Forecasting With Exponential Smoothing – What's The Right Smoothing Constant?

Handanhal V. Ravinder, Montclair State University, USA

ABSTRACT

This paper examines exponential smoothing constants that minimize summary error measures associated with a large number of forecasts. These forecasts were made on numerous time series generated through simulation on a spreadsheet. The series varied in length and underlying nature – no trend, linear trend, and nonlinear trend. Forecasts were made using simple exponential smoothing as well as exponential smoothing with trend correction and with different kinds of initial forecasts. We found that when initial forecasts were good and the nature of the underlying data did not change, smoothing constants were typically very small. Conversely, large smoothing constants indicated a change in the nature of the underlying data or the use of an inappropriate forecasting model. These results reduce the confusion about the role and right size of these constants and offer clear recommendations on how they should be discussed in classroom settings.

Keywords: Exponential Smoothing; Smoothing Constants; Forecast Error; Trend; Time Series

INTRODUCTION

Exponential smoothing is a very popular forecasting method. It is taught to graduate and undergraduate business students in introductory courses in operations, management science, marketing, and sometimes statistics. It is easy to understand and use and most commercial forecasting software products include it in their offerings.

Exponential smoothing forecasting methods use constants that assign weights to current demand and previous forecasts to arrive at new forecasts. Their values influence the responsiveness of forecasts to actual demand and hence influence forecast error. Considerable effort has focused on finding the appropriate values to use. One approach is to use smoothing constants that minimize some function of forecast error. Thus, in order to select the right constants for forecasting, different values are tried out on past time series, and the ones that minimize an error function like Mean Absolute Deviation (MAD) or Mean Squared Error (MSE) are the ones used for forecasting.

The purpose of this paper is to examine the relationship between the magnitude of error-minimizing constants and the underlying nature of the past time series. Specifically it looks at optimal values of the smoothing constants as a function of the underlying trends in the data (or their absence) as well the length of the time series. The role of initial forecasts is also examined. The results should reduce the confusion on the right magnitude of these smoothing constants that is seen in introductory treatment of this issue.

Exponential Smoothing

Two exponential smoothing models are popular, especially in classroom settings – simple exponential smoothing, and exponential smoothing with trend correction (often referred to as double exponential smoothing). Gardner (1985, 2006) provides a detailed review of exponential smoothing.

Simple Exponential Smoothing

Here, demand is level with only random variations around some average. The forecast F_{t+1} for the upcoming period is the estimate of average level L_t at the end of period *t*.

$$F_{t+1} = L_t = F_t + \alpha(D_t - F_t) = \alpha D_t + (1 - \alpha)F_t \tag{1}$$

where α, the smoothing constant, is between 0 and 1. The new estimate of level may be seen as a weighted average of D_t, the most recent information of average level, and F_t the previous estimate of that level. Small values of α imply that the revision of the old forecast, in light of the new demand, is small; the new forecast is not very different from the previous one. The method requires an initial forecast F_1 which has to be either assumed or estimated.

Exponential Smoothing with Trend Adjustment (Double Exponential Smoothing)

Here, the time series exhibits a trend; in addition to the level component, the trend (slope) has to be estimated. The forecast, including trend for the upcoming period *t+1,* is given by

$$F_{t+1} = L_t + T_t \tag{2}$$

Here, L_t is the estimate of level made at the end of period *t* and is given by

$$L_t = \alpha D_t + (1 - \alpha)F_t \tag{3}$$

T_t is the estimate of trend at the end of period *t* and is given by

$$T_t = \beta(F_t - F_{t-1}) + (1 - \beta)T_{t-1} \tag{4}$$

β is also a smoothing constant between 0 and 1 and plays a role similar to that of α.

Again, small values of α and β imply that consecutive estimates of level and trend components do not differ much from each other. Any revision in the light of the new demand is small. This method requires estimation of the initial level component L_1 and the initial trend component T_1 to start off the series of forecasts.

In both cases, the choice of initial forecasts has significant impact on the performance of forecasts (Ledolter and Abraham, 1984).

Smoothing Constants

Smoothing constants are key to successful forecasting with exponential smoothing, but there are no consistent guidelines in the forecasting literature on how they should be selected. Introductory treatments of forecasting will recommend that the smoothing constants be kept small, in the 0.1 to 0.3 range; see, for example, Schroeder, Rungtusanatham, & Goldstein (2013) and Jacobs & Chase (2013). Others (e.g., Paul, 2011) recommend the use of smoothing constants that minimize some function of error over past data, the minimization being done conveniently with a nonlinear optimizer (Bermudez, Segura, & Welcher, 2006; Chopra & Meindl, 2013). Often, this minimization will result in constants that are outside the range recommended.

In a previous paper, Ravinder (2013), we examined the error-minimization approach from a classroom teaching perspective. For textbook-type problems, the approach works well, but often produces smoothing constants that are zero or, if non-zero, well distributed over the entire 0-1 range. Researchers also recommend optimizing the starting forecast along with the smoothing constants (Bermudez, Segura, & Vercher, 2006). When this is done, there are even more problems where the constants are zeroes. The general conclusion of that paper was that when the data was well-behaved (i.e., there were no changes in trend), then a good initial estimate of level and/or slope would minimize the error function and the role of the smoothing constant would become inconsequential. A limitation of the previous paper was that it dealt with textbook problems involving small time series. It is difficult to find measurable departures from stability in small time series. The findings were thus of limited generalizability.

This paper seeks to remedy this by systematically examining longer time series that incorporate linear and nonlinear trends. It examines the optimal smoothing constants under various combinations of series duration and series behavior. It examines both simple exponential smoothing and double exponential smoothing. The results should provide guidance on the choice of the appropriate smoothing constants in a given forecasting situation.

The next section describes the approach used in this paper to generate time series of various lengths and characteristics, to make forecasts, and to examine the resulting error-minimizing smoothing constants. Next comes a discussion of the results and the conclusions that can be drawn from them.

APPROACH

For simple exponential smoothing, two types of demand series were generated using Excel's random variable generation capabilities. Details are shown in Appendix 1.

In the first type, demand was stable and normally distributed around an arbitrarily chosen mean and standard deviation. This situation represents the case for which simple exponential smoothing is traditionally recommended. The length of the series (number of periods for which demand was generated) varied from 12 to 60 in steps of 6.

In the second type, too, demand was normally distributed, but this time with a mean that was a linear function of time. The intercept and slope were arbitrarily picked. The standard deviation of demand was also arbitrarily picked and invariant with respect to time. It was also the same as in the first type of data set. Again, series length varied from 12 to 60. This situation is not considered ideal for simple exponential smoothing because of its slowness in responding to trend. This set-up allows us to examine the magnitudes of α that are thrown up by the optimization process for both types of data and to compare them. Examples of these two types of time series are shown in Appendix 2.

For each type of data set and each value of sample size, 100 data sets were generated. For each data set, simple exponential smoothing was used to generate forecasts for each time period. Forecast error was measured through Mean Absolute Deviation (MAD) and Mean Squared Error (MSE). Excel's Data Table feature was used to calculate MAD and MSE associated with different values of α. Thus, it was possible to identify optimal values of α - values that minimize MAD and MSE. The starting forecast F_1 has a significant impact on MAD and MSE and thus on optimal α. To assess this impact, two cases of F_1 were considered in each data set. In the first, F_1 was set equal to D_1. This is a fairly common assumption in textbooks. In the second case, F_1 was set equal to the average demand.

A similar approach was used for double exponential smoothing. Two types of data were generated. The first one, $D_t = a+bt$, had a linear trend, along with random fluctuations about this trend. The components a and b were arbitrarily selected. The second data set had a nonlinear trend, specifically, demand $D_t = a+b*\ln t$, where a and b were arbitrarily chosen. Again, there were random fluctuations about this trend. Examples of the types of series generated are shown in Appendix 2.

The double exponential smoothing process requires initial values of level (L_1) and trend (T_1). Two cases were explored. In the first case, the estimate of L_1 was the intercept of the least-squares line fitted to the data and the estimate T_1 was the slope of that least-squares line. In the other case, commonly seen in introductory textbooks, L_1 was simply set to D_1 and T_1 assumed to be zero. As in the case of simple exponential smoothing, series length varied from 12 periods to 60 periods. Here, too, the experimental set-up allowed comparison of the smoothing constants – α and β - that were best in each of the two cases.

RESULTS – SIMPLE EXPONENTIAL SMOOTHING

Figure 1 summarizes the results with the Mean Absolute Deviation criterion. Figures 1 and 2 provide more detail for the four cases discussed in the table. MSE results are similar and shown in Appendix 3, Exhibit 1.

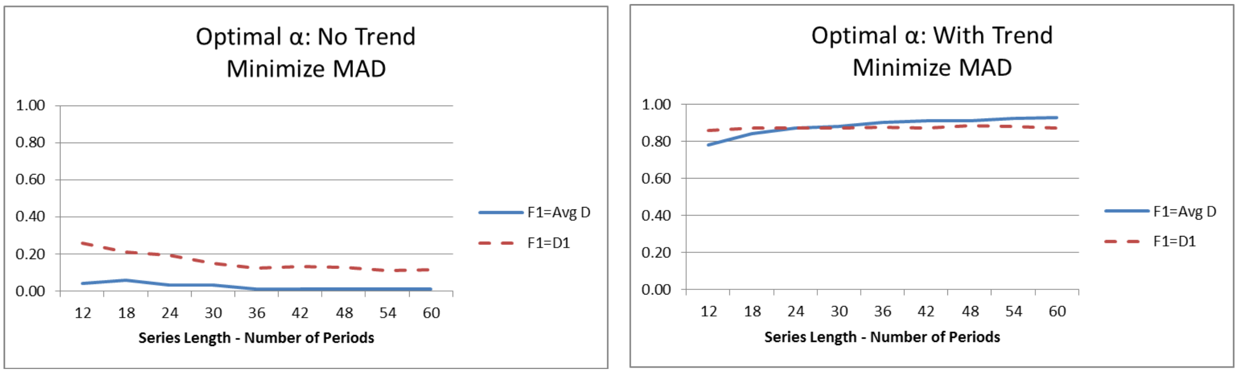


Figure 1: Optimal α And The Impact Of Series Length – No Trend, With Trend

Stable Series – No Trend

When the series is stable with no trend – the situation for which exponential smoothing has been traditionally prescribed – α values tend to be low, in the 0 – 0.30 range. For longer series, $n \geq 36$, the range is even narrower, 0 – 0.15.

The initial forecast F_1 makes a difference. If the initial forecast is just one value from the series, as when $F_1 = D_1$, the value of optimal α depends on how representative D_1 is of the entire series. With series of only a few periods, D_1 is likely to be unrepresentative. An unrepresentative D_1 means that forecasts need more revision as new demands become available; correspondingly, α has to be high. However, when F_1 is more representative of the series, as when $F_1 = \overline{D}$, little adjustment is needed to each forecast as the series unfolds; α values tend to be very low or even zero.

Stable Series – Linear Trend

When there is a trend in the time series, α is high (> 0.75), regardless of the length of the series or the kind of initial forecast used. With trend present, more weight is given to the actual demand than to the previous forecast; every forecast needs revision. With shorter series, $n \leq 24$, forecasts with $F_1 = \overline{D}$, tend to have smaller optimal α values; but as the series gets longer, the average across the entire series becomes more unrepresentative of the series. Subsequent forecasts need more revision and optimal α tends to be high.

RESULTS – DOUBLE EXPONENTIAL SMOOTHING

Series With Linear Trend

For the times series with a linear trend, Figure 2 shows values of α and β that minimize MAD. Data sets vary in size from n=12 to n=60.

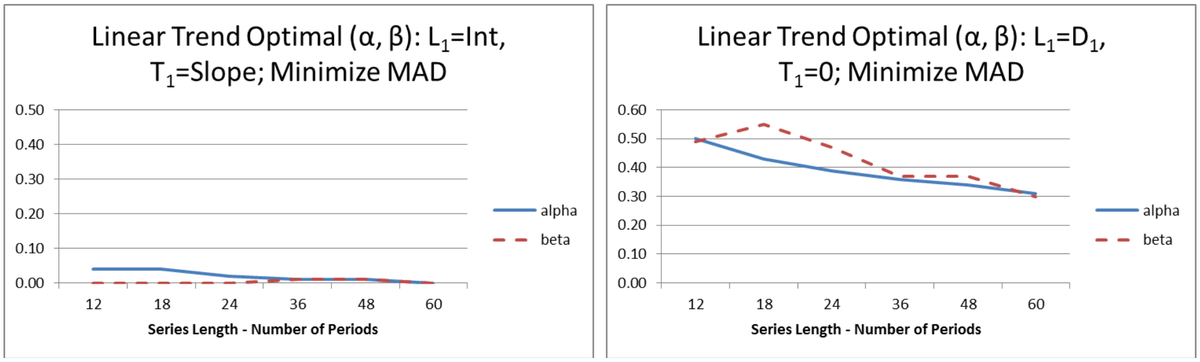


Figure 2: Optimal (α, β) And The Impact Of Series Length – Series With Linear Trend

The behavior of α and β are shown in the graphs above. The difference between the two graphs is in the initial estimates of L_1 and T_1. In the first graph, L_1 is estimated as the intercept of the least-squares line fitted to the data and T_1 as its slope. Both α and β are very small - less than 0.05 over the entire range of series lengths - 12 to 60. This makes sense. When the initial estimates of level and trend are good, as they would be if they were the least squares estimates of intercept and slope, then subsequent estimates of level and trend need little adjustment; so α and β tend to be small.

In the second graph, L_1 and T_1 are arbitrarily picked as L_1=D_1 and T_1=0. Both α and β are much bigger, in the 0.30 to 0.60 range, because of the greater need for revision as each new demand is received. The interesting result here is that both α and β decrease as the series length increases. The explanation is that the impact of arbitrary L_1 and D_1 is diluted as the series gets longer; each new forecast needs less adjustment. Results for MSE are very similar and are presented in Appendix 1.

Series With Nonlinear Trend

Figure 3 shows the results for the case where the series has an underlying nonlinear trend. Again, the two graphs relate to two different assumptions about the initial forecasts. In the first graph, L_1 is estimated as the intercept of the least-squares line fitted to the data and T_1 as its slope. In the second graph, L_1 and T_1 are arbitrarily picked to be D_1 and 0.

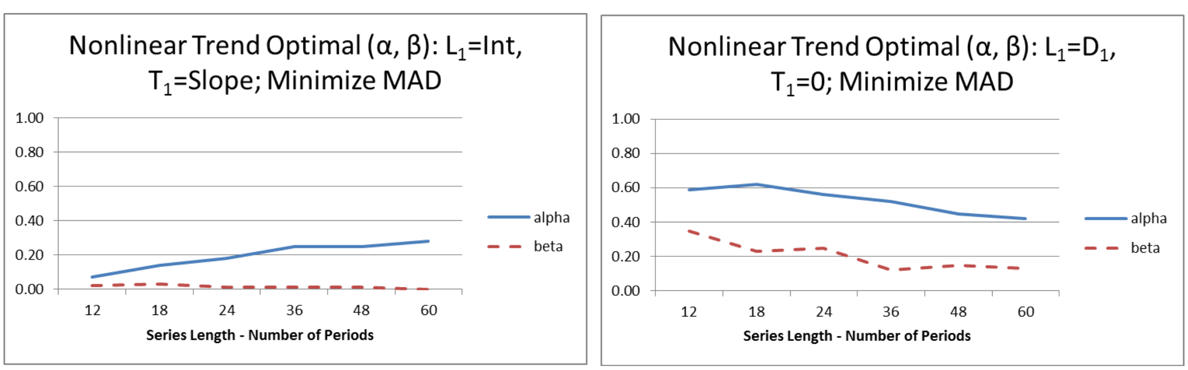


Figure 3: Optimal (α, β) And The Impact Of Series Length – Series With Nonlinear Trend

In the first graph, β is almost zero for all series lengths. Little adjustment is made to the trend component as each new forecast is calculated. The least-squares slope provides a good estimate of the average trend in the data, even though the series is not linear. Of course, the extent of departure from linearity matters and, in this case, it is not severe enough to disqualify the least-squares slope as a good approximation of trend. However, α increases as the series length increases. This means that the intercept becomes less representative as the series gets longer and the level estimate needs more adjustment in the form of a larger α.

In the second graph, L_1 and T_1 are arbitrarily picked to be D_1 and 0, respectively. Both α and β decrease slowly as the length of the series increases. One explanation for this has to do with the particular form of the nonlinearity underlying these series. Much of the nonlinearity is present in the earlier part of the series. For larger values of t, the function increases at a decreasing rate. For t larger than 12, the function is practically linear. One can imagine larger values of α and β being needed for the initial part of the series, but as the series straightens out and becomes more predictable, the values of α and β get smaller. If we can think of the optimization process as producing values of α and β that work well, on average, across the length of the series, these values will get smaller as the length of the series increases. Results for MSE are very similar and presented in Appendix 3, Exhibit 3.

This explanation brings up a larger point in the form of a caveat. The behavior of α and β, when the underlying series is nonlinear, depends on the particular form of the nonlinearity and no easy generalizations are possible. For illustration, this paper discussed one functional form that resulted in a concave non-decreasing demand series. It would be hazardous to generalize from the results with this function to situations where other forms of nonlinearity are present.

CONCLUSIONS

The results of this simulation-based study of optimal smoothing constants allow us to offer the following suggestions to teachers of forecasting, as well as to writers of introductory textbooks that discuss forecasting:

1. When there is no trend in the data, simple exponential smoothing will yield minimum error when α values are small, in the range 0.0 – 0.3. This is true of small series (n=12) as well as large (n=60). This confirms the recommendation of many textbooks. Up to this point, these kinds of recommendations have had an ad hoc flavor to them because there hasn't been published evidence. We believe our results provide this evidence.
2. When the initial forecast is good, α values will very often be zero. In fact, small non-zero values of α are indicative of local trends. Larger non-zero values of α are indicative of sustained trends which might be better accounted for with a technique, like double exponential smoothing.
3. Large values of the smoothing constants are certainly possible and should not be rejected without detailed examination of the underlying series or of the quality of the initial forecasts used.
4. When there is a linear trend in the data, the performance of double exponential smoothing depends on the initial estimates of the level and trend components. Where these are good, α and β will be very small. This is true of small as well as large series. Larger values might be indicative of poor initial estimates of level and trend. The impact of poor initial forecasts is felt less on longer series than on smaller ones. The values of α and β decrease with series length.
5. When there is a nonlinear trend in the data, the results are mixed and not easily generalizable. The best values of α and β depend on the particular kind of nonlinearity involved. The best approach is to graph the time series and pick appropriate starting values before finding the optimal values of α and β.

In summary, then, when exponential smoothing methods are used in the situations for which they are appropriate (simple exponential smoothing where there is no underlying trend and double exponential smoothing where there is an underlying linear trend), along with good starting forecasts, the best smoothing constants tend to be very small, if not zero. Significantly large smoothing constants signal the presence of either trend (simple exponential smoothing) or changes in trend (double exponential smoothing). This is a strong argument for the use of adaptive smoothing methods – methods that monitor forecast errors continuously and change the smoothing constants to keep them within predetermined limits – and more coverage of them in forecasting classes.

A limitation of the approach used is that it is based on simulated data. A disadvantage of simulation-based studies is that their results cannot be extrapolated easily to situations that are not explored in the simulation. Another drawback is that it gets cumbersome to add more variables to the model. For example, it is reasonable to think that the variability of the data around a trend would impact forecast errors and thus the optimal values of the smoothing constants. This could not be explored in this paper.

Despite these limitations, the results of this paper clarify the role of smoothing constants in exponential smoothing and offer useful guidelines to their selection.

AUTHOR INFORMATION

Handanhal Ravinder has been associate professor in the Information and Operations Management department in the School of Business at Montclair State University since Fall 2012. Prior to that, he spent 12 years in the healthcare industry in various market research and business analytics-related positions. Dr. Ravinder received his Ph.D. from the University of Texas at Austin and taught for many years at the University of New Mexico. His research interests are in operations management, healthcare supply chain, and decision analysis. Some of his previous publications have appeared in *American Journal of Business Education*, *Management Science*, *Decision Analysis*, and *Journal of Behavioral Decision Making*. E-mail: ravinderh@mail.montclair.edu

REFERENCES

1. Bermudez, J.D., Segura, J.V., and Velcher, E. (2006). Improving Demand Forecasting Accuracy Using Nonlinear Programming Software. *Journal of the Operational Research Society*, 57, 94-100.
2. Chopra, S. and Meindl, P. (2013). *Supply Chain Management: Strategy, Planning, and Operation*, 5th Edition, Prentice Hall, Page 195, Chapter 7.
3. Gardner, E.S. (1985). Exponential Smoothing – The State of the Art. *Journal of Forecasting*, 4, 1-28.
4. Gardner, E.S. (2006). Exponential Smoothing – The State of the Art – Part II, *International Journal of Forecasting*, 22, 637-666.
5. Jacobs, F. R. and Chase, R.B. (2013). *Operations and Supply Chain Management: The Core*, 3rd Edition, McGraw-Hill Higher Education, Page 59, Chapter 3.
6. Ledolter, J. and Abraham, B. (1984), Some comments on the initialization of exponential smoothing. *Journal of Forecasting*, Vol.3, Issue 1, January/March.
7. Paul, S.K. (2011). Determination of Exponential Smoothing Constant to Minimize Mean Square Error and Mean Absolute Deviation. *Global Journal of Research in Engineering*, Vol. 11, Issue 3, Version 1.0.
8. Ravinder, H.V. (2013). Determining the Optimal Values of Exponential Smoothing Constants – Does Solver Really Work? *American Journal of Business Education*, Vol. 6, No.3, May/June.
9. Schroeder, R., Rungtusanatham, M. J., and Goldstein, S. (2013). *Operations Management in the Supply Chain: Decisions and Cases*, 6th Edition, McGraw-Hill Higher Education, Page 261, Chapter 11.

APPENDIX 1

Simulation Details

Simple Exponential Smoothing

No Trend

Demand is assumed to be normally distributed with a mean of 100 and standard deviation of 15.

EXCEL: =NORMINV(rand(),100,15)

With Linear Trend

Demand is normally distributed with a mean of 100+10t and a standard deviation of 15.

EXCEL: =NORMINV(rand(),100+10*<t>,15), where <t> refers to the cell reference of the time period.

Double Exponential Smoothing

With Linear Trend

Demand is normally distributed with a mean of 100+10t and a standard deviation of 15.

EXCEL: =NORMINV(rand(),100+10*<t>,15), where <t> refers to the cell reference of the time period.

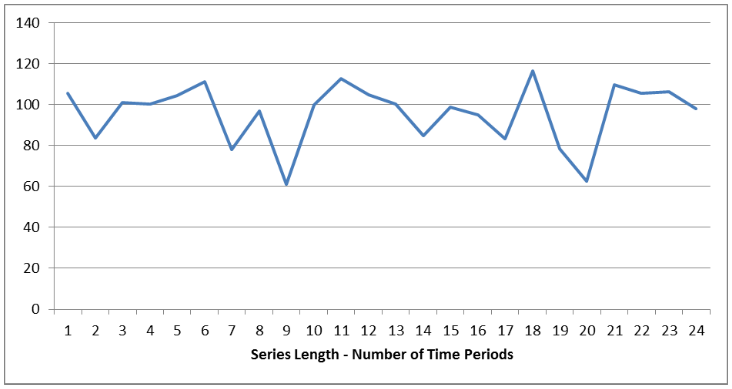
With Nonlinear Trend

Demand is normally distributed with a mean of 100+40Ln(t) and a standard deviation of 15.

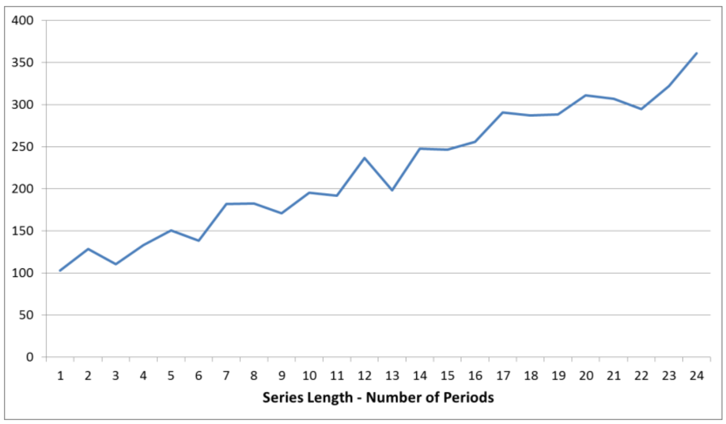
EXCEL: =NORMINV(rand(),100+40*Ln(<t>),15)

APPENDIX 2

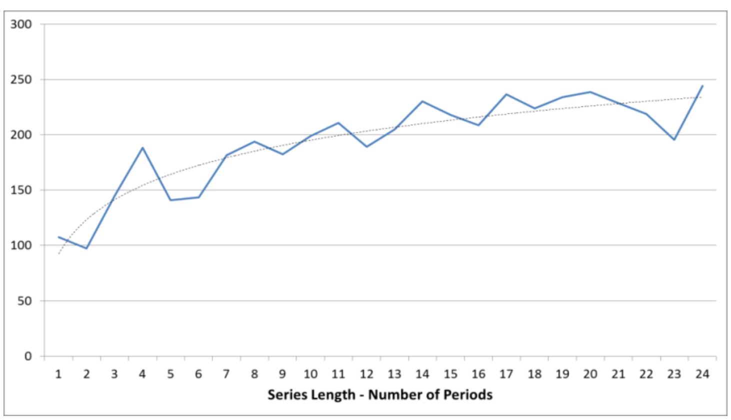
Examples Of Simulated Time Series



No Trend



Linear Trend
a+bt



Nonlinear Trend
a+bLn(t)

APPENDIX 3

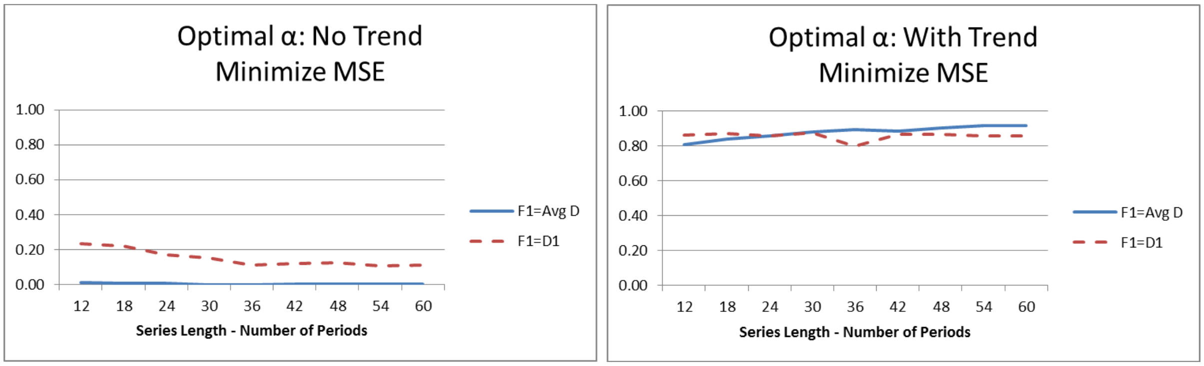


Exhibit 1

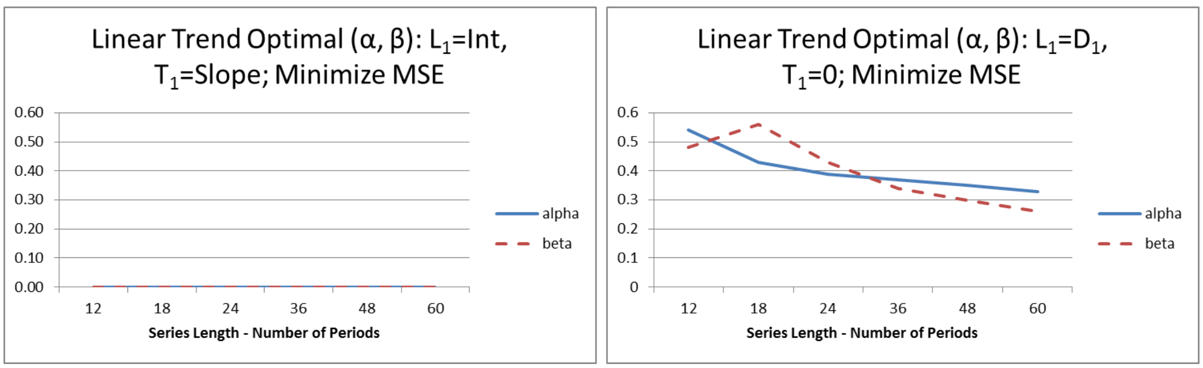


Exhibit 2

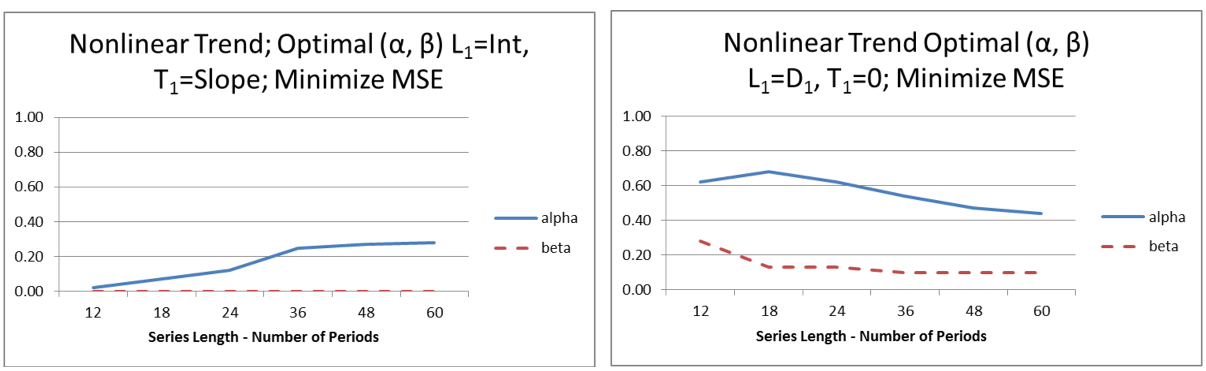


Exhibit 3