VALUATION

Three Short-term Econometric Sales Forecasting Models Part I: Time Series Regression Model

By Mark G. Filler, CPA/ABV, CBA, AM, CVA; and James A. DiGabriele, PhD/DPS, CPA/ABV/CFF

Very often in the arena of business interruption claims and commercial damages, whether arising from breach of contract or tort, the valuation analyst has the need for a reliable short-term sales forecasting tool. Unlike sales forecasting for business valuation purposes, where time units are measured in years, it is just next year's sales or a period of up to five–ten years until sales stabilize, commercial damages forecasts very often are measured in months, or even weeks.

To accommodate this need, a body of knowledge has been developed that presents itself in two econometric formats—explanatory models and time series models. This article provides a step-by-step walk-through for building a time series regression model for a specific need using Excel. It is designed so that the reader can develop his or her own follow-along spreadsheet in order to learn by doing. We believe that valuation analysts will find it useful since most discussions of this topic are very abstract and unusable.

EXPLANATORY MODELS: WHAT AND WHY

Explanatory models assume that one or more independent variables exhibit an explanatory relationship with the variable to be forecasted. For example, GDP is a function of monetary and fiscal policies, inflation, capital spending, imports, exports, and error. Notice that the relationship is not exact. There will always be changes in GDP that cannot be accounted for by the variables in the model, and thus some part of GDP changes will remain unpredictable. Therefore, we include the "error" term that represents random effects, beyond the variables in the model, which affect the GDP figures.

Explanatory models can be applied to many systems—a national economy, a company's market or a company's sales. The purpose of the explanatory model is to capture the form of the relationship and use it to forecast future values of the forecast variable. According to explanatory forecasting, any change in inputs will affect the output of the system in a predictable way, assuming the explanatory relationship has not changed.

Unlike explanatory forecasting, time series forecasting treats the system as a black box and makes no attempt to capture the factors affecting its behavior. Therefore, a prediction of the future is based on past values of a variable, but not on explanatory variables that may affect the system. The objective of such time series forecasting methods is to discover the pattern and/or trend in the historical data series and extrapolate that pattern and/or trend into the future.

There are three main reasons for wanting to treat a system as a black box. First, the system may not be understood, and even if it were understood it might be extremely difficult to measure the relationships assumed to govern its behavior. Second, the main concern may be only to predict what will happen and not know why it happens. And third, changes in a time series may be reasonable proxies for the entire cluster of explanatory variables that drive the system through repeated transaction patterns over time.

For example, if the only purpose were to forecast future values of GDP without concern as to why a certain level of GDP will be realized, a time series approach would be appropriate. It is known that the magnitude of GDP does not change drastically from one month to another or even from one year to the other. Thus the GDP of next month will depend upon the GDP of the previous month and possibly that of the months before. This makes the job of forecasting next month's GDP relatively easy since it requires no special input values as does explanatory forecasting for GDP. In fact, all we need to have is a reasonable amount of past GDP monthly history.

This series of articles will utilize the application of several econometric models. We will apply two time series models—a regression-based model and an autoregressive integrated moving average (ARIMA) model—and one explanatory model to an actual business interruption situation where a discount department store retailer (Rollie's Discount Department Store) suffered a fire loss during the months of November, December 2012 and January 2013. Our assignment was to determine the actual loss sustained by the insured for those months, which in turn, required us to predict what sales for those months would have been but for the casualty. We will demonstrate the applicability of the three models to the claim requirements by comparing and contrasting the forecast results for each, beginning with the regression-based time series model. Charts 6 through 11 do not include data from 2009 to 2011. To view that information please go to www.nacva.com/examiner/13-SO-charts.asp

INTRODUCTION TO THREE MODELS: REGRESSION-BASED TIME SERIES

Our historical data consists of a sequence of observations over time from November 2007 through October 2012, a period of sixty months. This sequence is called a time series, and the data set for Rollie's Discount Department Store is shown on Exhibit 1.

Some of the variation in a time series may be due to the variation in the number of days in each month. It is a good idea to adjust for this known source of variation to allow for study of other interesting features. Month length can have quite a large effect, since length can differ by as much as 10 percent ((31 – 28)/30). If this is not removed, it can affect seasonal patterns. As leap years come every four years [((365*3+366)/4) = 365.25], it is easily adjusted by multiplying each month's sales by: ((365.25/12)/no. of days for that month). For example, November 2001 would be adjusted as follows: *164,330 * ((365.25/12)/30) = $164,330 * 1.014583 = $166,726.*

The adjusted time series data set for Rollie's is shown on Exhibit 2 (on page 9). Also shown on Exhibit 2 is the percentage change from the prior year, indicating an upward trend but whose rate of change is diminishing over time.

EXHIBIT 1:
ROLLIE'S DISCOUNT DEPARTMENT STORE—ACTUAL MONTHLY SALES: NOVEMBER 2007–OCTOBER 2012

	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012
November	$164,330	$213,999	$260,667	$246,244	$235,708
December	264,400	291,119	311,513	305,520	308,677
January	84,906	82,849	93,205	93,430	96,997
February	75,390	82,485	88,736	86,226	88,157
March	75,328	87,838	90,085	97,369	101,207
April	77,102	94,337	109,497	113,309	111,176
May	101,519	132,548	141,435	131,180	136,851
June	114,315	132,788	145,444	150,194	170,377
July	119,700	141,325	147,984	160,488	168,906
August	130,284	158,056	158,522	171,851	184,999
September	105,470	119,961	124,592	134,631	152,787
October	143,399	172,229	174,212	194,255	178,025
Total	$1,456,143	$1,709,534	$1,845,892	$1,884,697	$1,933,867

EXHIBIT 2:
ROLLIE'S DISCOUNT DEPARTMENT STORE—ACTUAL MONTHLY SALES: NOVEMBER 2007–OCTOBER 2012

	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012
November	$166,726	$217,120	$264,468	$249,835	$239,145
December	259,602	285,837	305,861	299,976	303,076
January	83,365	81,346	91,514	91,735	95,237
February	79,127	89,666	96,461	93,732	92,527
March	73,961	86,244	88,450	95,602	99,371
April	78,226	95,713	111,094	114,961	112,797
May	99,677	130,143	138,869	128,800	134,368
June	115,982	134,724	147,565	152,384	172,862
July	117,528	138,761	145,299	157,576	165,841
August	127,920	155,188	155,646	168,733	181,642
September	107,008	121,710	126,409	136,594	155,015
October	140,797	169,104	171,051	190,730	174,795
Total	$1,449,921	$1,705,555	$1,842,686	$1,880,659	$1,926,676
% Change from Prior Year		17.6%	8.0%	2.1%	2.4%

VISUALIZING THE DATA: DEVELOPING TIME SERIES PLOTS

The single most important thing one can do when first exploring the data is to visualize the information via graphs. The basic features of the data, including patterns, trends, and unusual observations are most easily seen in graphic form. Additionally, graphs can suggest possible explanations for some of the variation in the data.

The type of data will determine which type of graph is the most appropriate. For time series, the most obvious form is a time plot in which data are plotted over time. For Rollie's, we graphed five-years of adjusted monthly sales data on Exhibit 3 (page 10). This time plot immediately reveals seasonal behavior, which is a regular repeating pattern in the data.

Along with seasonality, there is a systematic trend feature indicated by the fact that each year's succeeding data points are slightly higher than the year before. Trend is the long-term sweep or general direction of movement in a time series. It reflects the net influence of long-term factors that affect the time series in a fairly consistent and gradual way over time.

For seasonal time series data, it is often useful to produce a seasonal plot. Exhibit 4 (page 10) shows a seasonal plot of Rollie's adjusted sales. This graph consists of the data plotted against the individual "seasons" in which the data are observed (in this case a "season" is a month.) This is similar to a time plot except that the data from each season are overlapped. A seasonal plot enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified. In our case, there are no departures from the overall pattern, and we can clearly see that each succeeding year's sales are higher than the year before.

Our visual conclusion that there was a great deal of seasonality and a slight trend pattern is confirmed by an analysis

EXHIBIT 3: TIME PLOT

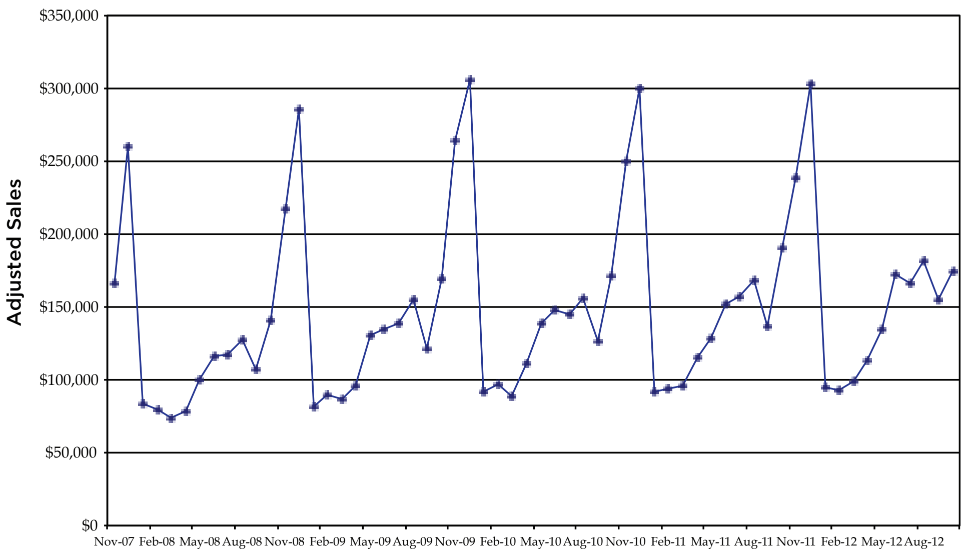
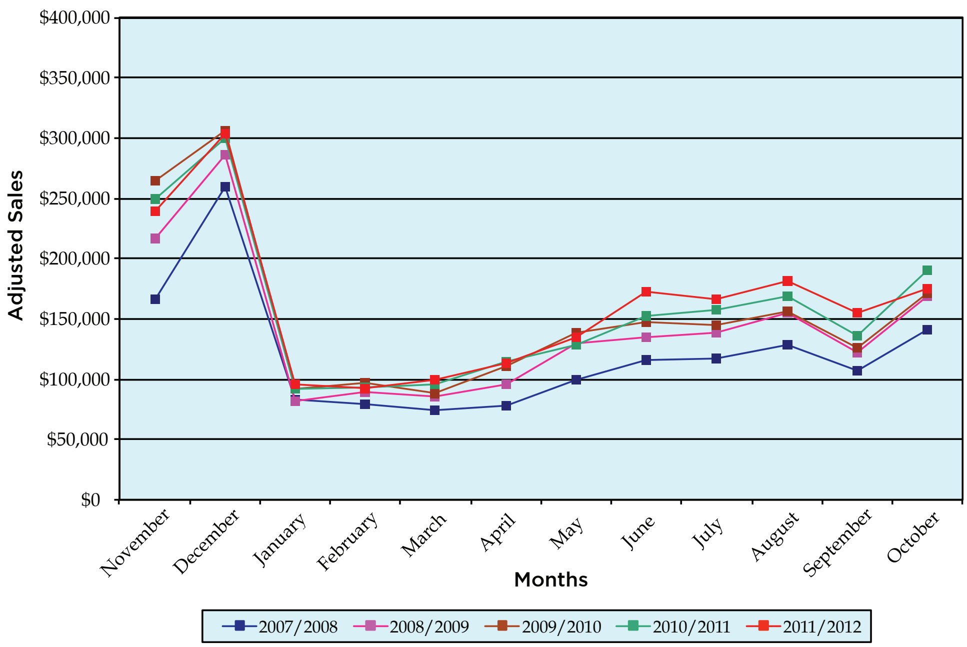


EXHIBIT 4: SEASONAL PLOT



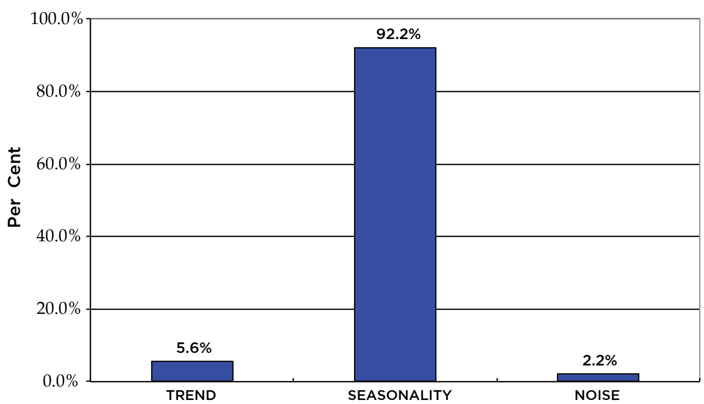
of variance (ANOVA), the results of which are shown on Exhibit 5 (page 11), which summarizes the degree of trend, seasonality, and noise. The ANOVA calculations are as follows: First, we computed the grand mean (the mean of all the data points). Then we computed the sum of squared differences between *each data point* and the grand mean (total variance). Next, we computed the trend variance: the sum of squared differences between the average value for *each year* and the grand mean. The ratio [(12*trend variance)/ (total variance)] is then defined as the proportion of variance due to trend.

The next step is to compute the seasonal variance: the sum of squared differences between the average value for *each month* and the grand mean. The ratio [(number of years of data*seasonal variance)/ (total variance)] is defined as the proportion of variance due to seasonality. Since the proportion must add up to 1.0, the proportion of noise is taken to be [1.0 – (proportion due to trend + proportion due to seasonality)]. The implication is that for the business interruption claim at hand we would need to account for a seasonal pattern as well as a trend pattern in our statistical model.

We next must choose a statistical model that will account for both seasonality and a trend pattern, albeit one that diminishes over time. Because we have a non-stationary time series, (since there is some upward trend in the data over time) forecasting techniques such as the moving average, weighted moving average, and exponential smoothing these techniques use some average of the previous values to forecast future values. They will consistently underestimate the predicted values. Therefore, we need to consider a technique that is appropriate for a non-stationary time series that involves an upward trend in the data over time coupled with a seasonal pattern.

Regression analysis is such a technique. We can build a regression model of a time series if data are available for the independent variables that account for the systematic movements of the time series. Even if the independent variables don't have a *causal* relationship

EXHIBIT 5: ANALYSIS OF VARIANCE - MONTHLY DATA



with the time series, some independent variables might have a *predictive* relationship with the time series. The behavior of a predictor variable might be correlated with that of the time series in a way that helps us forecast future values of the time series.

SELECTING INDEPENDENT VARIABLES

As mentioned earlier, trend is the long-term sweep or general direction of movement in a time series that reflects changes in the data over time. The mere passage of time does not cause the trend in the time series. However, like the consistent passage of time, the trend of a time series reflects the steady upward or downward movement in the general direction of the series. Thus, time itself might represent a predictor variable that could be useful in accounting for the trend in a time series.

SELECTING THE TREND LINE

If we were to superimpose a linear trend line on the time plot graph on Exhibit 3, we would see that while the line accounted for the upward trend in the data, the actual values would not be scattered randomly around the trend line. Since this scatter is a basic requirement for a good fitting regression-derived trend line, we can infer that a linear trend model might not be appropriate for this data.

As an alternative, a curved, or curvilinear, trend line will produce a better fit with the data, as we need to account for an upward trend that is diminishing over time. A model that will serve this purpose is a quadratic (or second-degree polynomial) model where time is the first degree and time squared is the initial second degree of the regression equation. Later we will see how we can modify the second degree to be any exponent that helps produce a better fitting model.

SELECTING AND CALCULATING SEASONAL INDICES

Now that we have determined how we will account for trend, we need to find an effective way of modeling seasonal effects in our time series. A simple way to do this is by use of "dummy", or indicator, variables for each of the months in a year. The downside to this method is that it uses up ten more degrees of freedom than necessary, hence we will choose another approach known as seasonal indices, which while more complex to develop, saves those precious degrees of freedom, which, *ceteris paribus*, allows for a smaller standard error of the estimate.

Seasonal indices reflect the average percentage by which observations in each "season" (month) differ from their projected trend values. For example, Rollie's sales in December are above the value predicted using a trend model, while January's sales fall below the value predicted using a trend model. Thus, if we can determine seasonal indices representing the average amount by which observations in a given month fall above or below the trend line, we can multiply our trend projections by these amounts and increase the accuracy of our forecasts.

Before we calculate seasonal indices for a quadratic trend model, we need to set up and validate the model. Exhibit 6 (page 13) shows the initial set up of the quadratic trend model for Rollie's sixty months of adjusted sales (years 2009–2011, rows 18–53 have been hidden for presentation purposes; to see these years (please go to the NACVA website www.nacva.com/examiner/13-SO-charts.asp to view these charts). If the reader wishes to replicate our computations in a follow-along spreadsheet, the formula for cell F4 is: *=TREND (($E$4:$E$63, $C$4:$D$63, C4:D4)*, which is then copied down to and through row 63.

We need to evaluate the accuracy of this model by seeing how well it explains past behavior of the time series variables. A visual, non-quantitative method, which we will introduce later, is to construct line plots that show the actual data versus the values predicted by the model. Several more formal quantitative measures of the accuracy of time series modeling techniques are *mean absolute deviation* (MAD), the *mean absolute percentage error* (MAPE), the *mean square error* (MSE), and the *root mean square error* (RMSE). We have chosen the MSE measure because it is somewhat easier to calculate then the other metrics. The formula for cell F65 is:*=SUMXMY2(F4:F63,E4:E63)/COUNT(F4:F63)*, and it computes an MSE of 3,549,850,132, which is an exceedingly large number as the trend line we have calculated only traverses the mid-point of the data set, and does not account for seasonal fluctuations. So, we need to include seasonal indices in the model.

The goal in developing seasonal indices is to determine the average percentage by which observations in each "season" (month) differ from the value predicted for them using the particular trend model selected. To accomplish this, in column G of Exhibit 7 (page 14) we calculate the ratio of each actual value (in column E) to its corresponding projected trend value shown (in column F) as: =E4/F4, and then copy this formula down to and through row 63. The results are shown on Exhibit 7.

The value in cell G4 indicates that the actual value in time period 1 was 135.6 percent of (or approximately 35.6 percent larger than) its estimated trend value. The value in cell G6 indicates that the actual value in time period 3 was 66.5 percent of (or approximately 33.5 percent smaller than) its estimated trend value. The remaining values in column G have similar interpretations.

On Exhibit 8 (page 15), in column K, we obtain the seasonal index for each month by computing the average of the values in column G on a month-by-month basis. For example, the seasonal index for month 11 equals the average of the cells in G4, G16, G28, G40, and G52. The seasonal index for month 12 equals the average of the values in cells G5, G17, G29, G41, and G53. Similar computations are required to calculate seasonal indices for months 1–10. We can use separate AVERAGE () functions for each month to compute these averages. However, for large data sets, such an approach would be tedious and prone to error. Thus, the averages shown in cells H4 through K5 are calculated as: *=SUMIF (B4$B$63, J4, $G$4:$G$63)/COUNTIF ($B$4:$B$63, J4)*, which is then copied down and through row 63.

The seasonal index for month 1 shown in cell K4 on Exhibit 8 indicates that, on average, the actual sales value in January of any given year will be 61.4 percent of (or 38.6 percent smaller than) the estimated trend value for the same time period. Similarly, the seasonal index for month 8 shown in cell K11 indicates that, on average, the actual sales value in August will be 105.7 percent of (or 5.7 percent larger than) the estimated trend value for the same period. The seasonal indices for all the other months have similar interpretations.

REFINING THE INITIAL QUADRATIC TREND MODEL

We can use the calculated seasonal indices to refine or adjust the original

	A	B	C	D	E	F
1	**EXHIBIT 6:** QUADRATIC TREND MODEL					
2			TIME	TIME	ADJUSTED	QUADRATIC
3	YEAR	MONTH	PERIOD	$PERIOD^2$	SALES	TREND
4	2007	11	1	1	166,726	122,962
5		12	2	4	259,602	124,195
6	2008	1	3	9	83,365	125,405
7		2	4	16	79,127	126,594
8		3	5	25	73,961	127,760
9		4	6	36	78,226	128,905
10		5	7	49	99,677	130,027
11		6	8	64	115,982	131,127
12		7	9	81	117,528	132,206
13		8	10	100	127,920	133,262
14		9	11	121	107,008	134,296
15		10	12	144	140,797	135,309
16		11	13	169	217,120	136,299
17		12	14	196	285,837	137,267
54	2012	1	51	2601	95,237	157,603
55		2	52	2704	92,527	157,734
56		3	53	2809	99,371	157,844
57		4	54	2916	112,797	157,931
58		5	55	3025	134,368	157,995
59		6	56	3136	172,862	158,038
60		7	57	3249	165,841	158,059
61		8	58	3364	181,642	158,058
62		9	59	3481	155,015	158,035
63		10	60	3600	174,795	157,990
64						
65					MSE	3,549,850,132

quadratic trend estimates shown on Exhibit 6. This is accomplished in column H of *Exhibit 9 (page 16)* as: *=F4*VLOOKUP (B4, $J$4:$K$15, 2)*, which is then copied down and through row 63. This formula takes the estimated quadratic trend value for each time period and multiplies it by the appropriate seasonal index for the month in which it occurs. The trend estimates for month one observations are multiplied by 61.4 percent; the trend estimates for month two are multiplied by 62.2 percent, and so on for months three through 12. Before we create a graph showing the actual sales data versus the seasonal forecast calculated in column H, we need to refine the seasonal indices. Cell H65 calculates MSE to be 89,444,067, an increase of 97.5 percent in accuracy from the pure quadratic trend model shown on Exhibit 6.

REFINING THE SEASONAL INDICES

While the approach for calculating seasonal indices illustrated in Exhibit 9 has considerable intuitive appeal, it is important to note that these seasonal adjustment factors are not necessarily optimal. We can refine the seasonal adjustment factors using Excel's Solver add-in to simultaneously determine the optimal values of the seasonal indices and the slope and intercept parameters of the quadratic trend model. On Exhibit 10 (page 17), cells K18, K19, and K20 on your worksheet (K54, K55, and K56 on the truncated demonstration sheet) are used to represent, respectively, the estimated values of the intercept, the time coefficient and the $time^2$ coefficient in the quadratic trend model. Label the cells as indicated on Exhibit 10, and enter, as placeholders, the amounts 1,000, 500 and 100 in each cell. Change the formula in cell F4 to: *$K$18+$K$19*C4+$K$20*D4,* and copy it down to and through row 63. Next, we will set up Solver to find the optimal values for the trend parameters in cells K18–K20 (K54, K55, and K56 on the truncated demonstration sheet) and seasonal parameters in cells K4:K15 that will minimize the MSE.

On the ribbon, click on the Data tab, and then to the far right of the ribbon,

	A	B	C	D	E	F	G
1	**EXHIBIT 7:** CALCULATING ACTUAL VALUE TO PROJECTED TREND VALUE						
2					ADJUSTED	QUADRATIC	ADJUSTED AS
3	YEAR	MONTH	TIME PERIOD	TIME PERIOD^2	SALES	TREND	A % OF TREND
4	2007	11	1	1	166,726	122,962	135.6%
5		12	2	4	259,602	124,195	209.0%
6	2008	1	3	9	83,365	125,405	66.5%
7		2	4	16	79,127	126,594	62.5%
8		3	5	25	73,961	127,760	57.9%
9		4	6	36	78,226	128,905	60.7%
10		5	7	49	99,677	130,027	76.7%
11		6	8	64	115,982	131,127	88.4%
12		7	9	81	117,528	132,206	88.9%
13		8	10	100	127,920	133,262	96.0%
14		9	11	121	107,008	134,296	79.7%
15		10	12	144	140,797	135,309	104.1%
16		11	13	169	217,120	136,299	159.3%
17		12	14	196	285,837	137,267	208.2%
54	2012	1	51	2601	95,237	157,603	60.4%
55		2	52	2704	92,527	157,734	58.7%
56		3	53	2809	99,371	157,844	63.0%
57		4	54	2916	112,797	157,931	71.4%
58		5	55	3025	134,368	157,995	85.0%
59		6	56	3136	172,862	158,038	109.4%
60		7	57	3249	165,841	158,059	104.9%
61		8	58	3364	181,642	158,058	114.9%
62		9	59	3481	155,015	158,035	98.15
63		10	60	3600	174,795	157,990	110.6%

click on Solver and select cell H65 to be minimized by changing cells K4:K15 and cells K18:K20 (K54, K55, and K56 on the truncated demonstration sheet) while constraining cell K16 to a value of 1 (or 100 percent). This constraint is necessary because if the seasonal indices do not average to 100 percent, there is some upward or downward bias in the trend component of the model. Click Solve, and the results are shown on Exhibit 10. Notice that the MSE has been reduced to 82,530,189, an improvement in model accuracy of 7.7 percent over the MSE of 89,444,067 shown on Exhibit 9.

What we have accomplished here is to minimize MSE, which is the average distance or deviation between adjusted sales and the seasonal forecast by using Excel's Solver optimizer add-in to change the two slope and intercept values of the quadratic trend line. At the same time, we adjusted the twelve seasonal indices while constraining them to still average 100 percent of monthly sales. Solver works by iterating possible solutions until it arrives at a combination of slope values, intercept value, and twelve seasonal indices that minimize MSE.

EXHIBIT 8: CALCULATING THE SEASONAL INDEX

	A	B	C	D	E	F	G	H	I	J	K
2–3	YEAR	MONTH	TIME PERIOD	TIME PERIOD^2	ADJUSTED SALES	QUADRATIC TREND	ADJUSTED AS A % OF TREND			MONTH	SEASONAL INDEX
4	2007	11	1	1	166,726	122,962	135.6%			1	61.4%
5		12	2	4	259,602	124,195	209.0%			2	62.2%
6	2008	1	3	9	83,365	125,405	66.5%			3	60.7%
7		2	4	16	79,127	126,594	62.5%			4	69.6%
8		3	5	25	73,961	127,760	57.9%			5	85.7%
9		4	6	36	78,226	128,905	60.7%			6	97.5%
10		5	7	49	99,677	130,027	76.7%			7	97.5%
11		6	8	64	115,982	131,127	88.4%			8	105.7%
12		7	9	81	117,528	132,206	88.9%			9	86.3%
13		8	10	100	127,920	133,262	96.0%			10	112.8%
14		9	11	121	107,008	134,296	79.7%			11	158.1%
15		10	12	144	140,797	135,309	104.1%			12	202.5%
16		11	13	169	217,120	136,299	159.3%				
17		12	14	196	285,837	137,267	208.2%				
54	2012	1	51	2601	95,237	157,603	60.4%				
55		2	52	2704	92,527	157,734	58.7%				
56		3	53	2809	99,371	157,844	63.0%				
57		4	54	2916	112,797	157,931	71.4%				
58		5	55	3025	134,368	157,995	85.0%				
59		6	56	3136	172,862	158,038	109.4%				
60		7	57	3249	165,841	158,059	104.9%				
61		8	58	3364	181,642	158,058	114.9%				
62		9	59	3481	155,015	158,035	98.1%				
63		10	60	3600	174,795	157,990	110.6%				

FURTHER OPTIMIZING THE MODEL

While this is an improvement over the previous model, we can optimize the model even further by transforming by exponentiation the dependent variable, sales, and modifying the exponent of the second degree of the quadratic model, time^2. We do this because very often time series data are (1) rarely linear, (2) infrequently homogeneous as to variance, and (3) not often distributed normally, or even, symmetrically. Fortunately, these three problems can be fixed with one procedure that involves the transformation of either or both the independent and dependent variables. This is so because data that is not normally distributed is also often neither linear nor homogeneous. Thus, transformation provides a simple way both to fix statistical problems (non-symmetrical, non-normal and heterogeneous distributions), and to better fit curves to data (curvilinear regression). To accomplish this we need to modify Exhibit 10 and then once more

	A	B	C	D	E	F	G	H	I	J	K
1	**EXHIBIT 9:** ESTIMATED QUADRATIC TREND VALUE FOR EACH TIME PERIOD										
2			TIME	TIME	ADJUSTED	QUADRATIC	ADJUSTED AS	SEASONAL			SEASONAL
3	YEAR	MONTH	PERIOD	PERIOD^2	SALES	TREND	A % OF TREND	FORECAST		MONTH	INDEX
4	2007	11	1	1	166,726	122,962	135.6%	194,359		1	61.4%
5		12	2	4	259,602	124,195	209.0%	251,496		2	62.2%
6	2008	1	3	9	83,365	125,405	66.5%	77,025		3	60.7%
7		2	4	16	79,127	126,594	62.5%	78,779		4	69.6%
8		3	5	25	73,961	127,760	57.9%	77,520		5	85.7%
9		4	6	36	78,226	128,905	60.7%	89,761		6	97.5%
10		5	7	49	99,677	130,027	76.7%	111,446		7	97.5%
11		6	8	64	115,982	131,127	88.4%	127,904		8	105.7%
12		7	9	81	117,528	132,206	88.9%	128,840		9	86.3%
13		8	10	100	127,920	133,262	96.0%	140,898		10	112.8%
14		9	11	121	107,008	134,296	79.7%	115,946		11	158.1%
15		10	12	144	140,797	135,309	104.1%	152,615		12	202.5%
16		11	13	169	217,120	136,299	159.3%	215,439			
17		12	14	196	285,837	137,267	208.2%	277,968			
54	2012	1	51	2601	95,237	157,603	60.4%	96,801			
55		2	52	2704	92,527	157,734	58.7%	98,158			
56		3	53	2809	99,371	157,844	63.0%	95,774			
57		4	54	2916	112,797	157,931	71.4%	109,973			
58		5	55	3025	134,368	157,995	85.0%	135,417			
59		6	56	3136	172,862	158,038	109.4%	154,153			
60		7	57	3249	165,841	158,059	104.9%	154,035			
61		8	58	3364	181,642	158,058	114.9%	167,114			
62		9	59	3481	155,015	158,035	98.1%	136,441			
63		10	60	3600	174,795	157,990	110.6%	178,197			
64											
65							MSE	89,444,067			

employ Excel's Solver add-in to optimize the transformation procedure by again minimizing MSE.

Set up Exhibit 11 (page 18) by copying columns A, B, and C from Exhibit 10. Enter the labels Sales Exp and Time^2 Exp and the placeholder amounts of 2 and 1 in cells J18:K19 (J54:K55 on the truncated demonstration sheet). In cell D4 enter the formula: =C4^$K$19 and copy it down to and through row 63. In column E enter the appropriate optimized seasonal index from cells K4:K15 on Exhibit 10. In column F bring over adjusted sales from column E on Exhibit 10. In cell G4 enter the formula: *=F4^$K$18* and copy it down and through row 63. In cell H4 enter the formula: *=TREND ($G$4:$G$63, $C$4:$E$63, C4:E4) ^ (1/$K$18)* and copy it down to and through row 63. Enter the MSE formula in cell H65, with cells H4:H63 and F4:F63

EXHIBIT 10: TRANSFORMATION OF INDEPENDENT AND DEPENDENT VARIABLES

	A	B	C	D	E	F	G	H	I	J	K
2–3	YEAR	MONTH	TIME PERIOD	TIME PERIOD^2	ADJUSTED SALES	QUADRATIC TREND	ADJUSTED AS A % OF TREND	SEASONAL FORECAST		MONTH	SEASONAL INDEX
4	2007	11	1	1	166,726	117,973	141.3%	188,122		**1**	61.0%
5		12	2	4	259,602	119,529	217.2%	240,542		**2**	61.9%
6	2008	1	3	9	83,365	121,056	68.9%	73,840		**3**	60.7%
7		2	4	16	79,127	122,553	64.6%	75,821		**4**	70.1%
8		3	5	25	73,961	124,021	59.6%	75,292		**5**	85.6%
9		4	6	36	78,226	125,460	62.4%	87,902		**6**	97.9%
10		5	7	49	99,677	126,870	78.6%	108,645		**7**	97.6%
11		6	8	64	115,982	128,250	90.4%	125,534		**8**	105.7%
12		7	9	81	117,528	129,601	90.7%	126,470		**9**	86.4%
13		8	10	100	127,920	130,922	97.7%	138,446		**10**	112.5%
14		9	11	121	107,008	132,215	80.9%	114,176		**11**	159.5%
15		10	12	144	140,797	133,478	105.5%	150,103		**12**	201.2%
16		11	13	169	217,120	134,711	161.2%	214,813		**Average**	100.0%
17		12	14	196	285,837	135,915	210.3%	273,519			
54	2012	1	51	2601	95,237	159,849	59.6%	97,503		Intercept	116,386.98
55		2	52	2704	92,527	159,939	57.9%	98,950		Slope 1	1,600.22
56		3	53	2809	99,371	159,999	62.1%	97,134		Slope 2	(14.67)
57		4	54	2916	112,797	160,030	70.5%	112,123			
58		5	55	3025	134,368	160,032	84.0%	137,044			
59		6	56	3136	172,862	160,004	108.0%	156,615			
60		7	57	3249	165,841	159,947	103.7%	156,082			
61		8	58	3364	181,642	159,860	113.6%	169,046			
62		9	59	3481	155,015	159,744	97.0%	137,950			
63		10	60	3600	174,795	159,599	109.5%	179,478			
64											
65							MSE	82,530,189			

as the necessary elements. Next, click on Solver and select cell H65 to be minimized by changing cells K18:K19, and then click Solve. Your results should look like Exhibit 11. Notice that the MSE has been reduced to 72,169,758, an improvement over the model on Exhibit 10 of 14.4 percent. We can now use the results of this model to produce a line plot graph on Exhibit 12 (page 19) showing adjusted sales versus forecasted sales.

A CONCLUDING LINE PLOT

From Exhibit 12's line plot we can see that the forecasted values match up very closely with the adjusted sales. In fact, the average distance between the adjusted and forecasted data points is only $8,793, which translates into a coefficient of variation of only 6 percent. This means that if we were to use this model to predict sales for period 61 as shown on Exhibit 13, we could be 95

	A	B	C	D	E	F	G	H	I	J	K
1	**EXHIBIT 11:** SALES EXP AND TIME^2										
2			TIME	TIME	SEASONAL	ADJUSTED	EXPONENTIATED	QUADRATIC			SEASONAL
3	YEAR	MONTH	PERIOD	PERIOD^2	INDEX	SALES	SALES	TREND		MONTH	INDEX
4	2007	11	1	1.000	159.5%	166,726	42,693	169,836		**1**	61.0%
5		12	2	0.466	201.2%	259,602	63,223	260,484		**2**	61.9%
6	2008	1	3	0.298	61.0%	83,365	23,091	64,771		**3**	60.7%
7		2	4	0.217	61.9%	79,127	22,047	70,024		**4**	70.1%
8		3	5	0.170	60.7%	73,961	20,766	71,106		**5**	85.6%
9		4	6	0.139	70.1%	78,226	21,824	85,694		**6**	97.9%
10		5	7	0.117	85.6%	99,677	27,056	108,840		**7**	97.6%
11		6	8	0.101	97.9%	115,982	30,946	127,528		**8**	105.7%
12		7	9	0.089	97.6%	117,528	31,311	128,265		**9**	86.4%
13		8	10	0.079	105.7%	127,920	33,754	141,058		**10**	112.5%
14		9	11	0.071	86.4%	107,008	28,813	114,287		**11**	159.5%
15		10	12	0.065	112.5%	140,797	36,750	152,681		**12**	201.2%
16		11	13	0.059	159.5%	217,120	53,958	223,758		**Average**	100.0%
17		12	14	0.055	201.2%	285,837	68,857	289,274			
54	2012	1	51	0.013	61.0%	95,237	25,984	104,254		Sales Exp	0.887
55		2	52	0.013	61.9%	92,527	25,328	106,055		Time^2 Exp	(1.102)
56		3	53	0.013	60.7%	99,371	26,982	105,016			
57		4	54	0.012	70.1%	112,797	30,191	118,783			
58		5	55	0.012	85.6%	134,368	35,259	141,696			
59		6	56	0.012	97.9%	172,862	44,083	160,151			
60		7	57	0.012	97.6%	165,841	42,492	160,323			
61		8	58	0.011	105.7%	181,642	46,063	172,966			
62		9	59	0.011	86.4%	155,015	40,023	145,148			
63		10	60	0.011	112.5%	174,795	44,520	184,165			
64											
65							MSE	72,169,758			

EXHIBIT 12: THREE MONTH PREDICTION

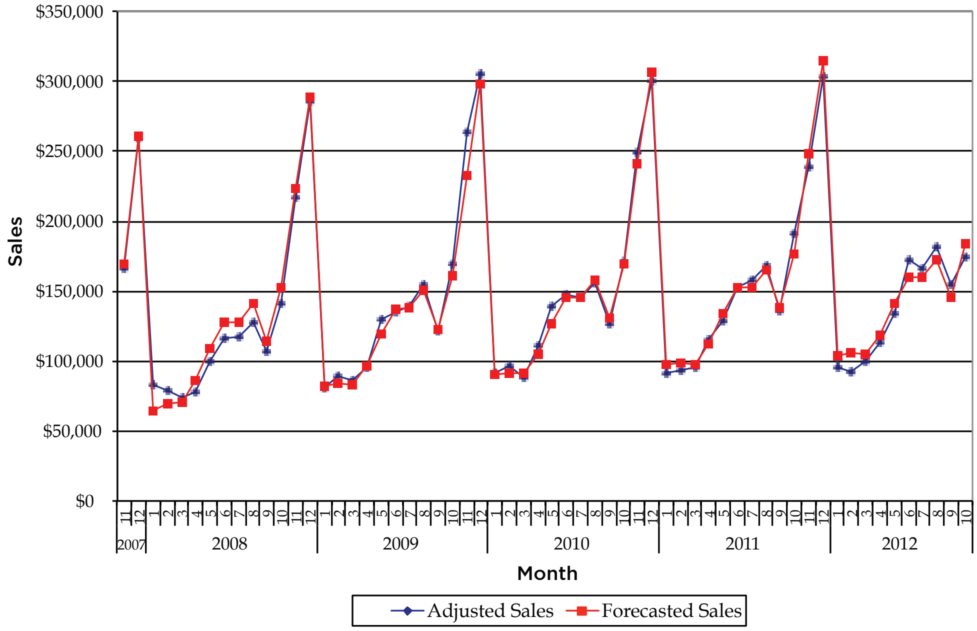
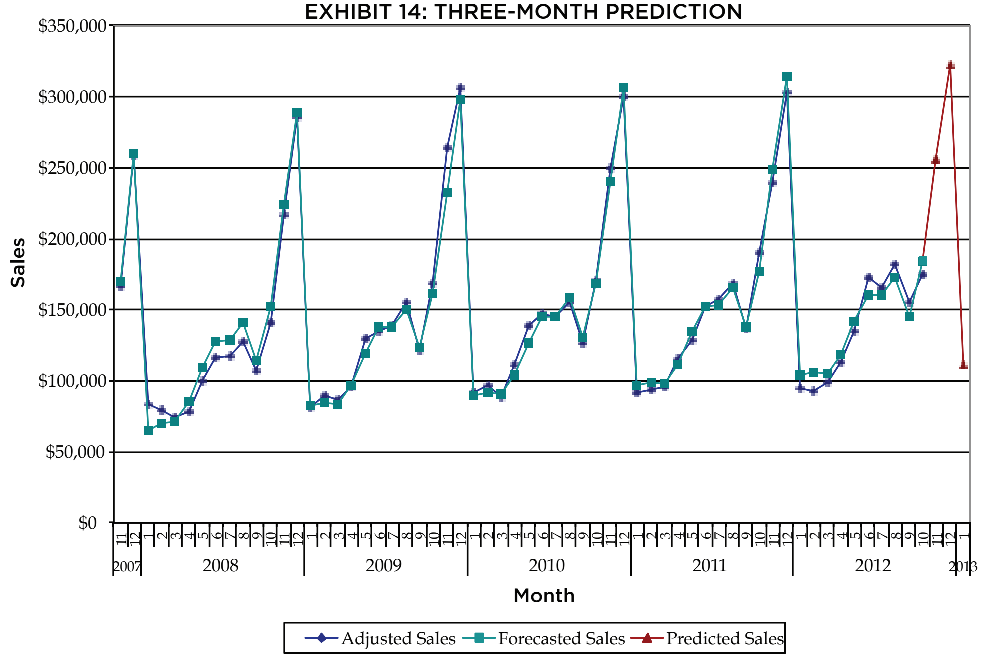


EXHIBIT 14: THREE-MONTH PREDICTION



percent confident that the actual level of sales observed would fall somewhere in the range from $238,682 to $273,854 ($256,268 ± $17,586).

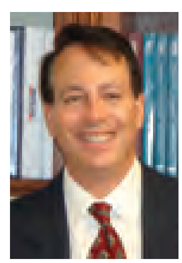
SUMMARY OUTPUT AND CONCLUSION

Exhibit 13 (page 20) presents Excel's summary output for its regression tool, and we also show the computations for the expected sales during the three-month period of interruption, including the conversion of sales back to actual days in a month.

Also shown is a description of the regression equation itself and how it works to produce a forecasted value for any particular month. The regression statistics are all excellent, and the absolute t-stats of the x coefficients are all way above 2, all of which indicates, along with the very low standard error of the estimate, an extremely good fitting model with highly accurate results. The end result can be seen on Exhibit 14 that shows the three-month prediction in relation to the sixty-month's of history on which it is based.

We can conclude that this regression time series model is a succinct method of predicting future values of a time series variable when seasonal and trend patterns are present. The goal was to demonstrate how a model can be fitted to the past behavior of a time series and then be used to predict future values using statistical techniques that can be accomplished using Excel. In future articles we will discuss and present the application of Auto Regressive Integrated Moving Average (ARIMA) models and explanatory forecasting models to Rollie's sales history. VE

Mark G. Filler, CPA/ABV, CBA, AM, CVA, leads Filler & Associates' litigation and claims support practice in Portland, ME. He has testified over 100 times on valuation and damages matters. He is past chairman of the Editorial Board of The Value Examiner. *E-mail: mfiller@filler.com.*



James A. DiGabriele, PhD, DPS, CPA/ABV/CFF, CVA, is an accounting professor at Montclair State University. He is also the managing director of DiGabriele, McNulty, Campanella & Co., LLC, in Fairfield, NJ, an accounting firm specializing in forensic/investigative accounting and litigation support. He has qualified as an expert in forensic accounting in federal and state courts. E-mail: digabrielej@mail.montclair.edu.

EXHIBIT 13: SUMMARY OUTPUT

REGRESSION STATISTICS	TRANSFORMED	BACK-TRANSFORMED
Multiple R	0.990	0.990
R Square	0.980	0.980
Adjusted R Square	0.979	0.979
Standard Error	2,017	8,793
Coefficient of Variation	5.33%	5.99%
Observations	60	60

ANOVA

	DF	SS	MS	F	SIGNIFICANCE F
Regression	3	11,021,252,254	3,673,750,751	903	0.0
Residual	56	227,929,305	4,070,166		
Total	59	11,249,181,559			

	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE	LOWER 95%	UPPER 95%
Intercept	925.379	893.368	1.036	0.305	(864.250)	2,715.009
Time Period	136.431	17.582	7.760	0.000	101.209	171.652
Time Period-1.10178	(11,037.043)	2,182.390	(5.057)	0.000	(15,408.893)	(6,665.194)
Seasonal Factor	33,470.974	659.070	50.785	0.000	32,150.700	34,791.248

PREDICTION

Regression Equation: Forecasted Monthly Sales = $(925.379 + \text{Time Period} \times 136.431 + \text{Time Period}^{-1.10178} \times -11037.043 + \text{Seasonal Factor} \times 33470.974)^{(1/.886701)}$

MONTH	INTERCEPT	TIME PERIOD	TIME PERIOD -1.10178	SEASONAL FACTOR	TRANSFORMED FORECAST	BACK-TRANSFORMED FORECAST$^{(1/.886701)}$	RECONVERTED
Nov-12	925.379	61	0.011	159.46%	62,502	256,268	252,585
Dec-12	925.379	62	0.011	201.24%	76,625	322,458	328,417
Jan-13	925.379	63	0.010	61.00%	29,822	111,243	113,298
Total							694,300

GLOSSARY

MSE (Mean Squared Error)	The sum of the squared difference between the actual response and the predicted response divided by the error degrees of freedom. Measures the degree of dispersion in the data set.
Degrees of freedom	Number of data points less the number of regression parameters, including the constant.
SEE (Standard Error of the Estimate)	The square root of the MSE. States the degree of dispersion in terms of the original response, or data set.
t statistic	The x coefficient value divided by its standard error. Indicates how many standard deviations the x coefficient is from zero.
p value	Indicates if the x coefficient is statistically significant at a particular level, i.e., if it is significantly different from zero.
COV (Coefficient of Variation)	The SEE divided by the average of the dependent variable. Allows for comparability of the degree of dispersion among differing data sets, models, etc.
r^2	The degree of explanatory power of the model. Takes a value between 0 and 1.
F statistic	Indicates if the model as a whole is statistically significant.
Grand mean	The mean of several subgroups; in this case the average of the twelve monthly averages.
Root Mean Square Error (RMSE)	A measure of forecast accuracy that has the same advantage of the MSE in that a penalty is assessed to large forecast errors, but, the value of RMSE is comparable in magnitude to other commonly used statistics for forecast accuracy.
Mean Absolute Percentage Error (MAPE)	A measure of forecasting accuracy that is used whenever knowledge of the size of the forecast error in relation to the size of the variable to be forecast is important.
Mean Percentage Error (MPE)	A measure of forecast accuracy that is used whenever the forecaster is interested in determining whether the forecast error is biased.
Mean Absolute Deviation (MAD)	A measure of forecasting accuracy that is used whenever the forecaster wishes to obtain a measure of the forecast error that is expressed in magnitudes comparable to the original observations.
Dummy Variable	A variable that indicates the category a given observation is in. For example, the category January would be coded 1, and all other months would be coded 0, and so on for each of the other months.