JOURNAL OF APPLIED MATHEMATICS AND DECISION SCIENCES, 8(2), 87–105
Copyright© 2004, Lawrence Erlbaum Associates, Inc.

Comparison of Efficient Seasonal Indexes

PETER T. ITTIG†
Management Science and Information Systems Department, University of Massachusetts, Boston, MA 02125-3393

Abstract. Estimates of a seasonal index in the standard manner (from a moving average) introduce systematic error into the seasonal estimates if a trend is present. This paper shows that a logarithmic modification of the standard moving average procedure will cause it to be consistent with a trend and is an efficient alternative. This paper also compares several other efficient seasonal indexing procedures appropriate for routine business applications and shows some numerical results. The results indicate that it is possible to achieve an improvement in the precision of the seasonal index, in the seasonally adjusted data and in forecasts based upon this data, by considering logarithmic alternatives to standard seasonal indexing procedures. This improvement may be accomplished without a substantial increase in complexity or in the associated computational burden. The opportunities for improvement are shown to be greatest when the data contain substantial trend and seasonal aspects and when the trend has a percentage form. Some suggestions for forecasters are offered.

Keywords: Seasonal Index, Forecasting.

1. Introduction

Seasonality is frequently an important element in forecasting for basic decision processes in business regarding production, inventory control, personnel, marketing and finance. It is common to deseasonalize the data before applying forecasting methods and then to reseasonalize the forecasts. Accuracy in a seasonal index can be very important in some applications. Errors in measuring the seasonal pattern translate into errors in forecasts and decisions made on the basis of those forecasts. Business forecasting applications often require highly efficient methods in order to handle very large numbers of inventory items and frequent updates. For example, a supermarket may have over 10,000 items with different product codes. The differences in methods are particularly important when there is a substantial trend and a substantial seasonal effect, as is common in business applications, and when a short data series is used, also common in business applications. Forecasters in business applications are typically reluctant to

† Requests for reprints should be sent to Peter T. Ittig, Management Science and Information Systems Department, College of Management, University of Massachusetts, 100 Morrissey Blvd, Boston, MA 02125-3393.

use more than 2 or 3 years of historical data, fearing that demand is changing too quickly for older data to be relevant. Both a short data series and a need for highly efficient methods (for large numbers of items) limit the use of complex procedures, such as seasonal ARIMA models, Census decomposition and nonlinear programming models. Complex procedures present additional difficulties when a forecaster is asked to explain the methods used.

Gardner (1986) noted that "Depending on customer service objectives, safety stocks are typically set in the range of two to three times the mean absolute deviation (MAD) of the forecast errors. If we think about the MAD in dollar terms, for every dollar it is reduced, safety stocks are automatically reduced by two to three dollars. That is, it takes less safety stock to maintain the same level of customer service." Wemmerlov (1989) commented that "The existence of forecast errors radically affects the behavior of the lot-sizing procedures... For example, forecast errors not only lead to stockouts, they also induce larger inventories."

In situations in which an accurate seasonal index is required for business applications there are advantages in considering alternatives to the most commonly used seasonal index procedures. Currently, if a seasonal index is needed for business applications, it is ordinarily obtained by *sequentially* separating trend and seasonal components of the time series. The most popular and standard way to do this is by first estimating the seasonal effect from a ratio of data to a centered moving average (CMA). This procedure is sometimes referred to as classical decomposition and is also the basis for the seasonal adjustment procedure of the US Census Bureau, referred to as Census II and experimental variations (X-11, X-12). For a discussion of this procedure see, for example, Makridakis et al. (1998, p. 109) or the latest Census Bureau reference manual (1999). An earlier paper (Ittig, 1997) showed that estimates of seasonals from a centered moving average introduce systematic error into the seasonal estimates if a trend is present. The standard procedure is only a heuristic and is *not* consistent with a trend. An alternate procedure that is sometimes used (for example, in the Minitab statistics package) is to first estimate a linear trend from the raw data and then de-trend the data. It is currently not well known that both of these methods introduce systematic error into the seasonal estimates if a trend is present and they incorrectly separate trend and seasonal influences, as will be discussed subsequently. The earlier paper (Ittig, 1997) also presented a method based upon a logarithmic regression that is often superior, *has a theoretical justification*, and *simultaneously* separates trend and seasonal influences while smoothing noise. It was further shown in that paper that errors in the seasonal index will propagate through to forecasts that are based upon seasonally adjusted data, even if the trend is correctly estimated. Of course, it is more difficult to accurately project the trend with bad seasonals.

It is relatively well known that a logarithmic transformation will sometimes aid forecasting. A few forecasters are aware that a seasonal index may be calculated from logged data (data that has been subjected to a logarithmic transformation). However, it is *not generally known* that a centered moving average applied to logged data allows the calculation of a seasonal index that is *consistent* with a trend. The existing forecasting literature does not describe this important property and as a consequence this approach is not popular. This property of a centered moving average of the logged data is presented and explored in this paper. An implementation procedure is presented that permits a seasonal index to be obtained in a simple manner that is related to the Ittig (1997) procedure, but a comparison of the two procedures displays a compromise between tracking and smoothing characteristics that is related to familiar issues in other areas of forecasting.

2. Note on Census Decomposition

A ratio of data to a centered moving average (CMA) is the basis for the seasonal adjustment procedure originally developed by the US Census Bureau, referred to as Census II and variations. The procedure contains adjustments that seek to separate the seasonal effect from the business cycle as well as the trend. The procedure is computationally intensive and difficult to understand, but is relatively popular with economists and is used to produce seasonally adjusted versions of some US Government data series. As with the standard CMA procedure, seasonals are not produced for the latest half-year as a consequence of computing centered annual moving averages. Recent versions of the Census method attempt to fix this defect by projecting the missing numbers using an ARIMA procedure.

Another disadvantage of the Census procedure is that it has become increasingly complex and more obscure as it has aged. Hylleberg & Pagan (1997, p. 329) commented that “Unfortunately, the way in which the separation of factors is accomplished has become increasingly a ‘black box’ and this lack of transparency has raised doubts about the procedure whenever odd movements in the series have been encountered.” ARIMA models share this lack of transparency. Recent versions of the Census method have also added an option to calculate the seasonals from logged data.

2.1. Decomposition models

In business forecasting it is common to view the data as a product of several components, including a multiplicative seasonal component, as shown in equation (1). The form shown is that of Makridakis, Wheelwright & Hyndman (1998, p.106) where the sales or demand data are Y, the seasonal component is S, the trend is T and the noise or random element is E, all at time period t. The multiplicative seasonal index S is percentage based and is defined to average 1.0 (or 100%). As noted by Makridakis and Wheelwright (1989, p.96), "practically all series in the economic and business domains consist of seasonality and cycle, which are proportional to the trend, and hence the multiplicative model is appropriate." In this paper, a multiplicative decomposition model is used, as in equation (1). This is also the model assumed in using the standard seasonal index from a centered moving average, which produces a multiplicative seasonal index.

$$Y_t = S_t * T_t * E_t \tag{1}$$

If the trend is linear with slope B and Y-intercept A, then $T_t = A + B * t$, and the data may be represented as shown below. Gardner (1987, p.175) noted that "A linear trend is the most popular assumption in practice."

$$Y_t = S_t \ * (A \ + B \ * t) \ * E_t \tag{2}$$

If the trend has a percentage form, rather than a linear form, then it may be represented by an exponential expression of the form, $T_t = P * (1 + i)^t$, where the parameter i is the growth rate and P is a constant. The combined decomposition model is shown as equation (3). This is the model used to obtain a seasonal index from a simultaneous decomposition in the Ittig (1997) paper.

$$Y_t = S_t * (P * (1 + i)^t) * E_t \tag{3}$$

An additive decomposition model, as shown in equation (4), is sometimes used by statisticians, but is not popular in business applications. The Ittig (1997) procedure has the effect of converting the multiplicative form of (1) to the additive form of (4) by the use of logarithms.

$$Y_t = S_t + T_t + E_t \tag{4}$$

3. Intel Example

In order to demonstrate that significant differences may arise from different methods of seasonal adjustment in a business application, an example is shown below for the Intel Corporation. Intel (INTC) is a large manufacturer of semiconductors. Quarterly revenues for Intel are shown in Table 1 for a four-year period (1996-1999). A seasonal index was generated from the first three years of data (1996-1998) and this was used to forecast revenues for each quarter of the fourth year (1999). Seasonals from several methods are shown in Table 2 for the first three years of data together with the percentage differences in comparison with the Logarithmic Regression procedure. The methods used are the Simultaneous (Logarithmic Regression) method from Ittig (1997), the standard procedure from a Centered Moving Average (Standard CMA) and a Logarithmic Centered Moving Average procedure (Log CMA) that will be described subsequently. The formulas for the Logarithmic Regression procedure are shown in the appendix. Note that the differences are substantial for some periods, as large as about 2%. Differences may be even larger in a situation in which there are particularly small seasonal values for some periods.

The differences in the seasonal index influence forecasts based upon these seasonals. To demonstrate this, forecasts were calculated for the fourth year from seasonally adjusted data for the first three years. The forecasting method used is a simple extrapolation of the trend from linear regression, based upon seasonally adjusted data. The forecasts were then reseasonalized in a conventional manner. Forecasts and forecast accuracy are shown in Table 3 using the correct revenues for the fourth year. This is true forecast accuracy. The differences are significant for some periods, as large as about 2% or about $120 million. The difference in the errors generally favors the logarithmic procedures. It may be shown that the Log Regression method also provides the best accuracy of fit to the historical data set.

Table 1. Revenues for Intel Corp. (millions of US dollars).

	1996	1997	1998	1999
Q1	4644	6448	6001	7103
Q2	4621	5960	5927	6746
Q3	5142	6155	6731	7328
Q4	6440	6507	7614	8212

One issue relates to whether the differences in methods are sufficiently great to be of interest even though the average errors of Standard CMA

Table 2. Seasonals for Intel for 3 Years of Data (1996-1998) and Differences (%).

	Q1	Q2	Q3	Q4	MAPD
Log Regression	0.988	0.930	0.986	1.096	
Standard CMA	1.010	0.938	0.969	1.082	
% difference	2.19%	0.92%	-1.71%	-1.21%	1.51%
Log CMA	1.008	0.937	0.972	1.083	
% difference	1.97%	0.81%	-1.48%	-1.13%	1.35%

MAPD= Mean Absolute Percentage Difference in comparison with Log Regression method.

procedure are proportionately much larger than those of the logarithmic procedures. If differences of this scale do not matter to the user, then an argument might be made for using familiar procedures even though some accuracy is being discarded. Another issue concerns accuracy for particular periods, regardless of the average accuracy indicated by MAPE. Accuracy is particularly important for periods with a seasonal value of less than one, since an error of .01 in such a value will produce an error of *more than* 1% in seasonally adjusted sales. This is because the adjustment procedure involves *dividing* by the seasonal index. The smaller the true seasonal value, the greater the magnification of errors in the seasonal value when adjusted sales are computed. Additional issues concern performance in smoothing noise and in tracking changing trend/seasonal components. A combination of these effects may produce significant differences in the seasonal index from different methods and in the resulting forecasts, as is seen in the Intel example.

Table 3. Forecasts and Forecast Accuracy for Intel fourth year (1999).

	Q1	Q2	Q3	Q4	MAPE	MAD	MSE	Bias
Standard CMA	7238.0	6889.2	7291.8	8334.6				
% error	1.90%	2.12%	0.49%	1.49%	1.50%	109.2	13765	91.1
Log Regression	7030.6	6769.5	7350.8	8352.6				
% error	1.02%	0.35%	0.31%	1.71%	0.85%	64.8	6523.4	28.6
Log CMA	7168.8	6824.4	7242	8258.3				
% error	0.93%	1.16%	1.17%	0.56%	0.96%	69.1	5005.2	26.1
Correct Values	7103	6746	7328	8212				

MAPE= Mean Absolute Percentage Error MAD= Mean Absolute Deviation MSE= Mean Squared Error

4. A Modified Method From a Centered Moving Average

As described earlier, it is not generally known that a centered moving average applied to logged data (data that has been subjected to a logarithmic transformation) allows the calculation of a seasonal index that is *consistent* with a trend. The existing forecasting literature is silent on this important point, which is demonstrated below. It is also not generally known that estimates of a seasonal index in the standard manner (from a centered moving average of the raw data) introduce systematic error into the seasonal estimates if a trend is present. However, that point is discussed in another paper (Ittig, 1997). The effect of using logged data in a modified CMA procedure is similar to multiple runs of the simultaneous regression based procedure for a very small data set, but *without* the use of regression. This modification is consistent with a multiplicative seasonal and a multiplicative trend, as in the regression based procedure. The modified procedure is convenient and is very similar to the standard method. The procedure permits the calculation of a seasonal index without the use of regression and the procedure has a sound theoretical basis, while the standard CMA procedure does not. This procedure will provide better tracking of changes in seasonal and trend effects than the Logarithmic Regression method, but weaker smoothing of noise (as will be seen subsequently). This procedure is referred to subsequently as the Log CMA procedure. The modified procedure involves the following steps:

i) Calculate centered moving averages for the logarithms of the sales data rather than for the raw data.

ii) Calculate the *difference* or deviation between the logarithm data and the centered moving averages of the logarithms. This step is similar to the conventional calculation of *ratios* of data to centered moving averages to obtain preliminary seasonals.

iii) Smooth the differences for each quarter by averaging these. This step is similar to the smoothing of preliminary ratios in the standard method.

iv) Exponentiate (raise to the power of e, the Naperian base) the smooth deviations for each season. This step produces a preliminary seasonal index, which is proportional to the final seasonal index.

v) Normalize the seasonal index to cause the values to average 1.0 to obtain the final seasonal index in the same manner as in the standard method. For quarterly data the seasonals should have a sum of 4.0.

The modified procedure is demonstrated in Table 4 for three years of quarterly data. Note that, as with the standard method, 3 years of data produce only 2 sets of seasonal estimates.

Table 4. Calculation of Seasonals from Centered Moving Average of Logarithms (Log CMA method).

Time	Sales	LogSales	Annual MA	Centered MA	difference	smooth	preliminary (EXP)	normalize	
1	180	5.19							
2	132	4.88	4.45						
3	140	4.94	4.59	4.52	0.42				
4	16	2.77	4.72	4.66	-1.89			Seasonals	
5	324	5.78	4.84	4.78	1	1	2.72	1.8	Q1
6	220	5.39	4.94	4.89	0.51	0.51	1.66	1.1	Q2
7	220	5.39	5.03	4.98	0.41	0.42	1.52	1	Q3
8	24	3.18	5.11	5.07	-1.89	-1.89	0.15	0.1	Q4
9	468	6.15	5.19	5.15	1		Sum	Sum	
10	308	5.73	5.26	5.23	0.5		6.04	4	
11	300	5.7							
12	32	3.47							

Entries rounded to 2 decimal places.

4.1. Origin of parameter estimates for log CMA method

This section demonstrates that the Log CMA method is consistent with a percentage trend and develops the formulas for the seasonal estimates. Both the Log CMA method and the Log Regression method are based upon a model of the data that assumes that the trend and seasonal relationships are both multiplicative or percentage based and may be expressed as shown below, where Sj is the seasonal index and i is the average rate of growth or decline in Sales per period. The parameter P is a constant and t is a time index. If the data consist of quarters, then j runs from 1 to 4.

$$Sales(t, j) = S_j * P * (1 + i)^t \tag{5}$$

The logarithmic form of (5) is given below in equation (6).

$$log\ (Sales) = log\ (P) + t * log(1 + i) + log\ (S_j) \tag{6}$$

It is easy to show that the first annual moving average, MA_1 , of the terms in equation (6) is as below for quarterly data.

$$MA_1 = log(P) + 2.5 * log(1 + i) + (1/4)\Sigma log(S_j) \tag{7}$$

Thus, an annual moving average centered at time t, CMA(t), may be expressed as:

$$CMA(t) = log(P) + t * log(1 + i) + (1/4)\Sigma log(S_j). \tag{8}$$

The difference or deviation between the logarithm data (6) and the centered moving averages for the same time period (8) may be expressed by subtracting equation (8) from equation (6), as shown below.

$$deviation(t, j) = log(S_j) - (1/4)\Sigma log(S_j). \tag{9}$$

If the differences or deviations in expression (9) are raised to the power of e, the Naperian base, the resulting expression is given below (10). This corresponds to step iv) in the procedure for the Log CMA seasonals.

$$exp(deviation) = S_j/(\Pi S_j)^{0.25} \tag{10}$$

The denominator in equation (10) is the geometric mean of the seasonals. Since the *same factor* is applied to the denominator of each term, it is removed when the seasonals are normalized in the last step of the procedure.

The equations shown are for quarterly data. More generally, if there are L periods in a year or cycle (e.g., 12 months), then there are L seasonals in equations (5) and (6), and the fraction in equations (7), (8), (9) and the power in equation (10) should be 1/L rather than 1/4.

5. Comparisons Without Noise

The performance of a seasonal indexing procedure depends upon several factors including the scale and shape of the trend, the scale of the seasonal effect, the level of noise in the data set, the length of the data series and the stability of the trend and seasonal effects. The interaction of these factors is complex. In order to explore aspects of the problem, a seasonal index was calculated from several different methods from constructed test data sets. The first cases are tests on data with no noise. Note that tests without noise favor CMA based procedures since these procedures provide only weak smoothing of noise and do not smooth noise at all when only two years of data are available. Data sets of varying lengths with both large and moderate trend/seasonal components are treated. In all of the tables for this section the errors reflect only systematic errors in the method of seasonal adjustment, since no noise is present in the data and the trend/seasonal pattern is stable. The data for the first case were constructed from a linear trend that increments by 20 units per period (from a base of 100), and from a multiplicative seasonal effect that has values of 1.8, 1.1, 1.0, 0.1 for the four quarters. The correct trend line is: *Yt = 80+20*t.* This data set is taken from the linear test case used in Ittig (1997). This data set contains large seasonal and trend components and represents a difficult test for a seasonal indexing procedure. The resulting seasonals and an error calculation are shown in Tables 5, 6 and 7 . The errors are reported as Mean Absolute Percentage Errors (MAPE) measured from the true known seasonals. In addition to the methods used previously, results are also shown for a de-trend procedure in Table 5. The Standard CMA and de-trend procedures both attempt to sequentially separate trend and seasonal effects and introduce systematic error into the estimates of the seasonals. The systematic error from the Standard CMA procedure was discussed in Ittig (1997). The systematic error from a de-trend procedure arises from fitting a regression line to the raw data and thereby incorrectly treating the seasonal effect as noise. Table 7 shows the errors resulting from extending the data set from 2 to 6 years in length. Note that the procedures based upon a centered moving average gradually improve with more data, in the absence of noise, due to the averaging of subsequent estimates, while the Logarithmic Regression procedure does not improve. Also, note that the only reason that there is any error at all for the logarithmic procedures in this case is that a linear trend is present in the test data rather than the percentage (multiplicative) trend assumed by the logarithmic methods, as shown in equation (3). If the trend were percentage based rather than linear, the logarithmic procedures would track the seasonals perfectly, as is shown in Table 7. For Table 7 the trend was set to 20% rather than 20 units. Note that with a percentage trend the standard seasonals are not improved by averaging over more years of data in this case. One advantage of the Logarithmic Regression procedure is that it allows for standard statistical tests, as may be seen in the results in the appendix Table A1. The large trend/seasonal case shows that systematic error alone in the standard CMA procedure contributes about 2 to 4 percent to average errors in the seasonals.

Table 5. Seasonals and Seasonals Accuracy for 2 Years of Quarterly Data with a Large Linear Trend/Seasonal.

	Q1	Q2	Q3	Q4	MAPE
De-trend	1.59	1.12	1.16	0.13	14.57%
Standard CMA	1.75	1.12	1.03	0.10	2.72%
Log Regression	1.78	1.11	1.01	0.10	0.84%
Log CMA	1.80	1.10	1.01	0.10	0.27%
Correct Values	1.8	1.1	1	0.1	0%

MAPE= Mean Absolute Percentage Error

Trend Line Equation from De-Trend Method: Yt = 162.571-1.23810 $* t$.

Table 6. Seasonal Accuracy (MAPE) for 2 to 6 years data with Large Linear Trend/Seasonal and No Noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	2.72	2.27	1.97	1.76	1.66
Log Regression	0.84	0.89	0.93	0.96	0.98
Log CMA	0.27	0.17	0.13	0.10	0.08

MAPE= Mean Absolute Percentage Error

Table 7. Seasonal Accuracy (MAPE) for 2 to 6 years data with Large Percentage Trend/Seasonal and No Noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	4.10	4.10	4.10	4.10	4.10
Log Regression	0	0	0	0	0
Log CMA	0	0	0	0	0

MAPE= Mean Absolute Percentage Error

The errors are sensitive to the size of the trend and seasonal effects. This may be seen in the next case (Tables 8-10) where a data set was used with more moderate seasonal and trend components. The trend was set to 5 units per period (from a base of 100) and the seasonals were changed to more moderate values of 1.3, 1.0, 0.9, 0.8, as may be found in retailing applications. The correct trend line is Yt = 95+5*t. Once again, if the trend was percentage based rather than linear, the logarithmic procedures would track the seasonals perfectly, as is shown in Table 10. For Table 10 the trend was set to 5% rather than 5 units and the moderate seasonals were used (1.3, 1.0, 0.9, 0.8). There is still no noise present in the data. The moderate trend/seasonal case shows that systematic error alone in the standard CMA procedure contributes about a quarter to a third of one percent to average errors in the seasonals.

Note that with a large linear trend in Table 6, there is a small, but noticeable, increase in the errors for the Logarithmic Regression method as more years of data are added, although the table shows that the Logarithmic Regression method is still superior to the Standard CMA method for *every* time period. In Table 9, with a smaller linear trend, this effect is not apparent but still present. This effect is due to the cumulative error associated with the difference between the linear trend in the data and the assumed percentage trend in the Logarithmic Regression procedure. For this reason, and because the trend and seasonal effects may change over time, there may be advantages in applying the Logarithmic Regression procedure to only a portion of the data at one time. The choice involves a tradeoff between smoothing and tracking that has parallels with the choice of the length of a moving average or the choice of a smoothing parameter in exponential smoothing. The Log CMA procedure has a similar effect to multiple runs of the Log Regression procedure for a very small data set.

Table 8. Seasonals and Seasonals Accuracy (MAPE) for 2 Years of Data with Moderate Trend and Seasonals.

	Q1	Q2	Q3	Q4	MAPE
De-trend	1.236	0.989	0.924	0.852	3.80%
Standard CMA	1.297	1.004	0.902	0.797	0.32%
Log Regression	1.299	1.001	0.901	0.800	0.09%
Log CMA	1.300	1.00	0.900	0.800	0.01%
Correct Values	1.3	1	0.9	0.8	0%

Trend Line Equation from De-Trend Method: Yt = 113.9286+ 0.571429* t

Table 9. Seasonal Accuracy (MAPE) for 2 to 6 years data with Moderate Linear Trend/Seasonal and No Noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	0.32	0.29	0.28	0.26	0.25
Log Regression	0.09	0.08	0.08	0.08	0.09
Log CMA	0.01	0.01	0.01	0.01	0.01

Table 10. Seasonal Accuracy (MAPE) for 2 to 6 years data with Moderate Percentage Trend/Seasonal and No Noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	0.36%	0.36%	0.36%	0.36%	0.36%
Log Regression	0	0	0	0	0
Log CMA	0	0	0	0	0

6. Tests With Simulated Noisy Data

In order to test the performance of seasonal indexing procedures on noisy data, simulations were run to generate data with the same trend and seasonal characteristics as shown previously, but with an additional noise factor. A multiplicative noise factor was introduced (in the manner indicated by equations 2 and 3) by multiplying each data point by a random number from a Normal distribution with a mean of 1.0 and a 2% standard deviation

(0.02). About two-thirds of the errors introduced in this manner should be less than 2%. Alternately, about half of the errors introduced should be less than 1.4% and half greater. The MAPE accuracy of the resulting seasonals is reported below in Table 11 for data with a large linear trend and seasonal. The data points for Table 11 are the same as those used to generate Tables 5 and 6, with the addition of noise factors. Each entry in the table represents an average of 10 simulations. With a large linear trend the results show that the errors from the standard CMA procedure are about 1% to 1.5% worse than those for the logarithmic procedures, with the larger differences for shorter data sets.

Table 12 shows the same information for data with a large *percentage* (multiplicative) trend rather than a linear trend. For the results in Table 12, the trend was changed from 20 units to 20%. In this case, with a large percentage trend, the errors from the standard CMA procedure are much worse, about 3% to 3.5% worse than those for the logarithmic procedures, with the larger differences for *longer* data sets.

Note that the resulting seasonals from the logarithmic procedures are consistently better whether the trend is linear or percentage based. Also note that, in both Tables 11 and 12 the MAPE accuracy for the logarithmic regression gradually *improves*. This differs from the earlier demonstration with no noise (Table 6). In this case the superior smoothing of noise associated with regression overcomes a portion of the cumulative errors in the trend. In Table 11, the CMA Log procedure is seen to be superior to the Log Regression procedure when more than 2 years of data are used due to superior tracking of the linear trend. The seasonals from the standard CMA procedure are much worse when the trend is percentage based, resulting in a more dramatic difference in Table 12. Table 12 also demonstrates the superior smoothing ability of the Log Regression method.

Table 11. Seasonal Accuracy (MAPE) for 2-6 years simulated noisy data with Large Linear trend/Seasonal and 2% noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	2.95	2.33	2.05	1.88	1.74
Log Regression	1.30	1.10	0.98	0.94	0.86
Log CMA	1.40	0.88	0.81	0.78	0.62

Each entry is an average of 10 simulations.

Table 12. Seasonal Accuracy (MAPE) for 2-6 years simulated noisy data with Large Percentage trend and 2% noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	4.28	4.16	4.16	4.19	4.14
Log Regression	1.13	0.82	0.80	0.67	0.56
Log CMA	1.46	0.87	0.83	0.79	0.63

Each entry is an average of 10 simulations.

6.1. Tests with moderate trend and seasonal effects plus noise

The errors in the standard CMA procedure are related to the magnitude of the trend and seasonal effects as well as the level of noise. Additional tests were run on data containing moderate trend and seasonal effects in order to demonstrate the scale of effects to be expected in such an application. The linear trend was reduced to 5 units per period (from a base of 100) and the seasonals were reduced to 1.3, 1.0, 0.9, 0.8, as may be found in retailing applications. The resulting test data were used previously. Again, noise was introduced by multiplying each data point by a random number from a Normal distribution with a mean of 1.0 and a 2% standard deviation (0.02). Each entry in Table 13 represents an average of 10 simulations. Note that the MAPE accuracy for the logarithmic regression gradually improves as the data are extended from 2 to 6 years. Once again, the superior smoothing of noise associated with regression overcomes a portion of the cumulative errors in the trend. The logarithmic regression procedure is seen to be superior to the *both* of the CMA based procedures for *every time period* in this case, due to superior smoothing capability, with a reduction of about one quarter percent in error. The Log CMA procedure provides only a small advantage over the standard CMA procedure in this circumstance (none at 2 years), though the advantage of the CMA Log procedure over the standard CMA procedure is seen to grow as more data points are added. Both of the CMA based methods fail to diminish noise when only 2 years of data are available. The Log Regression procedure provides some smoothing of noise if more than 5 quarters are available (one year plus one period).

Table 14 shows the same information for data with a moderate *percentage* trend rather than a linear trend. For the results in Table 14, the trend was changed from 5 units to 5%. Smoothing of noise in the CMA Log procedure is comparable to that of the standard CMA procedure and the accuracy of both methods are seen to improve as more data points are added. The advantage of the CMA Log procedure over the standard CMA procedure is seen to grow as more data points are added beyond 2 years due to superior tracking of the trend. The logarithmic regression procedure is again seen to be superior to the *both* of the CMA based procedures for *every time period* due to superior smoothing capability, with a reduction of about one quarter percent in error.

Table 13. Seasonal Accuracy (MAPE) for 2-6 years simulated noisy data with Moderate Linear Trend/Seasonal and 2% noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	1.30	0.93	0.87	0.85	0.77
Log Regression	1.05	0.81	0.77	0.63	0.53
Log CMA	1.33	0.87	0.80	0.77	0.61

Each entry is an average of 10 simulations.

Table 14. Seasonal Accuracy (MAPE) for 2-6 years simulated noisy data with Moderate Percentage Trend and 2% noise.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	1.32	0.96	0.91	0.91	0.83
Log Regression	1.07	0.81	0.78	0.64	0.56
Log CMA	1.33	0.87	0.80	0.77	0.61

MAPE= Mean Absolute Percentage Error
Each entry is an average of 10 simulations.

Additional tests were run on data containing no trend and the moderate seasonal component. The trend was reduced to zero units per period (from a base of 100) and the moderate seasonals were used (1.3, 1.0, 0.9, 0.8). Again, noise was introduced by multiplying each data point by a random number from a Normal distribution with a mean of 1.0 and a 2% standard deviation (0.02). Each entry in the table represents an average of 10 simulations. The results are shown in Table 15 for 2 to 6 years of noisy data. The standard CMA procedure is consistent if there is no trend present in the data and might be expected to perform well in this test. The logarithmic procedures attempt to separate out a trend that does not exist in this data set. These tests demonstrate the ability of several methods to smooth noise in this situation.

Note that the simultaneous logarithmic regression procedure offers superior performance to the CMA procedures *for every time period* due to superior smoothing of noise. The difference is greatest (about one quarter percent) with two years of data. The performance of the Log CMA procedure is comparable to that of the standard CMA procedure in smoothing noise, and thus offers no significant advantage for a data set with no trend.

With no trend present in the noisy data, the trend term in the logarithmic regression procedure will generally be found to fail standard significance tests. In this case, conventional statistical practice would be to remove the trend term from the regression equation and to rerun the regression to force a zero trend. If this is done, the resulting seasonal index is improved a bit, as shown in Table 15 in the row entitled "Log Regression with NO Trend Term". The improvement is most noticeable with a very short data series where some patterns of noise may be misconstrued as indicating a trend.

Table 15. Seasonal Accuracy (MAPE) for 2-6 years simulated noisy data with no trend and moderate seasonal.

	2 years	3 years	4 years	5 years	6 years
Standard CMA	1.34	0.87	0.80	0.77	0.69
Log Regression	1.07	0.81	0.78	0.64	0.56
Log Regression with NO Trend Term	1.00	0.77	0.76	0.64	0.56
Log CMA	1.33	0.87	0.80	0.77	0.61

Each entry is an average of 10 simulations.

7. Conclusions

When a seasonal index is needed for business applications, particularly when a substantial trend is present in the data, it is possible to achieve better results in the precision of the seasonal index, in the seasonally adjusted data and in forecasts based upon this data, by considering alternatives to standard seasonal indexing procedures. The standard CMA seasonals represent a rule of thumb approximation that works best when there is no trend present and provides only weak smoothing of noise. When a trend is present and there are more than 2 years of data, the seasonals offered by the Log CMA method appear to generally provide a convenient and superior alternative to the standard CMA procedure. The method is no more difficult to use than the standard method when applied in spreadsheet software or statistical software. These seasonals also provide a seasonal index with a sound theoretical justification. A disadvantage of the Log CMA procedure is that the convenience gained by avoiding regression results in the loss of standard statistical tests and no better smoothing of noise than the standard CMA procedure. Both CMA methods provide weaker smoothing of noise than the Log Regression procedure. A further difference is that procedures based on a moving average provide no results for the latest half-year and no smoothing of noise at all unless more than two years of data are available. Both CMA procedures will track changes in trend and seasonal effects to some degree. The Log Regression procedure will work with less data, as little as one year plus one period, and provides efficient smoothing of noise. The logarithmic procedures provide substantially better results when the trend and seasonal components of the data are large and particularly when the trend has a percentage form.

Ultimately, the choice of a seasonal index depends upon the balancing of several matters. One issue relates to convenience versus precision. The standard seasonals from a centered moving average are convenient in that they are familiar, moderately intuitive and moderately accurate under some circumstances. However, the standard seasonals are obtained from a heuristic that introduces systematic error into the seasonal estimates when a trend is present and provides only weak smoothing of noise. The manipulation of logarithms is slightly less convenient than manipulation of the raw data. A moving average is probably more convenient than one or more regression runs. However, the availability of fast regression routines substantially diminishes the difficulty of using regression in business applications. For example, the "Trend", "Linest" and "Logest" functions in Excel allow for the convenient use of regression in spreadsheet applications. Additional issues concern performance in the tradeoff between the smoothing of noise and the tracking of changes in trend or seasonals for the data being examined. Generally, the Log Regression procedure will provide better smoothing of noise than sequential procedures but may not track changes in trend or seasonals as well unless the range is broken into segments. For a short data series, the Log Regression procedure appears to provide superior results. The Log CMA procedure represents a compromise between convenience, simultaneous decomposition and smoothing of noise.

References

1. Bureau of the Census. *X-12-ARIMA Reference Manual*, Washington, DC 20233, 1999.
2. E. Gardner. A comment. *Interfaces*, 16: 106-107, 1986.
3. E. Gardner. Smoothing methods for short-term planning and control. Chapter 11 in S. Makridakis, & S. Wheelwright (Eds.), The Handbook of Forecasting, Wiley, NY, 1987.
4. S. Hylleberg and A. Pagan. Seasonal integration and the evolving seasonals model, *International Journal of Forecasting*, 13: 329-340, 1997.
5. P. Ittig. A Seasonal Index for Business, *Decision Sciences*, 28: 335-355, 1997.

6. S. Makridakis, & S. Wheelwright. *Forecasting Methods for Management*, 5th ed. Wiley, NY, 1989.
7. S. Makridakis, S. Wheelwright and R. Hyndman. *Forecasting: Methods and Applications*, 3^{rd} ed., Wiley, NY, 1998.
8. U. Wemmerlov. The behavior of lot-sizing procedures in the presence of forecast errors. *Journal of Operations Management*, 8: 37-45, 1989.

Appendix

Formulas for Logarithmic Regression Procedure

A simultaneous decomposition method was presented by Ittig (1997). This seasonal index is based upon a logarithmic regression using seasonal indicators and assumes a percentage (multiplicative) trend. This seasonal index simultaneously separates trend and seasonal effects while reducing noise through the use of regression. The method models the trend and seasonal relationship as below.

$Sales = S_j * P * (1 + i)^t$

Formulas for the seasonals on a quarterly basis are summarized below.

$$S_j = b'_j P, j = 1, 2, 3, 4 \tag{A.1}$$

$$P = (b'_1 + b'_2 + b'_3 + b'_4)/4 \tag{A.2}$$

$$b'_j = exp(b_j) \tag{A.3}$$

The regression coefficients b_j are obtained from the regression model,

$$log\ (Sales) = b_1 I_1 + b_2 I_2 + b_3 I_3 + b_4 I_4 + b_5 t. \tag{A.4}$$

In this model the I_j are seasonal indicators (0 or 1), and t is a time index. The regression model has no constant (i.e. the constant is forced to be zero). A spreadsheet implementation of this procedure is shown in Table A1. In the Excel spreadsheet it is possible to calculate the coefficients b'_j directly, without first calculating logarithms, by using the "LOGEST" spreadsheet function with the constant set to one.

Table A.1. Calculation of Seasonals from Simultaneous Decomposition by Logarithmic Regression.

Sales	Q1	Q2	Q3	Q4	Time	LogSales
130	1	0	0	0	1	4.87
105	0	1	0	0	2	4.65
99	0	0	1	0	3	4.60
92	0	0	0	1	4	4.52
156	1	0	0	0	5	5.05
125	0	1	0	0	6	4.83
117	0	0	1	0	7	4.76
108	0	0	0	1	8	4.68
182	1	0	0	0	9	5.20
145	0	1	0	0	10	4.98
135	0	0	1	0	11	4.91
124	0	0	0	1	12	4.82

	Coefficients	sum b_j'	sum b_j' /4	
	b_j'	(j= 1-4)	P	Seasonals
Q1	126.77	390.51	97.63	1.3
Q2	97.7			1
Q3	87.96			0.9
Q4	78.08			0.8
T	1.04	0.04042		
	(1+i)	i		

R Square 0.998 Standard Error 0.011

	Coefficients b_j	Standard Error	t	p
Intercept	0			
Q1	4.84	0.008	614.22	0
Q2	4.58	0.009	539.03	0
Q3	4.48	0.009	487.89	0
Q4	4.36	0.010	440.26	0
Time	0.04	0.001	41.34	0

Special Issue on Modeling Experimental Nonlinear Dynamics and Chaotic Scenarios

Call for Papers

Thinking about nonlinearity in engineering areas, up to the 70s, was focused on intentionally built nonlinear parts in order to improve the operational characteristics of a device or system. Keying, saturation, hysteretic phenomena, and dead zones were added to existing devices increasing their behavior diversity and precision. In this context, an intrinsic nonlinearity was treated just as a linear approximation, around equilibrium points.

Inspired on the rediscovering of the richness of nonlinear and chaotic phenomena, engineers started using analytical tools from "Qualitative Theory of Differential Equations," allowing more precise analysis and synthesis, in order to produce new vital products and services. Bifurcation theory, dynamical systems and chaos started to be part of the mandatory set of tools for design engineers.

This proposed special edition of the *Mathematical Problems in Engineering* aims to provide a picture of the importance of the bifurcation theory, relating it with nonlinear and chaotic dynamics for natural and engineered systems. Ideas of how this dynamics can be captured through precisely tailored real and numerical experiments and understanding by the combination of specific tools that associate dynamical system theory and geometric tools in a very clever, sophisticated, and at the same time simple and unique analytical environment are the subject of this issue, allowing new methods to design high-precision devices and equipment.

Authors should follow the Mathematical Problems in Engineering manuscript format described at http://www.hindawi.com/journals/mpe/. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	February 1, 2009
First Round of Reviews	May 1, 2009
Publication Date	August 1, 2009

Guest Editors

José Roberto Castilho Piqueira, Telecommunication and Control Engineering Department, Polytechnic School, The University of São Paulo, 05508-970 São Paulo, Brazil; piqueira@lac.usp.br

Elbert E. Neher Macau, Laboratório Associado de Matemática Aplicada e Computação (LAC), Instituto Nacional de Pesquisas Espaciais (INPE), São Josè dos Campos, 12227-010 São Paulo, Brazil ; elbert@lac.inpe.br

Celso Grebogi, Department of Physics, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK; grebogi@abdn.ac.uk