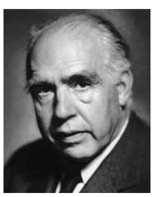
8

Big Data: Baseline Forecasting With Exponential Smoothing Models



PREDICTION IS VERY DIFFICULT, ESPECIALLY IF IT IS ABOUT THE FUTURE

NIELS BOHR (1885-1962), Nobel Laureate Physicist

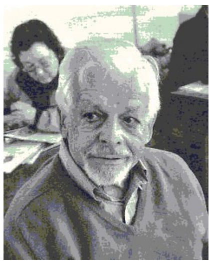
Exponential smoothing models provide a viable framework for forecasting large volume, disaggregate demand patterns. For short-term planning and control systems, these techniques are extremely reliable and have more than adequate track record in forecast accuracy with trend/seasonal data.

This chapter deals with the description and evaluation of techniques that

- are widely used in the areas of sales, inventory, logistics, and production planning as well as in quality control, process control, financial planning and marketing planning
- can be described in terms of a state-space modeling framework that provides prediction intervals and procedures for model selection are well-suited for large-scale, automated forecasting applications, because they require little forecaster intervention, thereby releasing the time of the demand forecaster to concentrate on the few problem cases
- are based on the mathematical extrapolation of past patterns into the future, accomplished by using forecasting equations that are simple to update and require relatively small number of calculations
- capture level (a starting point for the forecasts), trend (a factor for growth or decline) and seasonal factors (for adjustment of seasonal variation) in data patterns

What is Exponential Smoothing?

In chapter 3, we introduced forecasting with simple and weighted moving averages as an exploratory smoothing technique for short-term forecasting of level data. With exponential smoothing models, on the other hand, we can create short-term forecasts with prediction limits for a wider variety of data having trends and seasonal patterns; the modeling methodology offers prediction limits (ranges of uncertainty) and prescribed forecast profiles. Exponential smoothing provides an essential simplicity and ease of understanding for the practitioner, and has been found to have a reliable track record for accuracy in many business applications.



Exponential smoothing was invented during World War II by Robert G. Brown (1923–2013), *left*, who was involved in the design of tracking systems for fire-control information on the location of enemy submarines. Later on, the principles of exponential smoothing were applied to business data, especially in the analysis of the demand for service parts in inventory systems in Brown's book Advanced *Service Parts Inventory Contro*l (1982).

As part of the state-space forecasting methodology, exponential smoothing models provide a flexible approach to weighting past historical data for smoothing and extrapolation purposes. This exponentially declining weighting scheme contrasts with the equal weighting scheme that underlies the outmoded simple moving average technique for forecasting.

Exponential smoothing is a forecasting technique that extrapolates historical patterns such as trends and seasonal cycles into the future.

There are many types of exponential smoothing models, each appropriate for a particular forecast pattern or **forecast profile**. As a forecasting tool, exponential smoothing is very widely accepted and a proven tool for a wide variety of short-term forecasting applications. Most inventory planning and production control systems rely on exponential smoothing to some degree.

We will see that the process for assigning smoothing weights is simple in concept and versatile for dealing with diverse types of data. Other advantages of exponential smoothing are that the methodology takes account of trend and seasonal patterns in time series; embodies a weighting scheme that gives more weight to the recent past than to the distant past; is readily automated, making it especially useful for large-scale forecasting applications; and can be described in a modeling framework needed for deriving useful statistical prediction limits and flexible trend/seasonal forecast profiles.

When selecting a model for demand forecasting, focus on plausible forecast profiles, rather than fit statistics and model coefficients.

For demand forecasting, the disadvantages are that exponential smoothing models do not easily allow for the inclusion of explanatory variables into a forecasting model and cannot handle business cycles. Hence, when forecasting economic variables, such techniques are not expected to perform well on business data that exhibit cyclical turning points.

Smoothing Weights

To understand how exponential smoothing works, we need first to understand the concept of exponentially decaying weights. Consider a time series of production rates (number of completed assemblies per week) for a 4-week period in the table below. In order to predict next period's (*T* + 1) production rate without having knowledge of or information about future demand, we assume that the following week will have to be an average week for production. A reasonable projection for the following week can be based on taking an average of the production rates during past weeks. However, what kind of average should we propose?

: Week	Production
Three periods ago (T - 3)	266
Two periods ago (T - 2)	411
Previous (T - 1)	376
Current t = T	425

Equally Weighted Average. The simplest option. Described in Chapter 3, is to select an equally weighted average, which is obtained by given equal weight to each of the weeks of available data:

$$(425 + 376 + 411 + 266) / 4 = 370$$

This equally weighted average is simply the arithmetic mean of the data, the same basis underlying the moving average.. The forecast of next week's production rate is 370 assemblies. Implicitly, we are assuming that events of 2 and 3 weeks prior (e.g., the more distant past) are as relevant to what may happen next week as were events of the most current and prior week.

In Figure 8.1, a weight is denoted by w_i, where the subscript *i* represents the number of weeks into the past. For an equally weighted average, the weight given to each of the terms is 1/*n*, where *n* is the number of time periods. With *n* = 4, each weight in column 3 is equal to 1/4.

Week	Weights	Equal	Naïve_1	Linear Decay	Exponential Decay	Adjusted Weights
$T-3$	w_4	0.25	0	0.1	0.0625	0.0667
$T-2$	w_3	0.25	0	0.2	0.125	0.1333
$T-1$	w_2	0.25	0	0.3	0.25	0.2667
T	w_1	0.25	1	0.4	0.5	0.5333
Sum		1.00	1.00	1.00	0.9375	1.0000

Figure 8.1 Smoothing weights.

If we consider only the latest week, we have another option, shown in column 4 of Exhibit 8.1, which is the Naïve_1 forecast; it places all weight on the most recent data value. Thus, the forecast for next week's production rate is 425, the same as the current week's production. This forecast makes sense if only the current week's events are relevant in projecting the following week. Whatever happened before this week is ignored.

Exponentially Decaying Weights. Most business forecasters find a middle ground more appealing than either of the two extremes, equally weighted or Naïve_1. In between lie weighting schemes in which the weights decay as we move from the current period to the distant past.

$$w_1 > w_2 > w_3 > w_4 > \ldots$$

The largest weight, w_1, is given to the most recent data value. This means that to forecast next week's production rate, this week's figure is most important; last week's is less important, and so forth.

Many other patterns are possible with decaying weight schemes. As illustrated by column 5 of Figure 8.1, the weight starts at 40% for the most recent week and decline steadily to 10% for week T - 3. Our forecast for week $t = T + 1$ is the weighted average with decaying weights*:*

$$425 \times 0.4 + 376 \times 0.3 + 411 \times 0.2 + 266 \times 0.1 = 392$$

This weighted average gives a production rate forecast that is more than that of the equally weighted average and less than that of the Naïve_1, in this case*.*

An exponentially weighted average refers to a weighted average of the data in which the weights decay exponentially.

The most useful example of decaying weights is that of exponentially decaying weights*,* in which each weight is a constant fraction of its predecessor. A fraction of 0.50 implies a decay rate of 50%, as shown in column 6 of Figure 8.1. In forecasting next period's value, the current period's value is weighted 0.5, the prior week half of that at 0.25, and so forth with each new weight 50% of the one before. (These weights must be adjusted to sum to unity as in column 7.) From Figure 8.1, we can see that the adjusted weights are obtained by dividing the exponential decay weights by 0.9375.

ANALYSIS OF WEIGHTS ON PAST DATA
TITLE1: WEIGHTS ON PAST DATA — SMOOTHING WEIGHT = 0.64
TITLE2: SIMPLE EXPONENTIAL SMOOTHING — NBR. OF PERIODS = 12
X-AXIS: PERIOD
Y-AXIS: WEIGHT

Year	Month	Per	Data	Weight	Data*Weight
		0	14	0.000005	0.000066
1989	1	1	14	0.000008	0.000118
	2	2	11	0.000023	0.000257
	3	3	12	0.000065	0.000780
	4	4	14	0.000181	0.002528
	5	5	12	0.000502	0.006018
	6	6	14	0.001393	0.019504
	7	7	18	0.003870	0.069657
	8	8	17	0.010750	0.182742
	9	9	19	0.029860	0.567337
	10	10	16	0.082944	1.327104
	11	11	18	0.230400	4.147200
	12	12	16	0.640000	10.240000
	SUM			1.000000	16.563312

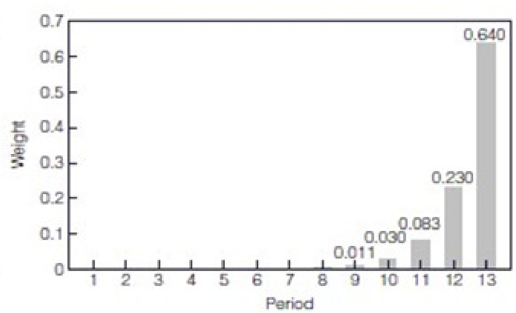


Figure 8.2 (*left*) Calculation of weighted averages of past data.

Figure 8.3 (*right*) Exponentially decaying weights for simple exponential smoothing..

Figure 8.2 illustrates the weighted average of all past data, with recent data receiving more weight than older data. The most recent data is at the bottom of the spreadsheet. The weight on each data value is shown in Figure 8.3. The weights decline exponentially with time, a feature that gives exponential smoothing its name.

The Simple Exponential Smoothing Method

All exponential smoothing techniques incorporate an exponential-decay weighting system, hence the term exponential. Smoothing refers to the averaging that takes place when we calculate a weighted average of the past data. To determine a one-period-ahead forecast of historical data, the projection formula is given by

$$Y_t\,(1) = \alpha\, Y_t + (1 - \alpha)\, Y_{t-1}\,(1)$$

where $Y_t\,(1)$ is the smoothed value at time t, based on weighting the most recent value Y_t with a weight α (α is a smoothing parameter) and the current period's forecast (or previous smoothed value) with a weight (1 - α). By rearranging the right-hand side, we can rewrite the equation as

$$Y_t\,(1) = Y_{t-1}\,(1) + \alpha\, [Y_t - Y_{t-1}\,(1)]$$

which can be interpreted as the current period's forecast $Y_{t-1}\,(1)$ adjusted by a proportion α of the current period's forecast error $[Y_t - Y_{t-1}\,(1)]$.

The simple exponential smoothing method produces forecasts that are a level line for any period in the future, but it is not appropriate for projecting trending data or patterns that are more complex.

We can now show that the one-step-ahead forecast $Y_t\,(1)$ is a weighted moving average of all past values with the weights decreasing exponentially. If we substitute for $Y_{t-1}\,(1)$ in the first smoothing equation, we find that:

$$Y_t\,(1) = \alpha\, Y_t + (1 - \alpha)\, [\alpha\, Y_{t-1} + (1 - \alpha)\, Y_{t-2}\,(1)]$$

$$= \alpha\, Y_t + \alpha\,(1 - \alpha)\, Y_{t-1} + (1 - \alpha)^2\, Y_{t-2}\,(1)$$

If we next substitute for $Y_{t-2}\,(1)$, then for $Y_{t-3}\,(1)$, and so, we obtain the result

$$Y_t\,(1) = \alpha\, Y_t + \alpha\,(1 - \alpha)\, Y_{t-1} + \alpha\,(1 - \alpha)^2\, Y_{t-2} + \alpha\,(1 - \alpha)^3\, Y_{t-3} + \alpha\,(1 - \alpha)^4\, Y_{t-4}$$

$$+ \;.\;.\;.\;. \; + \alpha\,(1 - \alpha)^{t-1}\, Y_1 + (1 - \alpha)^t\, Y_0\,(1)$$

The one-step-ahead forecast $Y_T\,(1)$ represents a weighted average of all past values. For three selected values of the parameter α, the weights that are assigned to the past values are shown in the following table:

Weight Assigned to:	α = 0.1	α = 0.3	α = 0.5	α = 0.9
Y_T	0.1	0.3	0.5	0.9
Y_{T-1}	0.09	0.21	0.25	0.09
Y_{T-2}	0.081	0.147	0.125	0.009

In Figure 8.4, we calculate a forecast of the production data, assuming that α = 0.5. (The production data are repeated in Figure 8.4, in the Actual column.) To use the formula, we need a starting value for the smoothing operation - a value that represents the smoothed average at the earliest week of our time series, here t = T - 3. The simplest choice for the starting value is the earliest data point. In our example, the starting value for the exponentially weighted average is the production rate for week t = T - 3, which was given as 266. The final result, $Y_T(1) = 391$ (rounded) for week t = T, is called the current level. It is a weighted average of 4 weeks of data, where the weights decline at a rate of 50% per week.

We defined a one-period-ahead forecast made at time t = T to be $Y_T(1)$. Likewise, the m-period-ahead forecast is given by $Y_T(m) = Y_T(1)$, for m = 2, 3, ... For a time series with a relatively constant level, this is a good forecasting technique. We called this simple smoothing in Chapter 3, but it is generally known as the simple exponential smoothing.

Week	Y	Actual	$Y_t(1)$	Formula	Error $Y_t - Y_{t-1}(1)$
$T-3$	Y_{T-3}	266	266	$Y_{(t-3)}(1) = Y_{(t-3)}$	
$T-2$	Y_{T-2}	411	339	$Y_{(t-2)}(1) = 0.5 \times Y_{(t-2)} + 0.5 \times Y_{(t-3)}(1)$	72
$T-1$	Y_{T-1}	376	357	$Y_{(t-1)}(1) = 0.5 \times Y_{(t-1)} + 0.5 \times Y_{(t-2)}(1)$	19
T	Y_T	425	391	$Y_t(1) = 0.5 \times Y_t + 0.5 \times Y_{(t-1)}(1)$	34

Figure 8.4 Updating an exponentially weighted average.

1	SIMPLE EXPONENTIAL SMOOTHING						
2	TITLE1: SIMPLE EXPONENTIAL SMOOTHING					WEIGHT: 0.64	
3	TITLE2: COMPANY TRAVEL EXPENSES					MSE: 4.7056	
4	X-AXIS: MONTH						
5	Y-AXIS: THOUSANDS OF $						
6	A	B	C	D	E	F	G
7	Year	Mon	Per	Data	Fcst	Error	Error²
8	1989	1	1	14	14.0	0.0	0.0
9		2	2	11	14.0	−3.0	9.0
10		3	3	12	12.1	−0.1	0.0
11		4	4	14	12.0	2.0	3.9
12		5	5	12	13.3	−1.3	1.7
13		6	6	14	12.5	1.5	2.4
14		7	7	18	13.4	4.6	20.7
15		8	8	17	16.4	0.6	0.4
16		9	9	19	16.8	2.2	5.0
17		10	10	16	18.2	−2.2	4.8
18		11	11	18	16.8	1.2	1.5
19		12	12	16	17.6	−1.6	2.4
20	1990	1	13		16.6		

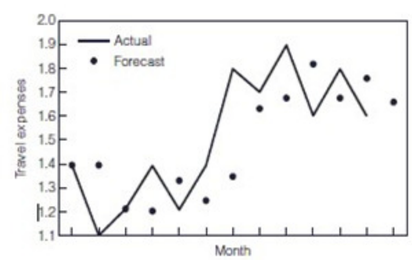


Figure 8.5 (*left*) Forecasting with simple exponential smoothing company travel expenses.

Figure 8.6 (*right*) Simple exponential smoothing: company travel expenses and one-period ahead forecasts.

The forecast profile of the simple exponential smoothing method is a level horizontal line.

Simple exponential smoothing works much like an automatic pilot or a thermostat. At each time period, the forecasts are adjusted according to the sign of the forecast error (actual data minus forecast.) If the current forecast error is positive, the next forecast is increased; if the error is negative, the forecast is reduced.

To get the smoothing process started (Figure 8.5), we set the first forecast (cell E8) equal to the first data value (cell D8). We can also use the average of the first few data values. Thereafter, the forecasts are updated as follows: In column F, each error is equal to actual data minus forecast. In column E, each forecast is equal to the previous forecast plus a fraction of the previous error. This fraction is called the **smoothing weight** (cell I2).

But how do we select the smoothing weight? The smoothing weight is usually chosen to minimize the mean square error (MSE), a statistical measure of fit. This smoothing weight is called *optimal,* because it is our best estimate based on a prescribed criterion (MSE). Forecasts, errors, and squared errors are shown in columns E, F, and G.

The one-step-ahead forecast (=16.6 in cell E20) extends one period into the future. The travel expense data, smoothed values, and the one-period-ahead forecast are shown graphically in Figure 8.6.

Forecast Profiles for Exponential Smoothing Methods

A system of exponential smoothing models can be classified by the type of trend and/or seasonal pattern generated as the forecast profile. The most appropriate technique to use for any forecasting should match the profile expected or desired in an application. Figure 8.7 shows the extended Pegels classification for 12 forecasting profiles for exponential smoothing developed by Everette S. Gardner, *left*, in a seminal paper *Exponential smoothing: The state of the art*, Journal of Forecasting. 1985.

A Pegels classification of exponential smoothing methods gives rise to 12 forecast profiles for trend and seasonal patterns.

After a preliminary examination of the data from a time plot, we may be able to determine which of the dozen models seems most suitable. In Figure 8.7, there are four types of trends to choose from (Nonseasonal column), and two types of seasonality (Additive and Multiplicative).

Each profile can be directly associated with a specific exponential smoothing model (Figure 8.8), as described the in the next section (some of which are referred to by a common name attributed to their authors). We now explain how each model works to generate forecasts; that is, we describe how each model produces the appropriate forecasting profile.

For a downwardly trending time series, multiplicative seasonality appears as steadily diminishing swings about a trend. For level data, the constant-level multiplicative and additive seasonality techniques give the same forecast profile.

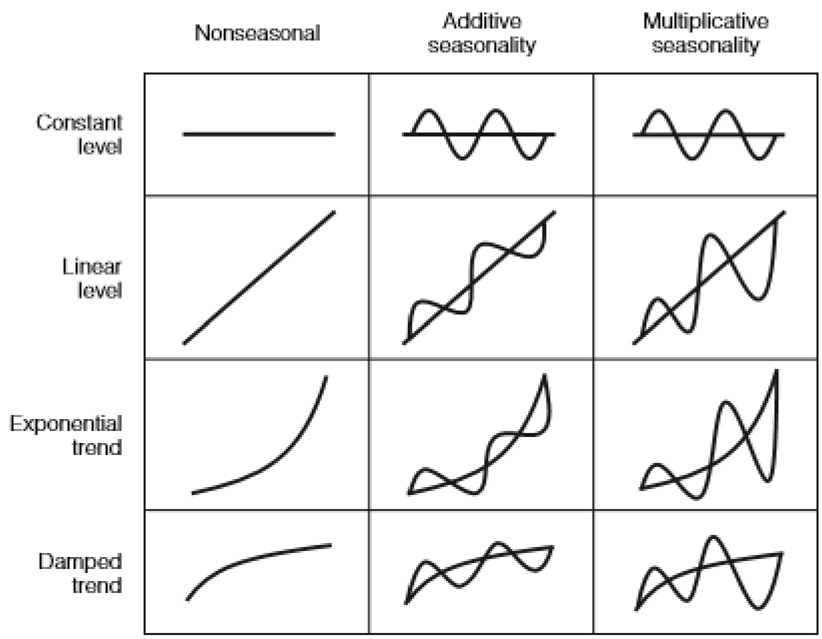


Figure 8.7 Pegels' classification of exponential smoothing methods extended to include the damped trend technique.

Model Name	Trend profile	Seasonal Profile	State Space Classification
Simple (single)	**N**one	**N**one	(N, N)
Holt	**A**dditive (Linear)	**N**one	(A, N)
Holt-Winters	**A**dditive (Linear)	**A**dditive or **M**ultiplicative	(A, A) or (A, M)

Figure 8.8 Most commonly implemented exponential smoothing methods.

The complete chapter can be found in

Change & Chance Embraced

ACHIEVING AGILITY WITH DEMAND FORECASTING IN THE SUPPLY CHAIN

HANS LEVENBACH, PhD

ISBN-10:0692945989
ISBN-13:978-069294588

Cover art: Rose Tannenbaum, Berkshire TypeGraphic, www.tgo.com

Cartoon art: Don Engel, Isographic LLC

Copy editor: David Coen, Freelance copyeditor

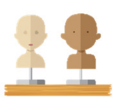
Printed by Createspace, www.createspace.com

Professional development: Certified Professional Demand Forecaster (CPDF®), www.cpdftraining.org

Contents

:

Chapter 4 - Taming Uncertainty: What You Need to Know about Measuring Forecast Accuracy... 80



Chapter 5 - Characterizing Demand Variability: Seasonality, Trend, and the Uncertainty Factor 114

Chapter 11 - Gaining Credibility Through Root-Cause Analysis and Exception Handling 283

Chapter 12 - The Final Forecast Numbers: Reconciling Change & Chance 316