

Exponential Smoothing Models¹

Professor Thomas B. Fomby
Department of Economics
Southern Methodist University
Dallas, TX
June 2008

Introduction

The formulation of **exponential smoothing** forecasting methods arose in the 1950's from the original work of Brown (1959, 1962) and Holt (1960) who were working on creating forecasting models for inventory control systems. **One of the basic ideas** of smoothing models is to construct forecasts of future values as weighted averages of past observations with the more recent observations carrying more weight in determining forecasts than observations in the more distant past. By forming forecasts based on weighted averages we are using a “smoothing” method. The adjective “exponential” derives from the fact that some of the exponential smoothing models not only have weights that diminish with time but they do so in an exponential way, as in $\lambda_j = \lambda^j$ where $-1 < \lambda < 1$ and $j = 1, 2, \dots$ represents the specific period in the past.

At least **three major points** can be raised about exponential smoothing models:

- As a methodology, **exponential smoothing methods suffer from not having an objective statistical identification and diagnostic system for evaluating the “goodness” of competing exponential smoothing models**. For example, the smoothing parameters of the smoothing models are determined by fit and are **not** based on any statistical criteria like tests of hypotheses concerning parameters or tests for white noise in the errors produced by the model. In this sense, exponential smoothing models are **ad hoc** models, statistically speaking. Of course, if one continues to monitor the forecasting performance of a given exponential smoothing model, and, if the model's forecasts become more and more inaccurate over time, then one has, in a sense, an ex post evaluation method for picking and choosing between competing exponential smoothing models. The

¹ The discussion here draws heavily from the SAS manual SAS/ETS Software: Time Series Forecasting System, Version 6, First Edition, Cary, NC. SAS Institute Inc., 1995, 225-235.

only problem is that this approach comes with a cost. Bad forecasting for a certain amount of time while learning can be expensive when, for example, dealing with inventories that run into the millions of dollars. But instead of pursuing this ex post **monitoring approach**, one can attempt to make a good choice of exponential smoother **before hand** by using **out-of-sample forecasting experiments**. In this approach, the forecaster reserves some of the available data for a “horse race” between competing exponential smoothing methods. To carry these horse races out, one divides the data into two parts: the **in-sample data** set (say 60% of the first part of the available time series data) and with the remaining latter part of the time series assigned to the **out-of-sample data** set. Then one “runs” the competing exponential smoothing methods through the out-of-sample data while forecasting h-steps ahead each time (we assume h is the forecast horizon of interest) while updating the “smoothing” parameter(s) as one moves through the out-of-sample data. In the process of generating these h-step-ahead forecasts for the competing methods we can compare the competing forecasts with the actual values that we withheld as we generated our forecasts and then use the standard forecasting accuracy measures like MSE, MAE, RMSE, PMAE, etc. to choose the best (most accurate) exponential smoothing forecasting method, as indicated by the out-of-sample forecasting experiment, for further use (subject to monitoring of course).

- Most exponential smoothing methods, as we will see below, can be shown to be **special cases of the class of Box-Jenkins models**. For this reason, Box-Jenkins forecasters have been critical of using exponential smoothing models for forecasting. They usually say, “Why be satisfied with a special case of a Box-Jenkins model when we can fit any Box-Jenkins model we want to the data? Moreover, we can use the various diagnostic tests that are available to choose a good Box-Jenkins model without being ad hoc in our model choice.” This, of course, is a very persuasive argument and is the reason why many forecasters use standard Box-Jenkins computer software for doing their forecasting.
- **A counterargument** to the traditional Box-Jenkins criticism of exponential smoothing methods is that, once basic decisions like the presence or absence of

trend in the time series and the presence or absence of seasonality is determined, pretty accurate forecasting can be obtained even to the point of being “almost” as accurate as the “fully flexible” and non-ad hoc Box-Jenkins models. The fact that carefully chosen exponential smoothing models do almost as well as Box-Jenkins models has been documented in two large-scale empirical studies by Makridakis and Hibon (1979) and Makridakis et. al. (1982). So if the Box-Jenkins models are more time consuming to build yet only yield marginal gains in forecasting accuracy relative to less time-consuming **well informed** choices of exponential smoothing models, we have an interesting trade off between time (which is money) and accuracy (which, of course, is also money). This trade-off falls in the favor of exponential smoothing models sometimes when, for example, one is working with 1500 product lines to forecast and has only a limited time to build forecasting models for them. In what we will argue below, an **informed choice** consists of knowing whether the data in hand has a trend in it or not and seasonality in it or not. Once these basic decisions are made (and if they are correct!), then pretty accurate forecasts are likely via the appropriately chosen exponential smoothing method.

So for those individuals that have heavy time constraints, it is probably worth investigating the possibility of using exponential smoothing forecasting models for the “less important” time series and saving the most important time series for modeling and forecasting with Box-Jenkins methods.

In the next section we are going to describe the basic **Additive Smoothing model** and how its smoothing parameters and prediction intervals are determined. In the following section we will discuss four sub-cases of this model that are available in SAS Proc Forecast.

II. The Additive Smoothing Model

Basic Notation

Given a time series, y_t , $t=1,2,\dots,T$, the **Additive Smoothing model** is of the form

$$y_t = \mu_t + \beta_t t + S_{t,p} + a_t \quad , \quad t = 1, 2, \dots, T, \quad (1)$$

where μ_t represents the **time-varying mean (level) term**, β_t represents the **time-varying slope term** (called “trend” term by exponential smoothers), $S_{t,p}$ denotes the **time-varying seasonal term** for the p seasons ($p = 1, 2, \dots, P$) in the year, and a_t is a **white noise error term**. For smoothing models without trend, $\beta_t = 0$ for all t , and for smoothing models without seasonal effects, $S_{t,p} = 0$ for all p . (This model is, of course, the Basic Structural Model of A.C. Harvey (1989) that we previously discussed when using Proc UCM in SAS.) At each time period t , each of the above time-varying components $(\mu_t, \beta_t, S_{t,p})$ are estimated by means of so-called **smoothing equations**. By way of notation, let L_t be the **smoothed level** that estimates μ_t , T_t be the **smoothed slope** that estimates β_t , and the **smoothed seasonal factor** S_t estimates the seasonal effect $S_{t,p}$ at time t .

The smoothing process begins with initial estimates of the smoothing states L_0, T_0 , and $S_{0,1}, S_{0,2}, \dots, S_{0,P}$. These smoothing states are subsequently updated for each observation by means of **smoothing equations**. These smoothing equations determine how the smoothing states change over time. **Knowledge of the smoothing state at time $t-1$ and the knowledge of the observation y_t at time t , along with smoothing weights uniquely determine the smoothing states at time t .** The smoothing equations for each smoothing model will be discussed subsequently.

Initialization of the Smoothing States

Given the time series y_t , $t = 1, 2, \dots, T$, the smoothing process begins by computing the smoothing states for time $t = 1$. However, this computation requires initial estimates of the smoothing states at time $t = 0$, L_0, T_0 , and $S_{0,1}, S_{0,2}, \dots, S_{0,P}$, even though there exist no observations on y_t on or before $t = 0$. These initial values of the smoothing states are

obtained by the method of **backcasting**. Consider the “Seasonal Factors Sum to Zero” parametrization of the Deterministic Trend/Seasonal model with AR(0) errors:

$$y_t = \alpha + \beta t + \gamma_1 D_{t,1} + \gamma_2 D_{t,2} + \cdots + \gamma_p D_{t,p} + a_t \quad (2)$$

where the restriction $\gamma_1 + \gamma_2 + \cdots + \gamma_p = 0$ is imposed. Now the choices for **the initial smoothing states for backcasting** are obtained by **re-ordering** the original y data from $t = T$ to $t = 1$ (in reverse order) and then running least squares on equation (2) and obtaining the least squares estimates $\hat{\alpha}, \hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_p$. Then the **initial slope state for backcasting purposes** is estimated as $T_{0,b} = \hat{\beta}$, the **initial seasonal states for backcasting** are estimated as $S_{0,1,b} = \hat{\gamma}_1$, $S_{0,2,b} = \hat{\gamma}_2, \cdots$, and $S_{0,p,b} = \hat{\gamma}_p$, while the **initial level for the backcast**, $L_{0,b}$, is simply set equal to y_T . Then, once the initial states of the backcasts are determined, the given model is smoothed in reverse order to obtain estimates (backcasts) of the initial states L_0, T_0 , and $S_{0,1}, S_{0,2}, \cdots, S_{0,p}$.

Prediction and Prediction Errors

Predictions are based on the last known smoothing state as we will see below. Denote the predictions of y made at time t for h periods ahead as \hat{y}_{t+h} and the h -step-ahead errors in prediction as $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$. The one-step-ahead prediction errors e_{t+1} are also the **model residuals** and the **sum of squares of the one-step-ahead prediction errors is the objective function** used in weight (parameter) optimization in the smoothing models.

The variance of the prediction errors are used to calculate the confidence limits of predictions made by the smoothing models. The equations for the variance of the prediction errors for each smoothing model will be listed in the following sections.

NOTE: $\text{Var}(a_t)$ is estimated by the mean square of the one-step-ahead predictions errors,

$$\hat{\text{var}}(a_t) = \sum_{t=0}^{T-1} (\hat{e}_{t+1}^2 / T). \quad (3)$$

Smoothing Weights

Depending on the particular smoothing model, the smoothing weights consist of the following:

- α is the level smoothing weight.
- γ is the trend smoothing weight. (4)
- δ is the seasonal smoothing weight.

Larger smoothing weights (less damping) permit the more recent data to have a greater influence on the predictions. Smaller smoother weights (more damping) give less weight to recent data. The above weights are restricted to “invertible” regions so as to guarantee stable predictions. As we will see, the smoothing models presented below have **ARIMA (Box-Jenkins) model equivalents**. That is, the smoothing models can be shown to be **special cases** of the more general Box-Jenkins models which we will study in detail later.

The above smoothing weights are chosen to minimize the sum of squared one-step-ahead prediction errors, $\sum_{t=0}^{T-1} e_{t+1}^2$. Usually this is done by doing a grid search across the invertible region of the smoothing weights and calculating the corresponding sum of squared errors by means of what is called the Kalman Filter. Being more specific, in the case of the Additive Smoothing Model (1), the previous states $L_t, T_t, S_t, L_{t-1}, T_{t-1}, S_{t-1} \dots$ are used to produce the predicted states $\hat{L}_{t+1}, \hat{T}_{t+1}$, and \hat{S}_{t+1} for next period. These predicted states then provide next period’s predicted value $\hat{y}_{t+1} = \hat{L}_{t+1} + \hat{T}_{t+1} + \hat{S}_{t+1}$ which in turn produces the one-step-ahead prediction error $e_{t+1} = y_{t+1} - \hat{y}_{t+1}$. Therefore, for fixed values of the smoothing weights, say α_i, γ_i , and δ_i , one can recursively generate the sum of squared one-step-ahead prediction errors $S(\alpha_i, \gamma_i, \delta_i) = \sum_{t=0}^{T-1} e_{t+1}^2$. Finally, one can do a grid search over the (α, γ, δ) invertible space and choose the smoothing weights, say α^*, γ^* , and δ^* , that minimize $S(\alpha_i, \gamma_i, \delta_i)$. Once these optimal smoothing weights are chosen, the exponential smoothing equations can be used to generate forecasts out-of-sample.

III. Four Simple Additive Exponential Smoothing Models

1. SIMPLE EXPONENTIAL SMOOTHING

(This model should be used when the time series data has no trend and no seasonality.)

The simple exponential smoothing model is given by the **model equation**

$$y_t = \mu_t + a_t . \quad (5)$$

The **smoothing equation** is

$$L_t = \alpha y_t + (1 - \alpha)L_{t-1} . \quad (6)$$

The **h-step-ahead prediction equation** is

$$\hat{y}_{t+h} = L_t , \quad h = 1, 2, \dots . \quad (7)$$

That is, you forecast y h-steps ahead by using the last available estimated (smoothed) level state, L_t .

The **ARIMA model equivalent** to the simple exponential smoothing model is the ARIMA(0,1,1) model

$$(1 - B)y_t = (1 - \theta B)a_t \quad (8)$$

where $\theta = 1 - \alpha$ and B represents the backshift operator such that $B^r x_t = x_{t-r}$ for any given time series x_t .

The **invertible region** for the simple exponential smoothing model is $0 < \alpha < 2$.

The **variance of the h-step-ahead prediction errors** is given by

$$\text{var}(e_{t+h}) = \text{var}(a_t) \left[1 + \sum_{j=1}^{h-1} \alpha^2 \right] = \text{var}(a_t) (1 + (h-1)\alpha^2) . \quad (9)$$

2. DOUBLE (BROWN) EXPONENTIAL SMOOTHING

(This model should be used when the time series data has trend but no seasonality.)

The Double exponential smoothing model is given by the **model equation**

$$y_t = \mu_t + \beta_t t + a_t . \quad (10)$$

The **smoothing equations are**

$$L_t = \alpha y_t + (1 - \alpha) L_{t-1} \quad (11)$$

$$T_t = \alpha(L_t - L_{t-1}) + (1 - \alpha) T_{t-1} . \quad (12)$$

The **h-step-ahead prediction equation is**

$$\hat{y}_{t+h} = L_t + ((h-1) + 1/\alpha) T_t , \quad h = 1, 2, \dots . \quad (13)$$

That is, you forecast y h-steps ahead by taking the last available estimated level state and multiplying the last available trend (slope), T_t , by $((h-1) + 1/\alpha)$.

The **ARIMA model equivalent** to the linear exponential smoothing model is the ARIMA(0,2,2) model

$$(1 - B)^2 y_t = (1 - \theta B)^2 a_t \quad (14)$$

where $\theta = 1 - \alpha$. The **invertible region** for the double exponential smoothing model is $0 < \alpha < 2$.

The **variance of the h-step-ahead prediction errors** is given by

$$\text{var}(e_{t+h}) = \text{var}(a_t) \left[1 + \sum_{j=1}^{h-1} (2\alpha + (j-1)\alpha^2)^2 \right] . \quad (15)$$

3. SEASONAL EXPONENTIAL SMOOTHING

(This model should be used when the time series data has no trend but seasonality.)

The seasonal exponential smoothing model is given by the **model equation**

$$y_t = \mu_t + S_{t,p} + a_t . \quad (16)$$

The **smoothing equations are**

$$L_t = \alpha(y_t - S_{t-p}) + (1 - \alpha) L_{t-1} \quad (17)$$

$$S_t = \delta(y_t - L_t) + (1 - \delta) S_{t-p} . \quad (18)$$

The **h-step-ahead prediction equation** is

$$\hat{y}_{t+h} = L_t + S_{t-P+h} \quad , \quad h = 1, 2, \dots \quad (19)$$

That is, you forecast y h-steps ahead by using taking the last available estimated level state and then add the last available smoothed seasonal factor, S_{t-P+h} , that matches the month of the forecast horizon.

The **ARIMA model equivalent** to the seasonal exponential smoothing model is the ARIMA $(0,1,P+1)(0,1,0)_p$ model

$$(1-B)(1-B^P)y_t = (1-\theta_1B - \theta_2B^P - \theta_3B^{P+1})a_t \quad (20)$$

where $\theta_1 = 1-\alpha$, $\theta_2 = 1-\delta(1-\alpha)$, and $\theta_3 = (1-\alpha)(\delta-1)$.

The **invertible region** for the seasonal exponential smoothing model is

$$\max(-P\alpha, 0) < \delta(1-\alpha) < 2-\alpha \quad (21)$$

The **variance of the h-step-ahead prediction errors** is given by

$$\text{var}(e_{t+h}) = \text{var}(a_t) \left[1 + \sum_{j=1}^{h-1} \psi_j^2 \right] \quad (22)$$

where

$$\psi_j = \begin{cases} \alpha, & j \bmod(P) \neq 0 \\ \alpha + \delta(1-\alpha), & j \bmod(P) = 0 \end{cases} \quad (23)$$

4. WINTERS METHOD – ADDITIVE VERSION WITH SEASONALITY

(This model should be used when the time series data has trend and seasonality.)

The seasonal exponential smoothing model is given by the **model equation**

$$y_t = \mu_t + \beta_t t + S_{t,p} + a_t \quad (24)$$

The **smoothing equations are**

$$L_t = \alpha(y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (25)$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \quad (26)$$

$$S_t = \delta(y_t - L_t) + (1-\delta)S_{t-p} \quad (27)$$

The **h-step-ahead prediction equation** is

$$\hat{y}_{t+h} = L_t + hT_t + S_{t-P+h} , \quad h = 1, 2, \dots \quad (28)$$

That is, you forecast y h -steps ahead by using taking the last available estimated level state and incrementing it h times using the last available trend (slope) T_t **while at the same time** adding the last available smoothed seasonal factor, S_{t-P+h} , that matches the month of the forecast horizon.

The **ARIMA model equivalent** to the seasonal exponential smoothing model is the ARIMA $(0,1,P+1)(0,1,0)_p$ model

$$(1-B)(1-B^P)y_t = (1 - \sum_{i=1}^{P+1} \theta_i B^i) a_t \quad (29)$$

where

$$\theta_j = \begin{cases} 1 - \alpha - \alpha\gamma, & j = 1 \\ -\alpha\gamma, & 2 \leq j \leq P-1 \\ 1 - \alpha\gamma - \delta(1-\alpha), & j = P \\ (1-\alpha)(\delta-1), & j = P+1 \end{cases} \quad (30)$$

The **invertible region** for the Winters method is

$$\begin{aligned} \max(-P\alpha, 0) &< \delta(1-\alpha) < 2-\alpha \\ 0 &< \alpha\gamma < [2-\alpha-\delta(1-\alpha)](1-\cos(\zeta)) . \end{aligned}$$

The **variance of the h-step-ahead prediction errors** is given by

$$\text{var}(e_{t+h}) = \text{var}(a_t) \left[1 + \sum_{j=1}^{h-1} \psi_j^2 \right] \quad (31)$$

where

$$\psi_j = \begin{cases} \alpha + j\alpha\gamma, & j \bmod(P) \neq 0 \\ \alpha + j\alpha\gamma + \delta(1-\alpha), & j \bmod(P) = 0 \end{cases} \quad (32)$$

THE KALMAN FILTER

The above smoothing models can be put into so-called “State Space Form” and then the Kalman Filter can be used to recursively solve for the one-step-ahead forecast errors of the model given its initial states. By so doing we can get the sum of squared errors of the one-step-ahead forecasts of the model for given smoothing weights. Then iteratively using the Kalman Filter the optimal smoothing weights can be determined by finding those weights that **minimize** the sum of squared errors of the one-step-ahead forecasts. For more discussion on the Kalman Filter and its updating equations one can refer to Hamilton (1994), Chapter 13.

IV. An Application and Some Comments

Appropriate Choice of Exponential Smoother By Identifying the Salient Characteristics of the Data

In some forecasting situations a forecaster must forecast many times series (think about 1500 product lines, for example) in a very short period of time. It is in this kind of situation then the above smoothing models can be very useful. Choosing an appropriate smoothing model is largely dependent on recognizing whether there is or is not a trend in the data and if seasonality is or is not present in the data. Recapping, the appropriate smoothing model can be chosen as in the below table where Y = Yes and N = No.

<u>Model to Choose</u>	<u>Trend Present</u>	<u>Seasonality Present</u>
Simple Exponential Smoothing	N	N
Double (Brown) Exponential Smoothing	Y	N
Seasonal Exponential Smoothing	N	Y
Winters Method – Additive Version	Y	Y

PROC FORECAST in SAS produces the forecasts implied by the above smoothing models. All you have to do is choose the right forecast option. Consider the following table.

Model

SAS PROC FORECAST OPTIONS

Simple Exponential Smoothing
(Trend = N, Season = N)

Method = EXPO Trend = 1

Double (Brown) Exponential Smoothing
(Trend = Y, Season = N)

Method = EXPO Trend = 2

Seasonal Exponential Smoothing
(Trend = N, Season = Y)

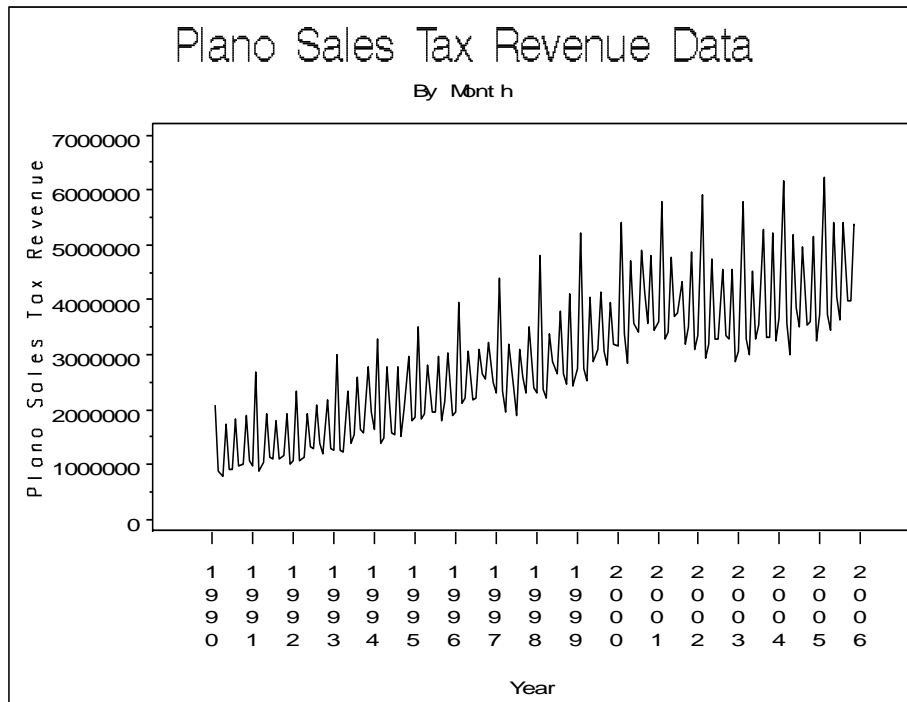
Method = Addwinters Trend = 1
Season = 12 (or Season = 4 if quarterly)

Winters Method – Additive Version
(Trend = Y, Season = Y)

Method = Addwinters Trend = 2
Season = 12 (or Season = 4 if quarterly)

An Application: The Plano Sales Tax Revenue Data

Recall the Plano Sales Tax Revenue data that we have previously analyzed. It is plotted in the graph below:



In terms of uncovering the nature of the Plano Sales Tax Revenue data we report below F-tests for trend, curvature of trend, seasonality, and a joint test for trend **and** seasonality using the autocorrelated error DTDS model

$$y_t = \alpha + \beta t + \theta t^2 + \gamma_1 D_{t1} + \gamma_2 D_{t2} + \gamma_3 D_{t3} + \dots + \gamma_{12} D_{t,12} + \varepsilon_t \quad (33)$$

with

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_r \varepsilon_{t-r} + a_t. \quad (34)$$

The relevant null hypotheses of these tests are

$$H_0 : \beta = \theta = 0 \quad (\text{No trend}) \quad (35)$$

$$H_0 : \theta = 0 \quad (\text{No curvature in trend}) \quad (36)$$

$$H_0 : \gamma_2 = \gamma_3 = \dots = \gamma_{12} = 0 \quad (\text{No seasonality}) \quad (37)$$

$$H_0 : \beta = \theta = \gamma_2 = \gamma_3 = \dots = \gamma_{12} = 0 \quad (38)$$

The outcomes of these tests produced by Proc Autoreg in SAS are as follows:

Test Trend				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	8.2658644E12	166.37	<.0001
Denominator	174	49682814486		

Test Curvature				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	125488100508	2.53	0.1138
Denominator	174	49682814486		

Test Season				
Source	DF	Mean Square	F Value	Pr > F
Numerator	11	3.4522094E12	69.48	<.0001
Denominator	174	49682814486		

Test Trend_Season				
Source	DF	Mean Square	F Value	Pr > F
Numerator	13	4.1476364E12	83.48	<.0001
Denominator	174	49682814486		

As we can see, trend is significant, curvature is not, seasonality is significant, and jointly speaking, both trend (in this case linear trend) and seasonality are significant. These results suggest that the **Additive Winters method with trend and seasonality** (24) is the recommended smoothing model for the Plano Sales Tax Revenue data.

In the case that the Plano City manager is sitting at December of 2005 wanting to forecast the 2006 revenues of the city, the appropriate SAS code for producing the necessary 13 forecasts are as follows:

```
proc forecast data = plano method = addwinters seasons = 12
  trend=2 lead=13 outlimit out=forecast;
var rev;
id date;
run;

proc print data = forecast;
run;
```

The resulting forecasts and their 95% confidence intervals are reported below:

Forecasting Plano Sales Tax Revenue Data
Using Additive Winters Smoothing
with trend and seasonality

Obs	date	_TYPE_	_LEAD_	rev
1	NOV05	FORECAST	1	3756519.60
2	NOV05	L95	1	3249487.85
3	NOV05	U95	1	4263551.35
4	NOV05	FORECAST	2	4031666.11
5	NOV05	L95	2	3521192.35
6	NOV05	U95	2	4542139.88
7	NOV05	FORECAST	3	6443738.67
8	NOV05	L95	3	5929164.54
9	NOV05	U95	3	6958312.80
10	NOV05	FORECAST	4	3993022.40
11	NOV05	L95	4	3473643.64
12	NOV05	U95	4	4512401.16
13	NOV05	FORECAST	5	3780119.96
14	NOV05	L95	5	3255190.81
15	NOV05	U95	5	4305049.10
16	NOV05	FORECAST	6	5473080.31
17	NOV05	L95	6	4941818.28
18	NOV05	U95	6	6004342.33
19	NOV05	FORECAST	7	4258830.36

20	NOV05	L95	7	3720421.25
21	NOV05	U95	7	4797239.47
22	NOV05	FORECAST	8	4127269.01
23	NOV05	L95	8	3580872.11
24	NOV05	U95	8	4673665.90
25	NOV05	FORECAST	9	5531792.38
26	NOV05	L95	9	4976545.76
27	NOV05	U95	9	6087038.99
28	NOV05	FORECAST	10	4222187.17
29	NOV05	L95	10	3657212.88
30	NOV05	U95	10	4787161.47
31	NOV05	FORECAST	11	4216839.10
32	NOV05	L95	11	3641248.20
33	NOV05	U95	11	4792430.00
34	NOV05	FORECAST	12	5635095.12
35	NOV05	L95	12	5047992.51
36	NOV05	U95	12	6222197.72
37	NOV05	FORECAST	13	4038422.10
38	NOV05	L95	13	3438623.43
39	NOV05	U95	13	4638220.76

These forecasts are generated by recursively solving forward the smoothing equations (25) – (27) and applying the prediction equation (28). The calculation of the confidence intervals is described in the next section.

Constructing Confidence Intervals for Exponential Smoothers

Consider the task of constructing forecasting confidence intervals for the smoothing models. In the above discussion we gave some expressions for the (population) variance of the h-step-ahead forecast errors of the various smoothing models labeled as $\text{var}(e_{t+h})$. Then the $(1 - \alpha)\%$ confidence interval for predicting y_{t+h} is given by

$$\Pr[\hat{y}_{t+h} - se(\hat{y}_{t+h})z_{\alpha/2} < y_{t+h} < \hat{y}_{t+h} + se(\hat{y}_{t+h})z_{\alpha/2}] = 1 - \alpha, \quad (39)$$

where $z_{\alpha/2}$ is that value of the standard normal random variable that satisfies the requirement that $\Pr(z > z_{\alpha/2}) = \alpha/2$ and the standard error of the forecast \hat{y}_{t+h} is calculated by taking the **square root** of the **estimate** of the variance of the h-step-ahead

forecast error, $\text{var}(e_{t+h})$. But $\text{var}(e_{t+h})$ is easily estimated by replacing $\text{var}(a_t)$ with

$$\hat{\text{var}}(a_t) = \sum_{i=0}^{T-1} (\hat{e}_{t+1}^2 / T)$$
 and replacing all of the unknown smoothing weights with the

estimated weights that minimize the one-step-ahead forecast errors of the model.

Upon examination of the forecast error variances of the four exponential smoothing models previously discussed (see (9), (15), (22), and (31)) it becomes clear that these error variances are a **monotonic function of the forecast horizon h** . The further out into the future the forecast is made, the more uncertain the prediction of y_{t+h} . Moreover, the forecast confidence intervals on the previous exponential smoothers are **unbounded in the limit ($h \rightarrow \infty$)**, unlike the forecast confidence intervals of the DTDS model, which are **bounded**. Of course this result obtains because of the **stochastic treatment** of the trend and seasonality in the smoothing models. Thus, even when the point forecasts of the deterministic trend DTDS models and exponential smoothing models do not differ very much, at longer horizons the “stochastic trends” of the exponential smoothing models will admit more uncertainty vis-à-vis their forecast confidence intervals than the DTDS models which have bounded forecast confidence intervals. The choice between “stochastic trend” confidence intervals and “deterministic trend” confidence intervals will then depend on experience with each individual series and how accurate the “coverages” are of the competing confidence intervals in forecasting applications. That is, do the 95% forecast confidence intervals of a forecasting method encompass roughly 95 out of the next 100 out-of-sample forecasts or not? If the coverage of a method’s 95% confidence intervals are only, say, 65% out of 100 trials, one can probably conclude that the method’s characterization of forecasting uncertainty is amiss and another forecasting method should be entertained if one’s interest goes beyond just accurate point forecasting and also focuses on the accuracy of one’s stated forecast uncertainty. See Corradi and Swanson (2005) for a nice discussion of the validation of competing forecasting methods by comparing not only the accuracies of the point forecasts of the competing methods but also the performance of competing confidence intervals in out-of-sample experiments.

Further Calibration of Exponential Models

If one wanted to even more finely “calibrate” which exponential smoothing model to use for forecasting, a very nice reference to consult is the survey paper by Everette S. Gardner, Jr. “Exponential Smoothing: The State of the Art,” Journal of Forecasting, vol. 4 (1985), pp. 1 – 38. In particular, consider the following reproduction of Exhibit 1 in that paper.

2 *Journal of Forecasting* Vol. 4, Iss. No. 1

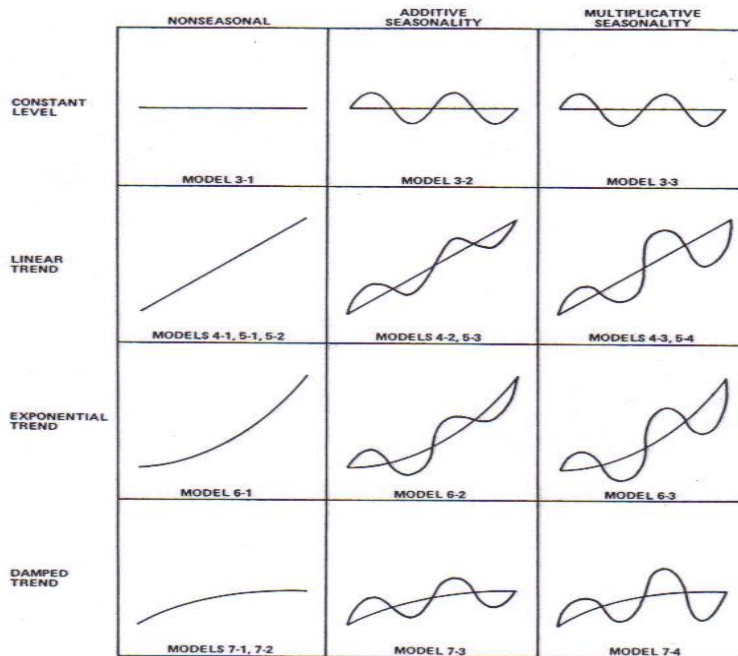


Exhibit 1. Forecast profiles from exponential smoothing

Source: Everette S. Gardner, Jr.
 “Exponential Smoothing: The State of the Art” Journal of Forecasting, Vol. 4 (1985)
 p. 2.

You will notice that the forecast profiles for various exponential smoothing models are broken out into 12 different “boxes” in the table. The rows are labeled by the type of trend present in the data: “constant level” (i.e. no trend), “linear trend”,

“exponential trend”, and “damped trend”. The columns are labeled by the choice of seasonality: “non-seasonal”, “additive seasonality”, and “multiplicative seasonality”. Under each forecast profile is the model number(s) (used in the paper) that map to that specific forecast profile. To use a fishing analogy, “This is like fishing for fish in a barrel!” All you have to do is figure out which box’s forecast profile best matches the “shape” of the data that you have at hand. Then a specific smoothing model can be correspondingly chosen. In the discussion here we have emphasized the methods appropriate for the four boxes in the northwest corner of Gardner’s diagram, the choices being only over trend and seasonality.

Of course, the DTDS model we previously discussed could be beneficial in discriminating between which of the above “boxes” we should choose. Consider the following **two** DTDS models, the first modeling the level of the time series, y_t , and the second modeling the logarithm of the time series, $\log(y_t)$:

$$y_t = \alpha + \beta t + \theta t^2 + \gamma_1 D_{t1} + \gamma_2 D_{t2} + \gamma_3 D_{t3} + \cdots + \gamma_{12} D_{t,12} + \varepsilon_t \quad (40)$$

$$\log(y_t) = \alpha + \beta t + \theta t^2 + \gamma_1 D_{t1} + \gamma_2 D_{t2} + \gamma_3 D_{t3} + \cdots + \gamma_{12} D_{t,12} + \varepsilon_t \quad (41)$$

both specifications having the usual AR(r) autocorrelated error term

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_r \varepsilon_{t-r} + a_t \quad (42)$$

Notice the quadratic term t^2 has been included to allow for testing for the curvature of trend. Going down the rows of the Gardner diagram, the standard test of trend in (40) will allow you to choose between the first two rows while if (41) fits the data better than (40) thus implying exponential growth then you could focus on the third row of “exponential trend”. Comparing the level and log fits of (40) and (41) would require that, when you estimate (41), you translate the fitted values of $\log(y_t)$ into fitted values of y_t and then compare goodness-of-fit criteria across the two equations based on the fitted values of y_t . For example, if one chooses R^2 as the goodness-of-fit criteria, the model with the highest value of R^2 (coefficient of determination) would be the preferred model. If the R^2 of the log model is higher, then you could focus on the exponential trend smoother. Otherwise, you can consider the other models. On the other hand, the damped

trend choice would require that there be some curvature in the data and thus that the quadratic term in (34) be statistically significant.

Distinguishing between the additive and multiplicative models is somewhat more problematic given the assumption of fixed seasonality in the DTDS model. The seasonal effects in the multiplicative models are increasing with time. One way to detect such changing seasonality is to apply tests of parameter constancy in the seasonal coefficients in (34) or (35) as in the Chow test or recursive Chow test. See J. Stock and M. Watson (Pearson, 2007) Introduction to Econometrics, 2nd edition, pp. 566 – 571 for more discussion on this point.

An Example of the Cost to Exponential Smoothing of Not Correctly Identifying the Salient Characteristics of the Time Series to be Forecast

To demonstrate the costs to exponential smoothing of not accurately identifying the salient characteristics of the time series that is being forecasted, an out-of-sample experiment on the Plano Sales Tax Revenue data was conducted whereby the forecasting performance of the preferred smoother, the Additive Winters method with trend and seasonality, is compared to the three other smoothers discussed above, namely the Simple Exponential Smoother (No Trend, No Season), the Linear (Holt) Exponential Smoother (Trend, No Season), and the Seasonal Exponential Smoother (No Trend, Season).

In conducting this experiment the Plano data was divided into two groups: the in-sample data spanning the time period February 1990 and ending December 2003 (167 observations) and the out-of-sample data spanning the time period January 2004 – November 2006 (23 observations). Three forecast horizons were examined, namely, one-step ahead, three steps ahead, and six steps ahead ($h = 1, 3, 6$). The four competing models were “rolled” through the out-of-sample data one observation at a time while updating the smoothing coefficients each time (i.e. refitting the models after each forecast) and then comparing the forecasts with the actual out-of-sample observations using the Percent Mean Absolute Error Criterion (PMAE) defined by

$$\text{PMAE} = 100 \sum_{t=1}^M \left| \frac{e_t}{y_t} \right| \quad (43)$$

where $e_t = y_t - \hat{y}_t$ is the out-of-sample forecast error of a given model and M is the number of out-of-sample forecasts produced by a given model.

The results of this out-of-sample forecasting experiment are reproduced in the table below.

**Out-of- Sample Forecasting Performances
Of Various Exponential Smoothing Methods
At Horizons (h) 1, 3, and 6
And using PMAE as the Measure of Accuracy**

TS	NTNS	TNS	NTS	M	h
4.30	19.87	20.57	4.61	23	1
3.84	14.76	15.34	4.58	21	3
3.50	13.37	13.83	4.88	18	6

Simple Exponential Smoothing = **NTNS** (No Trend, No Seasonality)

Seasonal Exponential Smoothing = **NTS** (No Trend, Seasonality)

Double Exponential Smoothing = **TNS** (Trend, No Seasonality)

Additive Winters Smoothing = **TS** (Trend, Seasonality)

From this table we can see that ignorance with respect to trend and seasonality in the data can be costly to exponential smoothing. For example, with respect to one-step-ahead forecasting ($h=1$), the preferred method (TS) has only a 4.30 PMAE while the other methods have larger PMAEs, especially the methods that ignore the seasonality in the data (NTNS (PMAE = 19.87) and TNS (PMAE = 20.57)). Ignoring the seasonality at horizon one causes the accuracy of the naïve models NTNS and TNS to have roughly 5 times less forecasting accuracy relative to the preferred method (TS). In this particular experiment ignorance of trend is not as costly as ignore of seasonality but this might not be the case in other data sets where the seasonality is “milder” and the trend is “steeper” than in the Plano Sales Tax Revenue data set. Similar conclusions can be drawn at the other horizons $h = 3$ and $h = 6$.

V. Conclusion

In summary, when using exponential smoothing models for forecasting, it is very important to know the salient characteristics of the time series that one is working with (i.e. trend and seasonality among other characteristics) before making a choice of exponential smoother. The better the diagnosis of the salient characteristics of the time series, the more accurate the exponential smoothing forecasts are likely to be. Of course, if there is more time to build forecasting models for a few time series of special importance, one might forego the somewhat mechanical smoothing models and build general Box-Jenkins models for forecasting the series. Since the above smoothing models are special cases of Box-Jenkins models, one is likely to gain forecasting accuracy by building a more general Box-Jenkins model rather than relying on “special case” Box-Jenkins models that are likely to be too restrictive in nature.

REFERENCES

- Archibald, B.C. (1990), "Parameter Space of the Holt-Winters Model," International Journal of Forecasting, 6, 199-209.
- Bowerman, B.L. and O'Connell, R.T. (1979), Time Series and Forecasting: An Applied Approach, Duxbury Press: North Scituate, Mass.
- Brown, R.G. (1959), Statistical Forecasting for Inventory Control, McGraw-Hill: New York, NY.
- Brown, R.G. (1962), Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall: New Jersey.
- Chatfield, C. (1978), "The Holt-Winters Forecasting Procedure," Applied Statistics, 27, 264-279.
- Corradi, Valentina and Norman R. Swanson (2005), "Predictive Density Evaluation," University of London Working paper to appear in the Handbook of Economic Forecasting (forthcoming).
- Gardner, E.S., Jr. (1985), "Exponential Smoothing: the State of the Art," Journal of Forecasting, 4, 1-38.
- Hamilton, James (1994), Time Series Analysis, Princeton University Press: Princeton, NJ.
- Harvey, A.C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter (Cambridge University Press: Cambridge, UK).
- Holt, C.C. et al., (1960), Planning Production, Inventories, and Work Force, Prentice-Hall: Englewood Cliffs, Chapter 14.
- McKenzie, Ed (1984), "General Exponential Smoothing and the Equivalent ARMA Process," Journal of Forecasting, 3, 333-344.
- Makridakis, S. and Hibon, M. (1979), "Accuracy of Forecasting: An Empirical Investigation (with discussion)", Journal of the Royal Statistical Society (A), 142, 97-145.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R. and Winkler, R. (1982), "The Accuracy of Extrapolation (Time

Series) Methods: Results of a Forecasting Competition,” Journal of Forecasting, 1, 111 – 153.

SAS Institute Inc., SAS/ETS Software: Time Series Forecasting System, Version 6, First Edition, Cary, NC: SAS Institute Inc., 1995, pp. 225 – 234.

Sweet, A.L. (1985), “Computing the Variance of the Forecast Error for the Holt-Winters Seasonal Models,” Journal of Forecasting, 4, 235-243.

Stock, J. and M. Watson (2007), Introduction to Econometrics, 2nd edition (Pearson: NY), 566 – 571.

Yar, M. and Chatfield, C. (1990), “Prediction Intervals for the Holt-Winters Seasonal Models,” Journal of Forecasting, 4, 235-243.