPONENTIAL SMOOTHING METHODS

Optimality of exponential smoothing methods for some stochastic processes and ARIMA models was soon recognized (Muth, 1960; Nerlove and Wage, 1964; Theil and Wage, 1964; Roberts, 1982). We present some results:

Simple exponential smoothing

Note that $\hat{a}_t = \alpha y_t + (1-\alpha)\hat{a}_{t-1}$ can also be written (the error correction form) as (1) $\hat{a}_t = \hat{a}_{t-1} + \alpha e_t$ where $e_t = y_t - \hat{a}_{t-1}$ is the last one step ahead forecast error.

From (1) we have $\hat{a}_t = \frac{\alpha e_t}{(1-B)}$ and from the error definition $y_t = \hat{a}_{t-1} + e_t$. So, by substitution, $y_t = \frac{\alpha e_{t-1}}{(1-B)} + e_t$, or

$$(1-B)y_t = e_t - (1-\alpha)e_{t-1}.$$

Provided the errors are white noise, y_t is an *ARIMA(0,1,1)* with parameter $\theta = 1-\alpha$. As $0 < \alpha < 1$, also $0 < \theta < 1$ and simple exponential smoothing is equivalent to adjusting a restricted *ARIMA(0,1,1)*. We can also see that simple exponential smoothing is stable (invertible) for $0 < \alpha < 2$ and not only for $0 < \alpha < 1$.

Simple exponential smoothing is also optimal for the structural model:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t \end{aligned}$$

Holt's method

Holt's estimators in the error correction form are:

$\hat{a}_t = \hat{a}_{t-1} + \hat{b}_{t-1} + \alpha e_t$, and

$\hat{b}_t = \hat{b}_{t-1} + \alpha\beta e_t$, where $e_t = y_t - (\hat{a}_{t-1} + \hat{b}_{t-1})$.

So we have $\hat{b}_t = \frac{\alpha\beta e_t}{(1-B)}$ and $\hat{a}_t = \frac{\alpha\beta e_{t-1}}{(1-B)^2} + \frac{\alpha e_t}{(1-B)}$. Substitution of $\hat{a}_{t-1}$ and $\hat{b}_{t-1}$ in $y_t = \hat{a}_{t-1} + \hat{b}_{t-1} + e_t$ gives

$$(1-B)^2 y_t = e_t - (2-\alpha-\alpha\beta)e_{t-1} - (\alpha-1)e_{t-2}.$$

Provided errors are white noise, y_t is following an *ARIMA(0,2,2)* with parameters $\theta_1 = 2-\alpha-\alpha\beta$ and $\theta_2 = \alpha - 1$. Applying Holt's method is equivalent to adjusting a restricted *ARIMA(0,2,2)*.

Holt's method is also optimal for the structural model:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\
\beta_t &= \beta_{t-1} + \xi_t
\end{aligned}
$$

Damped exponential smoothing

As a similar analysis shows, applying damped exponential smoothing with $0 < \phi < 1$ is equivalent to adjusting an *ARIMA(1,1,2)*, with $\theta_1 = 1 + \phi - \alpha - \phi\alpha\beta$ and $\theta_2 = \phi(\alpha - 1)$.

Holt-Winters

Additive Holt-Winters is equivalent to adjusting an $ARIMA(0,1,L+1)(0,1,0)_L$ (see Roberts, 1982). Additive Holt-Winters ARIMA equivalent is similar to Harvey's basic structural model ARIMA equivalent (Harvey, 1989).

Multiplicative Holt-Winters is nonlinear and does not have any ARIMA equivalent.

8 - PREDICTION INTERVALS

When we do not define clearly the model, it is hard to construct prediction intervals for more then one step ahead (as we don't kown how to obtain forecast variances for longer horizons). Intervals derived by assuming deterministic trends and seasonals seem "unhelpful and potentially misleading" (Chatfield and Yar, 1988). So, except for the multiplicative Holt-Winters, intervals derived from ARIMA model's for wich exponential smoothing methods are optimal are an option. For multiplicative Holt-Winters see Chatfield and Yar (1991).

An empirical approach to prediction intervals, based on computation of error's variances for different lead times during model-fitting, was proposed by Gardner (1988). He also proposed the use of Chebyshev's inequality in defining the confidence level. This seems to give rise to too much wide prediction intervals, which would be not very helpfull.

9 - ROBUSTNESS

Although easy to use and understand and despite its optimal limited range of application compared to Arima modelling exponential smoothing methods compared very well with model based methods in forecasting competitions (see Makridakis *et al.*, 1982). Recently, Chen (1997) investigated robustness properties of four time series forecasting methods for seasonal series (wich include Holt-Winters, ARIMA, structural components, and regression on polinomial time trends and dummies with stationary ARMA errors). He concluded that "the strongly robust properties of the Holt-Winters method are due to the reasonable structure of the predictor with parsimonious parametrization. This has flexibility to handle a wide class of time series having stochastic/deterministic linear trend and seasonal components. This flexibility and parsimony are even more important in the case where the time series has structural changes".

References:

Brown, R.G., 1959, *Statistical Forecasting for Inventory Control*, New York, McGraw-Hill.

Chatfield, C., Yar, M., 1988, Holt-Winters forecasting: some practical issues, *The Statistician*, 37, 129-140.

Chatfield, C., Yar, M., 1991, Prediction intervals for multiplicative Holt-Winters, *International Journal of Forecasting*, 7, 31-37.

Chen, C.,1997, Robustness properties of some ferecasting methods for seasonal time series: a Monte Carlo study, *International Journal of Forecasting*, 13, 269-282.

Gardner, Jr. E.S.,1985, Exponential smoothing: The state of the art, *Journal of Forecasting*, 4, 1-28.

Gardner, Jr. E.S.,1988, A simple method of computing prediction intervals for time series forecasts, *Management Science*, 34, 541-546.

Gardner, Jr. E.S., McKenzie, E., 1985, Forecasting trends in time series, *Management Science*, 31, 1237-1246.

Gardner, Jr. E.S., McKenzie, E., 1988, Model identification in exponential smoothing, *Journal of the Operational Research Society*, 3, 863-867.

Harvey, A.C., 1989, *Forecasting, structural time series models and the Kalman filter*, Cambridge.

Holt, C.C. *et al.* , 1960, *Planning Production, Inventories and Wok Force*, Englewood Cliffs, NJ:Prentice-Hall, Chap. 14.

Makridakis,S. *et al.,* 1982, The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting*, 1, 111-153.

Muth, J.F.,1960, Optimal properties of exponentially weighted forecasts, *Journal of the American Statistical Association*, 55, 299-306.

Nerlove, M. and Wage, S., 1964, Some observations on adaptive forecasting, *Management Science*, 28, *Management Science*, 10, 207-224.

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

Tashman, L.J., and Kurk, J.M., 1996, The use of protocols to select exponential smoothing procedures: a reconsideration of forecasting competitions, *International Journal of Forecasting*, 12, 235-253.

Theil,H. and Wage,S., 1964, Some observations on adaptive forecasting, *Management Science*, 10, 198-206..

Winters, P.R., 1960, Forecasting sales by exponentially weighted moving averages, *Management Science*, 6 ,324-342.