ISSN 1440-771X
ISBN 0 7326 1055 9

MONASH UNIVERSITY

AUSTRALIA

Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method

Anne B. Koehler, Ralph D. Snyder, J. Keith Ord.

Working Paper 1/99
January 1999

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method

by

Anne B.Koehler
Department of Decision Sciences
and Management Information Systems
Miami University
Oxford, OH 45056
Tel: (513) 529-4826
E-mail: KOEHLEAB@MUOHIO.EDU

Ralph D. Snyder
Department of Econometrics and Business Statistics
Monash University
Clayton, Victoria 3168
Australia
E-mail: RALPH.SNYDER@BUSECO.MONASH.EDU.AU

J. Keith Ord
Department of Management Science
and Information Systems
Pennsylvania State University
303 Beam Business Administration Building
University Park, PA 16802
E-mail: JKO@PSUVM.PSU.EDU

Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method

Abstract

A new class of models for data showing trend and multiplicative seasonality is presented. The models allow the forecast error variance to depend on the trend and/or the seasonality. It can be shown that each of these models has the same updating equations and forecast functions as the multiplicative Holt-Winters method, regardless of whether the error variation in the model is constant or not. While the point forecasts from the different models are identical, the prediction intervals will, of course, depend on the structure of the error variance and so it is essential to be able to choose the most appropriate form of model. Two methods for making this choice are presented and examined by simulation.

1. Introduction

Among the most widely known and used forecasting techniques for seasonal time series are the methods proposed by Winters (1960), one for additive seasonality (additive Holt-Winters method) and one for multiplicative seasonality (multiplicative Holt-Winters method). Of these two methods, the one for multiplicative seasonality has been implemented more often in computer forecasting software. The seasonality is multiplicative if the effect of the season increases with an increase in the level of the time series. It is additive if the seasonal effect does not depend on the current level of the time series and can simply be added or subtracted from a forecast that depends only on level and trend. While the multiplicative Holt-Winters method provided reasonable point forecasts, there was no way to justify the choice of prediction intervals because no underlying statistical model on which to base the variance of the forecast error had been found. By an underlying model for a forecasting method we mean the following: when the initial conditions and the parameters are the same, the conditional expected values for future values computed from the model agree with the predictions from the forecasting method. An ARIMA model that underlies the additive Holt-Winters method was identified by McKenzie (1976). However, there is no ARIMA model for the multiplicative Holt-Winters method (see Abraham and Ledolter, 1986).

Ord, Koehler, and Snyder (1997) present a general class of nonlinear models among which is a state space model (with a single source of error) for the multiplicative Holt-Winters model. The previous work that led to the discovery of this class of models included Ord and Koehler's proposed state space model (with multiple sources of error) for the multiplicative Holt-Winters method (1990) and Snyder's state space models (with a single source of error) for linear smoothing methods (1985, 1988). Independently, others were working on the same problem of how to provide a statistical basis for computing prediction intervals instead of just the point forecasts. Assuming that an underlying model exists but without specifying it, Chatfield and Yar (1991) show that the variance of the forecast error is likely to depend on the season or on both the season and the level. They do not, however, suggest how to choose between these possible forms for the variances of the forecast errors. In a working paper, Archibald (1994) studies three models for the multiplicative Holt-Winters method. He investigates procedures for choosing among his three models, and one of the procedures is the maximum likelihood method of this paper. The comparisons in his simulation studies are made without a reference to a 'true model.'

The general framework for nonlinear models in Ord, Koehler, and Snyder (1997) can be used to show that there are, in fact, many models that underlie the multiplicative Holt-Winters method. In this paper, we will first present the original model from the 1997 paper in two forms, one of which clearly shows the connection to exponential smoothing, and examine this particular model carefully. Then we will present the general form for the models and choose three more specific models to compare with the original model. We derive the variances of the forecast errors for the general model and investigate the coverage of future values by the prediction intervals that correspond to the four selected models. Next we use both a maximum likelihood method and a correlation method to identify models for simulated time series. Finally, we discuss the implications of these results for practitioners.

2. A Model for the Multiplicative Holt-Winters Method

We will begin this section by presenting the state space model (OKS) for the multiplicative Holt-Winters method in essentially the same form as Example 4 in Ord, Koehler, and Snyder (1997). Then we will show an equivalent exponential smoothing form for the transition equations that will reveal an obvious relationship to the Holt-Winters smoothing equations. Still focussing on this one model, we will show that the minimum mean square forecast function and updating equations for the model agree with the forecast function and updating equations for the multiplicative Holt-Winters method. We will conclude the section by presenting and discussing how to estimate the smoothing parameters.

Model 1: OKS Model

Observation Equation

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m} + (\ell_{t-1} + b_{t-1})c_{t-m}\varepsilon_t \tag{2.1}$$

Transition Equations of State Equations

level $\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1(\ell_{t-1} + b_{t-1})\varepsilon_t$ (2.1a)

trend $b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})\varepsilon_t$ (2.1b)

season $c_t = c_{t-m} + \alpha_3 c_{t-m}\varepsilon_t$ (2.1c)

The observation equation shows the relationship between the time series y_t and the underlying state variables at time t. The state equations show the transition of the state variables, which can be given the following interpretations at time period t:

ℓ_t is the underlying level for the time series,

b_t is the underlying growth rate, and

c_t is the seasonal factor.

The number of seasons in a given year is m. It is assumed that the error terms, ε_t, are NID$(0,\sigma^2)$ and that they are independent of past values of the time series and past values of the state variables. The α_1, α_2, and α_3 are parameters that correspond to the smoothing constants in the Holt-Winters Method. Necessary conditions for stability of the model (which may not be sufficient, Ord, Koehler, and Snyder (1997)) are $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_3 > 0$, $2\alpha_1 + \alpha_2 < 4$, and $\alpha_3 < 1$.

We now introduce the notation e_t for the forecast error (with perfect information) as opposed to the relative error ε_t. We define e_t as follows:

$$e_t = y_t - (\ell_{t-1} + b_{t-1})c_{t-m} \tag{2.2}$$

In contrast,

$$\varepsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})c_{t-m}}{(\ell_{t-1} + b_{t-1})c_{t-m}} \tag{2.3}$$

$$= \frac{e_t}{(\ell_{t-1} + b_{t-1})c_{t-m}}$$

In order to obtain the exponential smoothing form for the transition equations, we substitute the right-hand side of Equation 2.3 into the transition equations 2.1a, 2.1b, and 2.1c. The result is

the following form for these equations.

Exponential Smoothing Form of State Equations:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1 e_t / c_{t-m} \tag{2.4a}$$

$$b_t = \quad b_{t-1} + \alpha_2 e_t / c_{t-m} \tag{2.4b}$$

$$c_t = \quad c_{t-m} + \alpha_3 e_t / (\ell_{t-1} + b_{t-1}) \tag{2.4c}$$

The minimum forecast mean square error for h periods in the future from time n is always the expected value of y_{n+h} conditioned on the information up through time n. If we let the information at time n be

$$I_n = \{y_1,...,y_n, \ell_0, b_0, c_{j-m} \, (j = 1,...,m)\} \tag{2.5}$$

then

$$E(y_{n+h}|I_n) = (\ell_n + hb_n)c_{n+h-m} \tag{2.6}$$

In practice, we would know $y_1, y_2,...,y_n$ and need to find estimates for ℓ_0, b_0, and c_{j-m} (j = 1,2,...,m). The form for our forecasts at time n would be

Forecast Function

$$\hat{y}_{n+h} = (\hat{\ell}_n + h\,\hat{b}_n)\hat{c}_{n+h-jm} \quad \text{where} \quad j = [h/m] + 1 \tag{2.7}$$

Updating Equations for t = 1,2,...,n

$$\hat{\ell}_t = \hat{\ell}_{t-1} + \hat{b}_{t-1} + \hat{\alpha}_1 \hat{e}_t / \hat{c}_{t-m} \tag{2.7a}$$

$$\hat{b}_t = \hat{b}_{t+1} + \hat{\alpha}_2 \hat{e}_t / \hat{c}_{t-m} \tag{2.7b}$$

$$\hat{c}_t = \hat{c}_{t-m} + \hat{\alpha}_3 \hat{e}_t / (\hat{\ell}_{t-1} + \hat{b}_{t-1}) \tag{2.7c}$$

where $\hat{e}_t = y_t - \hat{y}_{t-1}$ and $\hat{y}_{t-1} = (\hat{\ell}_{t-1} + \hat{b}_{t-1})\hat{c}_{t-m}$

This forecast function with the updating equations is the error correction form of the multiplicative Holt-Winters method (Gardner, 1985) except for a minor difference in Equation 2.7c. In the usual smoothing method $\hat{e}_n$ (i.e. the residual) is divided by $\hat{\ell}_n + \hat{b}_n$ in Equation (2.7c).

We turn now to estimation of initial states and parameters. If we specialize the concentrated likelihood function from Ord, Koehler, and Snyder (1997) to model (2.1), we find it to be

$$L(\alpha, x_0, s|y_1^n) = (2\pi s^2)^{-n/2} \prod_{t=1}^{n} |(\ell_{t-1} + b_{t-1})c_{t-m}|^{-1} \tag{2.8}$$

where $y_1^n = (y_1, y_2, \ldots, y_n)$

$$s^2 = (\sum_{t=1}^{n} \varepsilon_t^2)/n$$

$$\alpha = (\alpha_1, \alpha_2, \alpha_3)$$

$$x_0 = (\ell_0, b_0, c_0, c_{-1}, \ldots, c_{-m+1})$$

We now examine the log likelihood function to compare maximum likelihood to the usual criterion for choosing α. The usual criterion is to minimize $\sum_{t=1}^{n} \hat{e}_n^2$ (i.e. sum of the squared residuals or estimated forecast errors). The negative of 2 times the log likelihood function (ignoring constants) is

$$F_1(\alpha, x_0, s)|y_1^n) = n \log s^2 + 2 \sum_{t=1}^{n} \log|(\ell_{t-1} + b_{t-1})c_{t-m}| \tag{2.9}$$

Minimizing (2.9) is equivalent to maximizing the likelihood function (2.8). If we ignore the second term in (2.9), we would recommend minimizing $\sum_{t=1}^{n} \hat{\varepsilon}_t^2$ (i.e. sum of the squared estimated relative errors) rather than the usual criterion. We have found that it works well to estimate the initial values of the states prior to maximizing the likelihood. The estimates that we use are found by the procedures originally proposed by Winters (1960) for his initial estimates.

If one uses the maximum likelihood estimation procedure, the estimate for σ^2 will be

$$\hat{s}^2 = (\sum_{t=1}^{n} \hat{\varepsilon}_t^2)/n \quad \text{where} \quad \hat{\varepsilon}_t = \frac{\hat{e}_t}{(\hat{\ell}_{t-1} - \hat{b}_{t-1})\hat{c}_{t-m}} .$$

3. General Model and Three More Specific Models for the Multiplicative Holt-Winters Model

The model in the previous section is a special case of a general model.

General Model:

Observation Equation

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m} + (\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma}_{t-m}\varepsilon_t \tag{3.1}$$

Transition Equations

(3.1c)

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1(\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma-1}_{t-m}\varepsilon_t \tag{3.1a}$$

$$b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma-1}_{t-m}\varepsilon_t \tag{3.1b}$$

$$c_t = c_{t-1} + \alpha_3(\ell_{t-1} + b_{t-1})^{\beta-1} c^{\gamma}_{t-m}\varepsilon_t \tag{3.1c}$$

where the notation is the same as that of Model 1 of the previous section with the addition of the parameters β and γ where $0 \le \beta \le 1$ and $0 \le \gamma \le 1$. Model 1 is obviously the case where $\beta = 1$ and $\gamma = 1$. Equation (2.3) can now be generalized to

$$\varepsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})c_{t-m}}{(\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma}_{t-m}} = \frac{e_t}{(\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma}_{t-m}} \tag{3.2}$$

Using Equation (3.2) to substitute for ε_t in Equation (3.1a), (3.1b), and (3.1c), will produce exactly the same exponential smoothing form of the state equations in (2.4). Furthermore, the forecast function and updating equations for this model are the same as those of Equations in (2.7). Finally, Function (2.9), which is minimized to choose $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ becomes

$$F_g(\alpha, x_0, s|y^n_1) = n \log s^2 + 2 \sum_{t=1}^{n} \log| (\ell_{t-1} + b_{t-1})^{\beta} c^{\gamma}_{t-m}| \tag{3.3}$$

In addition to Model 1, there are three other models that we believe are the most important special cases of the general model (3.1). All four models have the same transition or state equations when they are put in the exponential smoothing form in (2.4). Hence, we only need to specify the observation equation. For each of these three additional models, we now present both the observation equation and the form of Function (3.3) that corresponds to the appropriate likelihood function.

Model 2: $\beta = 1$ and $\gamma = 0$

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m} + (\ell_{t-1} + b_{t-1})\varepsilon_t \tag{3.4}$$

$$F_2(\alpha, x_0, s|y_1^n) = n \log s^2 + 2 \sum_{t=1}^{n} \log|(\ell_{t-1} + b_{t-1})| \tag{3.5}$$

Model 3: $\beta = 0$ and $\gamma = 1$

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m} + c_{t-m}\varepsilon_t \tag{3.6}$$

$$F_3(\alpha, x_0, s|y_1^n) = n \log s^2 + 2 \sum_{t=1}^{n} \log|c_{t-m}| \tag{3.7}$$

Model 4: $\beta = 0$ and $\gamma = 0$

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m} + \varepsilon_t \tag{3.8}$$

$$F_4(\alpha, x_0, s|y_1^n) = n \log s^2 \tag{3.9}$$

Only for Model 4 is $s^2 = \sum_{t=1}^{n} e_t^2/n$. In the other three cases (i.e. Models 1, 2, and 3) e_t # ε_t. Only for Model 4 will the maximum likelihood estimates for α_1, α_2, and α_3 agree with those from minimizing $\hat{s}^2$, when ε_t has a normal distribution. Hence, only for Model 4 would the usual criterion of minimizing $\sum_{t=1}^{n} \hat{e}_t^2$ be recommended.

4. Prediction Intervals

We believe that the most important differences between Models 1, 2, 3, and 4 are the variances of the forecast error and their effect on the prediction intervals for future values. If we have the same estimates for the initial values of the level, trend, and seasonal factors and for α_1, α_2, and α_3, the point forecasts for y_{n+h} at time n are the same for all models (see the forecasting function (2.7)). Prediction intervals for h periods ahead at time n given $y_1, y_2, ..., y_n$ have the form

$$\hat{y}_{n+h} \pm k \, Var(y_{n+h} - \hat{y}_{n+h}).$$

The k is determined by the appropriate forecasting distribution.

In order to study the effect of the model on the prediction intervals through the variances of the forecast errors, we assume that we have perfect information at time n. In particular, we assume that we know the values of α_1, α_2, α_3, σ and the initial values for the level trend, and m seasonal factors. This information is summarized by I_n in (2.5). Thus the forecast from all models at time n, given I_n, will be

$$\hat{y}_{n+h} = (\ell_n + hb_n)c_{n+h-m} \quad \text{for} \quad 1 \le h \le m$$

The variance of the forecast error for h-periods ahead at time period n is derived in Appendix 1. We now specialize the results of the Appendix 1 to the four selected models.

Model 1: $\beta = 1$ and $\gamma = 1$

h = 1 $\quad Var(y_{n+1} - (\ell_n + b_n)c_{n+1-m}|I_n) = (\ell_n + b_n)^2 c^2_{n+1-m}\sigma^2$ (4.1a)

$2 < h \le m$ $\quad Var(y_{n+h} - (\ell_n + hb_n)c_{n+h-m}|I_n)$ (4.1b)

$$= [(\alpha_1 + (h - 1)\alpha_2)^2(\ell_n + b_n)^2 + ... + (\alpha_1 + \alpha_2)^2(\ell_n + (h - 1)b_n)^2$$

$$+ (\ell_n + hb_n)^2]c^2_{n+h-m}\sigma^2$$

where $\quad \sigma^2 = Var(e_t/(\ell_{t-1} + b_{t-1})c_{t-m})$ (4.1c)

Model 2: $\beta = 1$ and $\gamma = 0$

h = 1 $\quad Var(y_{n+1} - (\ell_n + b_n)c_{n+1-m}|I_n) = (\ell_n + b_n)^2\sigma^2$ (4.2a)

$2 < h \le m$ $\quad Var(y_{n+h} - (\ell_n + hb_n)c_{n+h-m}|I_n)$ (4.2b)

$$= [(\alpha_1 + (h - 1)\alpha_2)^2(\ell_n + b_n)^2 / c^2_{n+1-m} + ... + (\alpha_1 + \alpha_2)^2(\ell_n + (h - 1)b_n)^2/c^2_{n+(h-1)-m}$$

$$+ (\ell_n + hb_n)/c^2_{n+h-m}]c^2_{n+h-m}\sigma^2$$

where $\sigma^2 = Var(e_t/(\ell_{t-1} + b_{t-1}))$ (4.2c)

Model 3: $\beta = 0$ and $\gamma = 1$

$h = 1$ $Var(y_{n+1} -(\ell_n + b_n)c_{n+1-m}|I_n) = c^2_{n+1-m}\sigma^2$ (4.3a)

$2 < h \leq m$ $Var(y_{n+h} - (\ell_n + hb_n)c_{n+h-m}|I_n)$ (4.3b)

$$= [(\alpha_1 + (h - 1)\alpha_2)^2 + ... + (\alpha_1 + \alpha_2)^2 + 1]c^2_{n+h-m}\sigma^2$$

where $\sigma^2 = Var(e_t/c_{t-m})$ (4.3c)

Model 4: $\beta = 0$ and $\gamma = 0$

$h = 1$ $Var(y_{n+1} - (\ell_n + b_n)c_{n+1-m}|I_n) = \sigma^2$ (4.4a)

$2 < h \leq m$ $Var(y_{n+h} - (\ell_n + hb_n)c_{n+h-m}|I_n)$ (4.4b)

$$= [(\alpha_1 + (h - 1)\alpha_2)^2/c^2_{n+1-m} + ... + (\alpha_1 + \alpha_2)^2/c^2_{n+(h-1)-m} + 1/c^2_{n+h-m}]c^2_{n+h-m}\sigma^2$$

where $\sigma^2 = Var(e_t)$ (4.4c)

A simulation was run to investigate the impact of choosing the wrong model on prediction intervals. In particular, it is important to see if the coverage (i.e., chances of including future observations) is reduced if the wrong model is used to compute the prediction intervals. The steps in the simulation were as follows:

A. Generate a time series $y_1,...,y_n, y_{n+1},...,y_{n+12}$ with Model j, j = 1, 2, 3, 4, by using (3.1) specialized with appropriate β and γ. The initial values for the seasonal factors are $c_{j'-m} = 1 + A\sin(2\pi j'/m)$, $j' = 1,2,...,m$ where A is the seasonal amplitude. All parameters and initial values are specified in Table 1. A value of 5(.05) indicates that $\sigma = 0.05$ for Models 1 and 2 and $\sigma = 5$ for Models 3 and 4.

B. Using the first n values from each generated time series, compute the variance of the h-period-ahead forecast error, h = 1,2,...,12, that corresponds to Model i, i = 1,2,3,4. (See Equations (4.1), (4.2), (4.3), and (4.4).) A value for σ^2 is computed using

$$s^2 = \frac{\sum_{i=1}^{n} \varepsilon_t^2}{n} \quad \text{where} \quad \varepsilon_t = \frac{e_t}{(\ell_{t-1} + b_{t-1})^\beta c_{t-m}^\gamma}.$$

Using this variance and the value of 1.96 for k, compute the upper and lower limits for the prediction interval that corresponds to Model i, i = 1, 2, 3, 4. (See Section 4.)

C. For each time series from Model j, i = 1, 2, 3, 4, count the number of times y_{n+h}, h = 1,2,...,12, is contained in the prediction interval computed with Model i, i = 1, 2, 3, 4.

D. Replicate 1,000 times.

E. Compute the percentage of times the future values y_{n+h}, h = 1,2,...,12, from Model j, j = 1, 2, 3, 4, are contained in the prediction intervals that are calculated from Model i, i = 1,2,3,4.

The results of this simulation are contained in the upper third of Table 1. They show that Model 1 and 2 provide the same coverage for each other as do Models 3 and 4. However, using Models 3 or 4 to compute prediction intervals for time series generated by Models 1 or 2 provides coverage that is reduced from the 95% level to 63 to 65%. On the other hand, using Models 1 or 2 to predict time series generated by Models 3 or 4 raises the level of coverage by 3%. We believe it is the upward trend that plays the major role in this difference. The middle third of Table 1 is the same simulation except that $\alpha_2 = 0$ (i.e., the trend or growth rate is not changing). Since the level is changing, the time series has an upward trend with a drift. The bottom third of the table removes the trend from the models (i.e., growth rate $b_t = 0$). By comparing all parts of the table, one can see that the choice of the model will be important if there is upward trend.

----- Table 1 -----

The formulas in Equations 4.1, 4.2, 4.3, and 4.4 indicate that intervals for Model 1 will be wider at higher levels and higher seasonal peaks of the time series and narrower at lower levels and seasonal troughs. Hence, in the bottom third of the table, some of these differences are being averaged out. Thus choosing the correct model could be even more important than our one simulation already shows.

5. Choosing a Model

From the last section, we know that the coverage rate of future values by prediction intervals is affected by the choice of the model. Hence, we now examine a maximum likelihood method (ML) and a correlation method (CORR) for choosing among Models 1, 2, 3, and 4. The maximum likelihood method chooses the model which has the largest value for the likelihood function. The correlation method chooses the model by comparing the correlations of the absolute value of the residuals $|\hat{e}_t|$, with estimates for the level and seasonal factors in the term with ε_t. This correlation method follows from a suggestion in Neter, Kutner, Nachtscheim, and Wasserman (1996) for determining nonconstant variance in regression models. For a given time series y_t, t = 1,2,...,n, the methods will now be described more precisely.

Maximum Likelihood Method (ML)

i) Using Model i, i = 1,2,3,4, find $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ that minimize Function (3.3), $F_i(\alpha, x_0, s|y_1^n)$, for the corresponding values of β and γ, where

$$\hat{s}^2 = \Big[\sum_{t=1}^{n} \hat{e}_t^2/((\hat{\ell}_{t-1} + \hat{b}_{t-1})^{\beta}\hat{c}_{t-m}^{\gamma})\Big]/n$$

The specific functions are given in (2.9), (3.5), (3.7), and (3.9).

ii) Choose Model i for which F_i is the smallest of F_1, F_2, F_3, and F_4.

Correlation Method

i) Residuals $\hat{e}_t$ are found by estimating α_1, α_2, and α_3 with the usual criterion (i..e.

minimize $\sum_{i=1}^{n} \hat{e}_t^2 / n$ as in the case for Model 4).

i=1

ii) Compute the following correlations

cor 1: $|\hat{e}_t|$ vs $(\hat{\ell}_{t-1} + \hat{b}_{t-1})\hat{c}_{t-m}$

cor 2: $|\hat{e}_t|$ vs $\hat{\ell}_{t-1} + \hat{b}_{t-1}$

cor 3: $|\hat{e}_t|$ vs $\hat{c}_{t-m}$

iii) Model i, i = 1,2,3 is selected if

a) cor i is highest among cor 1, cor 2, and cor 3

and b) cor i $\geq$ Critical Correlation Value

iv) Model 4 is selected if

cor i < Critical Correlation Value for all i = 1,2,3

We tested these two methods with simulations for the values of the parameter, number of seasons, seasonal amplitude, and number of time periods shown in Tables 2, 3, and 4. The steps for these simulations are the following:

Step 1 (Generate time series)

For Model j, j = 1,2,3,4, generate a time series y_1, y_2,...,y_n by using (3.1) specialized with the appropriate β and γ. The initial values are $\ell_0 = 100$, $b_0 = 2$, and $c_{j'-m} = 1 + A\sin(2\pi j'/m)$, $j' = 1,2,...,m$, where A is the seasonal amplitude. All other values are specified in Tables 2, 3, and 4.

Step 2 (Estimates of initial values)

For each time series, compute estimate, $\hat{x}_0 = (\hat{\ell}_0, \hat{b}_0, \hat{c}_{j'-m,} j' = 1,2,\ldots,m,)$, for the vector of initial state values by using Winters original method (1960).

Step 3 (ML Choice)

For each time series generated from Model j, use the Maximum Likelihood Method to select the most likely Model i, except use $\hat{x}_0$ from Step 1 instead of the optimal value for x_0.

Step 4. (Estimates with usual criterion)

For each time series generated by model j, use Step 3 to estimate the $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ and $\hat{\ell}_t$, $\hat{b}_t$, and $\hat{c}_t$, t = 1,2,...,n, that minimize F_4 for Model 4. In other words, use the usual criterion to select the estimates for the smoothing parameters and states at each time t.

Step 5. (CORR choice)

For each time series generated from Model j in step 1 and estimates from Step 4, use the Correlation Method with a Critical Correlation Value of 0.20 to pick Model i. This critical value was determined by trying to balance Type I and Type II errors for no correlation. Experimentation indicates that 0.20 is a reasonable choice.

----- Tables 2, 3, 4 -----

The results are shown in Tables 2, 3, and 4. There are 1,000 replications for Model j, j = 1,2,3,4, and each specification of the number of time periods, the number of seasons, the seasonal amplitude, σ, α_1, α_2, and α_3. The tables show that time series generated by Models 1 or 2 are identified correctly by both ML and CORR most of the time and in the remaining cases usually confused with each other. In the previous section, we demonstrated that these two models will have similar coverage by each other′s prediction intervals. Model 3 is usually misidentified as Model 1, and similarly Model 4 is misidentified as Model 2. An increase in seasonal amplitude reduces this problem for Models 3 and 4. On the positive side, for Models 3 and 4, they are rarely chosen if they are not the correct models. From the coverage study in the last section, we believe that if we have upward trend that the prediction intervals that are computed using Model 3 or 4 will be narrower than prediction intervals computed from Models 1 or 2. Thus, a misidentification of Model 3 or 4 as Model 1 or 2 for a time series with an upward trend will result in intervals that are wider than necessary while no trend time series will have comparable coverage by all the models.

An increase in seasonal amplitude improves the identification of all models by both methods for montly time series. Increases in the variance or in the smoothing constants did not show clear patterns of effects on the model identification.

An interesting observation is that Correlation Method performed reasonably well in comparison to the Maximum Likelihood Method for choosing models. The Correlation Method seems to do better than the Maximum Likelihood for Model 4. However, there are two cautions to be given here. The estimates for the parameters and the states were made using Model 4, that is, the usual criterion of minimizing $\sum_{i=1}^{n} \hat{e}_t^2$ was employed. Thus Model 4 probably has a slight edge in being chosen. In addition, the Critical Correlation Value is somewhat arbitrary. If it is increased, more time series will be assigned to Model 4 and vice-versa.

Other simulations with the growth rate b = 0.5 produced qualitatively similar results. Selections were not quite as clear-cut as is to be expected with a smaller slope.

6. Summary and Conclusions

There are many models, not just one, that underlie the multiplicative Holt-Winters method. Choosing the correct model is important for the prediction intervals to have the potential to reach the desired coverage of future values. While the formulas for the variances of the forecast errors already say that it will matter if the wrong interval is used, the simulation for the prediction intervals demonstrated how significant this difference in the percentage of coverage can be. Both the maximum likelihood and the correlation methods provide help in identifying the correct models, especially for Models 1 and 2. One can feel confident that if Model 3 or 4 is selected,then it is likely to be the correct model. When there is upward trend and Model 1 is selected instead of Model 3, and Model 2 instead of Model 4, the coverage is better than is expected. However, the intervals are wider than necessary for the desired percentage of coverage. Thus, it is only the desire to obtain small intervals that makes it important to identify Model 3 and 4 correctly in this case. The formulas for the variances of the forecast errors say that the reverse will be true for downward trend, that is, using Model 1 for Model 3, or Model 2 for Model 4 will produce intervals that are too narrow with reduced coverage. Thus, one will not be as confident about using the results of the identification methods with downward trend. When there is no trend, the overall coverage is the same for all the models, but the intervals will be different for the different models at the high and low points of the model. Hence, there is a need to find even better ways to identify the correct models.

There is some information about these models for real data. Archibald (1994) studied models that are equivalent to Models 1, 3, and 4. He found that the maximum likelihood estimation of the smoothing parameters by these three models produced point forecasts that did not differ in accuracy. He also found that for the data from the 1982 Makridakis competition the maximum likelihood method tended to favor Model 1 with Model 1 being chosen about 50% of the time and Model 3 less than 20% of the time. Building on the paper by Chatfield and Yar (1991), Robert G. Goodrich has implemented a correlation procedure in his commercial forecasting software to choose between one-period-ahead variances that correspond to Model 1 and 3 (i.e., 4.1a and 4.3a). He has reported that all of his time series point to Model 1.

On the basis of evidence from this paper and these other sources, we make the following recommendations if your time series shows variation that increases when the level of your data increases. If you intend to implement one model without any identification procedure, select Model 1. This means you should minimize Equation (2.9) or minimize the related (but not equivalent) sum of the relative squared error and compute the variance for your prediction intervals with Equations (4.1a) and (4.1b). If you plan to identify the appropriate model from among Models 1, 2, 3, and 4, you can use either the maximum likelihood method or the correlation method for this identification. Then you would use the estimation criterion and variance formula that are appropriate for the chosen model, but take heed of information in the first paragraph of this summary about when this effort will be beneficial.

Appendix 1

Derivation of the Variance of the Forecast Error
for Model when β and γ are restricted to 0 or 1.

Model: See Equations 3.1, 3.1a, 3.1b, and 3.1c.

σ^2 changes with the model as follows:

$$\sigma^2 = \mathrm{Var}(\varepsilon_t) = \mathrm{Var}\left[\frac{y_t - (\ell_{t-1} + b_{t-1})c_{t-m}}{(\ell_{t-1} + b_{t-1})^{\beta} c_{t-m}^{\gamma}} \right] \qquad \text{A.1}$$

This derivation will assume perfect information for α_1, α_2, α_3, β, γ, ℓ_0, b_0, c_0, $c,\ldots,c_{1-m}$ and σ. Hence, at time n we would also know ℓ_n, b_n, and c_{n+h-m}, $h = 1,2,\ldots,m$. In addition, the values of β and γ are restricted to 0 or 1. When $\beta = 1$, the derived variance will be an approximation because we will ignore products of the error terms, ε_t. This approximation will be reasonable because when $\beta = 1$, the values for ε_t will be proportions of the current level of y_t (see Equation A.1), and the products of numbers much less than 1 will be trivially small.

For all models, the minimum mean square error forecast for y_{n+h} at time n is

$E(y_{n+h}|I_n) = (\ell_n + hb_n)c_{n+h-m}$ where $I_n = (y_1,y_2,\ldots,y_n, \ell_0, b_0, c_{j-m}\ (j = 1,\ldots,m))$.

Consider the forecast horizon h = 1 at time t.

$$\mathrm{Var}(y_{n+1} - (\ell_n + b_n)c_{n+1-m}|I_n) = (\ell_n + b_n)^{2\beta} c_{n+1-m}^{2\beta} \sigma^2$$

Consider the forecast horizon h = 2 at time n, and use Equations (3.1), (3.1a), and (3.1b).

$$\mathrm{Var}[y_{n+2} - (\ell_n + 2b_n)c_{n+2-m}|I_n)]$$

$$= \mathrm{Var}[(\ell_{n+1} + b_{n+1})c_{n+2-m} + (\ell_{n+1} + b_{n+1})^{\beta} c_{n+2-m}^{\gamma}\varepsilon_{n+2} - (\ell_n + 2b_n)c_{n+2-m}|I_n]$$

$$= c_{n+2-m}^2\ \mathrm{Var}[\alpha_1(\ell_n + b_n)^{\beta} c_{n+1-m}^{\gamma-1}\varepsilon_{n+1} + \alpha_2(\ell_n + b_n)^{\beta} c_{n+1-m}^{\gamma-1}\varepsilon_{n+1}$$

$$+ (\ell_n + b_n + \alpha_1(\ell_n + b_n)^{\beta} c_{n+1-m}^{\gamma-1}\varepsilon_{n+1} + b_n + \alpha_2(\ell_n + b_n)^{\beta} c_{n+1-m}^{\gamma-1}\varepsilon_{n+1})^{\beta} c_{n+2-m}^{\gamma-1}\varepsilon_{n+2}|I_n]$$

$$.c^2_{n+2-m}\mathrm{Var}[\alpha_1(\ell_n + b_n)^{\beta}c^{\gamma-1}_{n+1-m}\varepsilon_{n+1} + \alpha_2(\ell_n + b_n)^{\beta}c^{\gamma-1}_{n+1-m}\varepsilon_{n+1} + (\ell_n + 2b_n)^{\beta}c^{\gamma-1}_{n+2-m}\varepsilon_{n+2}|I_n]$$

$$= c^2_{n+2-m}[(\alpha_1 + \alpha_2)^2(\ell_n + b_n)^{2\beta}c^{2(\gamma-1)}_{n+1-m} +(\ell_n + 2b_n)^{2\beta}c^{2(\gamma-1)}_{n+2-m}]\sigma^2$$

This same line of reasoning can be followed for $2 \le h \le m$ to show

$\mathrm{Var}[y_{n+h} - (\ell_n + hb_n)c_{n+h-m}|I_n]$

$$c^2_{n+h-m}[(\alpha_1 + (h-1)\alpha_2)^2(\ell_n + b_n)^{2\beta}c^{2(\gamma-1)}_{n+1-m} + (\alpha_1 + (h-2)\alpha_2)^2(\ell_n + 2b_n)^{2\beta}c^{2(\gamma-1)}_{n+2-m}$$

$$+ \ldots + (\alpha_1 + \alpha_2)^2(\ell_n + (h-1)b_n)^{2\beta}c^{2(\gamma-1)}_{n+h-1-m} + \ell_n + hb_n)^{2\beta}c^{2(\gamma-1)}_{n+h-m}]\sigma^2$$

Comment: Interested readers may wish to look at Ord, Koehler, and Snyder (1997) to see simulation methods that account for the estimation error in prediction intervals.

References

Abraham, B. and J. Ledolter, 'Forecast Functions Implied by Autoregressive Integrated Moving Average and other Related Forecast Procedures', *International Statistical Review*, 54 (1986), 51-66.

Archibald, Blyth C., 'Parameter Space of the Holt-Winters' Model,' *International Journal of Forecasting*, 6 (1990), 199-209.

Archibald, Blyth C., 'Winters Model: Three Versions, Diagnostic Checks, and Forecast Performances,' Working Paper WP-94-4, School of Business Administration, Dalhousie University, Halifax, Canada, (1994).

Bowerman, B. L. and R. T. O'Connell, *Forecasting and Time Series: An Applied Approach*, Duxbury, Belmont, CA, 1993.

Chatfield, C. and M. Yar, 'Prediction Intervals for Multiplicative Holt-Winters,' *International Journal of Forecasting*, 7 (1991), 31-37.

Gardner, Jr., E. S., 'Exponential Smoothing: the State of the Art,' *Journal of Forecasting*, 4 (1985), 1-28.

McKenzie, E., 'A Comparison of Some Standard Seasonal Forecasting Sysems', *The Statistician*, 25 (1976), 3-14.

Neter, J., M. H. Kutner, C. J. Nachtscheim, and W. Wasserman, *Applied Regression Models*, Irwin, Chicago, IL, 1996.

Ord, J. K. and A. B. Koehler, 'A Structural Model for the Multiplicative Holt-Winters Method', *Proceedings of the Decision Sciences Institute*, 1990, 555-557.

Ord, J. K., A. B. Koehler, and R. D. Snyder, 'Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models', *Journal of the American Statistical Association*, 92 (1997), 1621-1629.

Snyder, R. D., 'Recursive Estimation of Dynamic Linear Statistical Models', *Journal of the Royal Statistical Society, Series B*, 47 (1985), 272-276.

Snyder, R. D., 'Statistical Foundations of Exponential Smoothing', Working Paper 5/88, Department of Econometrics, University of Monash, 1988.

Winters, P. R., 'Forecasting Sales by Exponentially Weighted Moving Averages,' *Management Science*, 6 (1960), 324-342.

Table 1

Percentage of Future Values from True Model
within Prediction Limits of Four Potential Models

True Model Specifications

Initial Level	Initial Trend	m	n	σ	Seas Ampl.(A)	α_1	α_2	α_3
100	2	12	72	5(.05)	0.30	0.20	0.05	0.10

Potential Model	True Model 1	2	3	4
1	95	96	98	99
2	94	96	98	99
3	63	65	95	96
4	63	65	94	96

Initial Level	Initial Trend	m	n	σ	Seas Ampl.(A)	α_1	α_2	α_3
100	2	12	72	5(.05)	0.30	0.20	0.00	0.10

1	95	94	100	100
2	94	95	100	100
3	56	56	95	94
4	56	55	94	95

Initial Level	Initial Trend	m	n	σ	Seas Ampl.(A)	α_1	α_2	α_3
100	0	12	72	5(.05)	0.30	0.20	0.00	0.10

1	95	94	94	94
2	94	95	93	95
3	95	94	95	94
4	94	95	94	95

Table 2

Effect of Seasonal Amplitude and Variance
on Model Identification

True Model Specifications*

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
12	72	1(.01)	0.30	0.20	0.05	0.10

Percentage of Time True Model Identified

Identified Model	Maximum Likelihood Method: True Model 1	2	3	4	Correlation Method: True Model 1	2	3	4
1	82.3	17.2	64.9	18.6	73.3	14.7	54.7	15.8
2	16.8	82.4	9.7	55.7	23.6	80.3	6.1	31.1
3	0.7	0.1	21.6	3.9	1.0	0.0	23.2	2.5
4	0.2	0.3	3.8	21.8	2.1	5.0	16.0	50.6

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	5(.05)	0.30	--	--	--

Identified Model	Maximum Likelihood Method: True Model 1	2	3	4	Correlation Method: True Model 1	2	3	4
1	77.9	12.8	60.0	15.7	66.6	14.1	49.5	15.2
2	16.2	80.8	7.7	51.9	23.0	69.3	5.1	28.1
3	5.1	2.0	27.4	5.3	4.9	0.5	25.8	2.5
4	0.8	4.4	4.9	27.1	5.5	16.1	19.6	54.2

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	1(.01)	0.60	--	--	--

Identified Model	Maximum Likelihood Method: True Model 1	2	3	4	Correlation Method: True Model 1	2	3	4
1	98.6	5.8	52.1	5.9	91.9	13.7	35.4	15.2
2	0.2	88.2	0.1	28.5	0.3	71.9	0.0	9.3
3	1.2	0.2	47.4	4.1	7.8	0.5	64.4	12.7
4	0.0	5.8	0.4	61.5	0.0	13.9	0.2	62.8

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	5(.05)	0.60	--	--	--

Identified Model	Maximum Likelihood Method: True Model 1	2	3	4	Correlation Method: True Model 1	2	3	4
1	93.6	3.4	68.1	3.5	92.7	10.2	60.5	11.2
2	1.1	89.5	0.9	47.4	1.6	72.9	0.2	25.2
3	5.3	0.3	30.7	1.2	5.2	0.4	38.8	4.7
4	0.0	6.8	0.3	47.9	0.5	16.5	0.5	58.9

*All true models have initial level = 100 and initial trend = 2. Critical value for correlation criteria is 0.20. For each specification of true model i, i = 1,2,3,4, there are 1,000 replications.

Table 3

Effect of Smoothing Parameters and Variance
on Model Identification

True Model Specifications

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
12	72	1(.01)	0.30	0.30	0.05	0.10

Percentage of Time True Model Identified

Identified Model	Maximum Likelihood Method: True Model 1	2	3	4	Correlation Method: True Model 1	2	3	4
1	82.9	18.0	68.0	19.2	74.0	15.5	57.4	17.9
2	16.5	81.5	9.1	58.6	23.2	79.4	6.4	31.8
3	0.5	0.2	19.1	4.1	0.9	0.0	21.6	3.1
4	0.1	0.3	3.8	18.1	1.9	5.1	14.6	47.2

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	5(.05)	--	0.30	0.05	0.10

Identified Model	ML 1	2	3	4	Corr. 1	2	3	4
1	79.7	15.0	62.0	17.8	67.8	15.7	54.6	16.9
2	15.4	76.7	8.7	50.1	21.5	66.9	5.9	28.8
3	4.0	2.3	24.9	6.1	4.6	0.4	23.3	2.3
4	0.9	6.0	4.4	26.0	6.1	17.0	16.2	52.0

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	1(.01)	--	0.20	0.10	0.10

Identified Model	ML 1	2	3	4	Corr. 1	2	3	4
1	84.3	24.9	52.1	20.5	73.8	19.2	43.3	15.0
2	11.4	70.1	7.1	38.6	17.3	65.3	4.5	21.0
3	3.2	1.9	35.0	11.8	4.2	1.3	33.8	6.0
4	1.1	3.1	5.8	29.1	4.7	14.2	18.4	58.0

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
--	--	5(.05)	--	0.20	0.10	0.10

Identified Model	ML 1	2	3	4	Corr. 1	2	3	4
1	79.8	23.1	43.6	15.3	71.9	20.2	37.5	13.8
2	12.3	70.5	4.6	31.3	17.5	65.9	2.2	19.3
3	7.2	1.0	43.8	14.5	6.5	0.4	36.1	6.2
4	0.7	5.4	8.0	38.9	4.1	13.5	24.2	60.7

Table 3 (continued)

--	--		1(.01)		--		0.20		0.05		0.20
1	82.3	16.8	63.3	18.1		73.2	14.7	53.5	15.4		
2	16.6	82.6	9.4	56.3		23.9	79.9	5.0	30.5		
3	0.7	0.0	22.8	4.3		0.6	0.0	23.7	2.9		
4	0.4	0.6	4.5	21.3		2.3	5.4	17.8	51.2		
--	--		5(.05)		--		0.20		0.05		0.20
1	78.3	13.3	59.3	16.2		66.4	13.6	48.6	13.5		
2	15.5	79.4	7.9	50.5		23.0	68.7	5.0	28.3		
3	5.3	1.5	27.7	5.3		4.8	0.6	25.9	2.3		
4	0.9	5.8	5.1	28.0		5.8	17.1	20.5	55.9		

Table 4

Effect of Seasonal Periods (m), Length of Series (n), and Variance
on Model Identification

True Model Specifications

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
12	48	1(.01)	0.30	0.20	0.05	0.10

Percentage of Time True Model Identified

Identified Model	Maximum Likelihood Method True Model 1	2	3	4	Correlation Method True Model 1	2	3	4
1	75.6	22.3	62.1	26.2	64.7	17.7	56.6	24.1
2	18.7	73.3	13.3	50.0	25.0	70.9	12.0	39.5
3	4.2	0.9	19.2	6.9	4.1	5.0	18.4	4.0
4	1.5	3.5	5.4	16.9	6.2	10.9	13.0	32.4

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
12	48	5(.05)	--	--	--	--

Identified Model	ML 1	2	3	4	Correlation 1	2	3	4
1	70.7	17.9	60.1	22.8	57.2	16.1	55.6	22.5
2	16.5	69.6	11.7	48.2	23.4	63.0	10.8	36.4
3	9.3	3.6	22.7	8.6	8.7	1.3	20.4	3.5
4	3.5	8.9	5.5	20.4	10.7	19.6	13.2	37.6

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
4	48	1(.01)	--	--	--	--

Identified Model	ML 1	2	3	4	Correlation 1	2	3	4
1	72.7	20.0	27.6	9.0	59.9	17.7	23.4	7.8
2	15.5	69.4	6.3	23.3	15.2	53.9	2.1	10.4
3	9.2	1.9	52.8	16.8	10.7	12.0	42.8	8.3
4	2.6	8.7	13.3	50.9	14.2	27.2	31.7	73.5

m	n	σ	seas. ampl.(A)	α_1	α_2	α_3
4	48	5(.05)	--	--	--	--

Identified Model	ML 1	2	3	4	Correlation 1	2	3	4
1	73.3	18.0	37.8	10.9	59.8	14.5	34.9	8.5
2	18.8	74.3	10.3	35.5	22.5	62.5	6.8	17.3
3	6.0	1.3	41.5	11.5	5.9	0.4	30.2	4.9
4	1.9	6.4	10.4	42.1	11.8	22.6	28.1	69.3

Table 4 (continued)

4	36		1(.01)		--		--		--	--
1	67.0	22.1	37.1	13.9		48.9	15.4	28.3	9.9	
2	16.6	61.4	9.9	32.0		17.9	42.8	4.1	13.4	
3	12.3	4.0	40.5	16.4		14.8	2.3	38.5	11.5	
4	4.1	12.5	12.5	37.7		18.4	39.5	29.1	65.2	
4	36		5(.05)		--		--		--	--
1	69.4	21.8	46.0	17.1		49.8	13.7	36.9	11.7	
2	18.9	66.7	16.8	47.9		26.4	56.2	11.2	26.6	
3	8.8	2.6	27.4	9.7		9.1	1.5	24.5	6.3	
4	2.9	8.9	9.8	25.3		14.7	28.6	27.4	55.4	