LONGITUDINAL REGRESSION FORECASTING USING EXCEL

John F. Kros and S. Scott Nadler

Forecasting is an integral business activity that is covered to varying degrees in most marketing textbooks. Texts generally present the material from a theoretical perspective, but many students need to experience the material from an active learning. To bridge this pedagogical gap this article provides a step-by-step tutorial for teaching forecasting using Excel based regression analysis to forecast seasonal demand. Three seasonal regression models are presented and interpreted including a quarterly model, a bi-monthly model, and a monthly model. The results of the models are compared and each model’s efficacy discussed. This assignment has been well received by students.

Keywords: seasonal forecasting models; spreadsheet forecasting models, active learning; linear regression; seasonality

Introduction

One of the most common recurring complaints from Business School alumni is that many new graduates lack technical skills such as the ability to use standard statistical packages and spreadsheets necessary to complete basic tasks such as forecasting. While these new employees are well versed in forecasting from a theoretical perspective, they all too frequently lack the technical expertise associated with active learning (Davis, Misra, and Auken 2002).

Forecasting is an important part of all successful marketing programs and there have been numerous articles which speak to the growing need for accurate forecasts (Keating and Wilson 1988; Hanke and Weigand 1994; Wilson and Daubeck 1989; Loomis and Cox 2003; and Armstrong and Brodie 1999). Related studies have shown that computer usage in forecasting courses has grown and that a large portion of faculty teaching forecasting use some type of software in their curriculum (Wilson and Daubeck 1989; Armstrong and Brodie 1999; and Loomis and Cox 2003).

For more than twenty years, forecasters have used computerized spreadsheets in an attempt to produce more accurate forecasts. During this time, numerous researchers have published work ranging from selecting the appropriate method of forecasting (see Chase 1997) to how to use specific forecasting approaches in spreadsheets (Albright et al. 2005; Kros 2007; Ragsdale 2004; Savage 2003; Grossman 1999). Other studies have shown that the most commonly used method of forecasting is based on linear multiple regression (Wilson and Daubeck 1989; Hanke 1984, 1988; and Loomis and Cox 2000). Radovilsky and Eyck (2000) provided a much needed discussion on forecasting with Excel. Although detailed, most of these works failed to address the issue of seasonality and the creation of in depth seasonal forecasting models via Excel and even fewer provide marketing instructors with the means to teach seasonally influenced forecasting effectively.

During this period there has also been a trend to educate college students on the use of spreadsheets to solve problems within the business curriculum (Grossman 1997; Powell 1997; Fylstra et al. 1998; Conway and Ragsdale 1997; Savage 1997). Many of the textbooks for the traditional management science course now include spreadsheet-modeling applications (Anderson et al. 2005; Ragsdale 2004). Some textbook authors utilize spreadsheet modeling as a major focus of business problem solving (Kros 2005; Moore et al. 2001). Consequently, the use of spreadsheets in financial and process analysis has become commonplace for many managers today, and the problem solving process now includes thinking in terms of spreadsheets rather than mathematical functions (Powell 1997; Sonntag and Grossman 1999). Since many marketing managers use spreadsheets extensively, the authors feel it is extremely important to educate students in the building spreadsheet forecasting models. There are several benefits to a spreadsheet models for forecasting: (a) spreadsheet use is ubiquitous in and around corporate settings, (b) data

Marketing Education Review, Volume 18, Number 3 (Fall 2008).
input is simple and the functionality of the data easily interpretable, (c) most firms currently own some form of spreadsheet software, and (d) since most firm’s mid-level through senior managers tend to be familiar with using spreadsheets for problem solving, they should also view spreadsheet based forecast models as credible.

**Purpose and Contribution**

In this article we provide readers with a step-by-step guide to forecasting with Excel using regression analysis under conditions characterized by seasonality. To accomplish the stated objective the authors present a discussion of seasonal regression models in order to familiarize readers with this important concept. This is followed by an overview of using Excel based regression analysis to forecast seasonal demand using Imported Beer sales data from the Beer Institute. This data set was chosen because it is available for download free of charge, it is replicable, and the results are readily interpretable. The data set also represents a realistic snap-shot of an industry with close ties to the marketing discipline as well as many jobs associated with student employment such as wait staffing, hospitality management, and retail sales. While other data sets can be obtained from organizations such as the U.S. Department of Agriculture and the United Nations, the authors’ students have generally found that working with these data sets is more difficult. Three seasonal regression models are presented and interpreted including a quarterly model, a bi-monthly model, and a monthly model. The results of the models are compared and a discussion of each model’s efficacy is provided.

It has been noted that there are a myriad of forecasting texts and articles that include real world examples, demonstrations, and justifications for the use of sophisticated statistical tools such as SPSS and SAS. However, there remains a significant pedagogical gap in what the needs and perceptions of managers in industry are and what is actually being taught at the university level. To illustrate this point and justify the manuscript’s contribution in these two areas, the authors conducted a content analysis of current textbooks and two surveys (one from the academic side and one from the industry side). The results of these surveys drove the content analysis. The academic survey provided important insight into the type of texts to review, software used, and the most commonly taught forecasting techniques. The industry survey provided insights into the types of software and forecasting techniques used. The academic survey contained 116 usable surveys and obtained a 35% response rate, whereas the industry survey contained 112 usable surveys with a 32% response rate. Overall, the results of this combined research approach indicate the following: that there is a significant gap between what is being taught in universities and what industry practitioners need. In addition, the preponderance of industry respondents (80.5%) report that the primary forecasting tool available to them is Excel. Finally, none of the texts currently on the market provides a concise step-by-step tutorial for the use of regression with seasonal constraints. Consequently, we believe it is very important to provide marketing instructors and industry managers with a step-by-step process for employing Excel to develop a model for seasonal regression forecasting.

**Literature Review**

The marketing education literature is replete with examples that exemplify both the need to incorporate written (technical) assignments into the marketing curriculum and examples of these assignments (e.g., West 2006). The impetus for these studies stems from the fact that about 40% of college students are visual learners. These students prefer to be taught through pictures, diagrams, flow charts, timelines, films, and demonstrations. Yet marketing instruction remains heavily reliant on presenting content primarily through lectures or written assignments. Consequently, some students may be underperforming because the teachers’ teaching styles and students’ learning styles are not in synch (Clark, Flaherty, and Yankey 2006).

It has been suggested that many college students learn better from visual stimuli and efforts to match teaching and learning styles could offer significant learning benefits. Given these finding the purpose of this tutorial is to assist marketing instructors teach students about seasonal forecasting with multiple regression models through the use of Excel spreadsheets. To accomplish the stated objective the authors will begin by providing readers with an overview of seasonal regression models.

**Seasonal Regression Models**

Corporations typically need forecasts that cover different time spans in order to achieve operational, tactical, and strategic intents. Firms typically use monthly data from the last one or two years to achieve operational or short term forecasting. Tactical forecasting is generally based on quarterly data from the last five to six years or comparable annual data. Strategic forecasting generally requires additional periods in order to make projections for 25 or more years into the future (Lapide 2002; Zhou 1999). Tactical and strategic forecasting is made more complex when the issue of seasonality is added to the analysis.
One goal of a forecasting model is to account for the largest amount of systemic variation in the behavior of a time series data set as possible. Moving average, exponential smoothing, and linear regression models all attempt to account for systemic variation. However, each of these models may fail due to additional systemic variation that is not accounted for. In many business time series data sets a major source of systemic variation comes from seasonal effects. Seasonal variation is characterized by increases or decreases in the time series data at regular time intervals, namely calendar or climatic changes (Ragsdale 2006).

For example, sales demand for beer in the United States has increased over time but tends to vary during the year and to be higher in the Spring and Summer months than in the Fall and Winter months (http://www.foodandbeveragereports.com). The most dramatic seasonality occurs in what are appropriately referred to as seasonal beers. These beers tend to get very popular during the fall and early winter season and during the holidays. Darker beers like porters and stouts tend to be more popular during the cooler months, while lighter styles like Wheat beers and Pale Ales spike up during the warmer months. It has been said that the thirst for beer tends to change with nature’s seasons. Some who enjoy beer stay loyal throughout the year, but statistics have shown that beer consumption changes during the years in regular cycles (Northern Colorado Beer Examiner 2008). Therefore, time is not the only variable that has an impact on beer sales; multiple factors play a role.

Multiple linear regression models are a commonly used technique in forecasting when multiple independent variables impact a dependent variable. From the earlier example, beer sales could be considered the dependent variable while time and a seasonal factor could be considered independent variables, and is represented in the following general model for multiple linear regression:

\[
Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} + \ldots + b_nX_{nt} + \epsilon_t
\]

(1)

where \(X_{1t}\) = time and \(X_{id}\) through \(X_{nt}\) are seasonal indicators. The \(X\)'s denote the independent variables while the \(Y\) denotes the dependent variable. For example, the \(X_{1t}\) term represents the first independent variable for the time period \(t\) (e.g., \(X_{1t}=1, X_{1t}=2, X_{1t}=3, \ldots\)). The \(\epsilon_t\) term denotes the random variation in the time series not accounted for by the model. Since the values of \(Y_t\) are assumed to vary randomly around the regression function, the average or expected value of \(\epsilon_t\) is 0.

Therefore, if an ordinary least squares estimator is employed the best estimate of \(Y_t\) for any time period \(t\) is:

\[
Y_t = \hat{b}_0 + \hat{b}_1X_{1t} + \hat{b}_2X_{2t} + \hat{b}_3X_{3t} + \ldots + \hat{b}_nX_{nt}
\]

(2)

Equation (2) represents the line passing through the time series that minimizes the sum of squared differences between actual values \((Y_t)\) and the estimated values \((\hat{Y}_t)\). In the case when \(n=1\) the equation represents simple regression.

However, if a data set contains seasonal variation a standard multiple linear regression model generally does not provide very good results. With seasonal effects, the data tend to deviate from the trend lines in noticeable patterns. Forecasts for future time periods would be much more accurate if the regression model reflected these drops and ascents in the data. Seasonal effects can be modeled using regression by including indicator variables, which are created to indicate the time period to which each observation belongs. The next sections present three seasonal regression forecasting models. The first is a quarterly model followed by a bimonthly and then a monthly model.

Many publicly held companies are required to submit quarterly (data occurring in cycles of three months—a common business cycle) reports regarding the status of their businesses. It is also very common for quarterly forecasts to be used to develop future quarterly projections. Therefore, if quarterly data were being analyzed, the indicator variables for a quarterly regression model could be stated as follows:

\[
X_{1t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the first quarterly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(3)

\[
X_{2t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the second quarterly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(4)

\[
X_{3t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the third quarterly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(5)

\[
X_{4t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the fourth quarterly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(6)

Notice the unique coding used to define \(X_{1t}, X_{2t}, X_{3t}, X_{4t}\) and \(X_{id}\) in equations 3 through 6. Table 1 summarizes the coding structure.

At times, bimonthly data (occurring in cycles of two months) data is required. In turn, if bimonthly date were being analyzed, the indicator variables could be stated as follows:

\[
X_{1t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the first bimonthly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(7)

\[
X_{2t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the second bimonthly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(8)

\[
X_{3t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the third bimonthly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(9)

\[
X_{4t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the fourth bimonthly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(10)

\[
X_{5t} = \begin{cases} 
1, & \text{if } Y_t \text{ is an observation from the fifth bimonthly period of any year, otherwise} \\
0, & \text{otherwise}
\end{cases}
\]

(11)
Notice the unique coding used to define $X_{2t}$, $X_{3t}$, $X_{4t}$, $X_{5t}$, $X_{6t}$, and $X_{7t}$ in equations 7 through 12. Table 2 summarizes the coding structure.

However, some businesses forecast on a monthly basis. For the case of a monthly cycle, the indicator variables could be stated as follows:

$$X_{9t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the sixth bimonthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Notice the unique coding used to define $X_{2t}$, $X_{3t}$, $X_{4t}$, $X_{5t}$, $X_{6t}$, and $X_{7t}$ in equations 7 through 12. Table 2 summarizes the coding structure.

However, some businesses forecast on a monthly basis. For the case of a monthly cycle, the indicator variables could be stated as follows:

$$X_{2t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the first monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$X_{3t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the second monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

$$X_{4t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the third monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$X_{5t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the fourth monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

$$X_{6t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the fifth monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

$$X_{7t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the sixth monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

$$X_{8t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the seventh monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

$$X_{9t} = \begin{cases} 1, & \text{if } Y_t \text{ is an observation from the eighth monthly period of any year} 0, \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

Table 3 summarizes the coding structure.

Overall, this coding structure can apply to any time frame the forecaster chooses (e.g., monthly seasonal periods, or peak versus non-peak seasons) and each of the models that will be discussed herein.

### The Seasonal Model

In general, if we identify $p$ seasons, we only need $p-1$ indicator variables in the regression model. Therefore, any one of the seasonal indicators can be omitted and the choice of which indicator to omit is left to the modeller. For example, in our monthly seasonal example although twelve monthly periods are identified only eleven indicator variables are needed. The excluded monthly period will simply serve as a base-level period to measure changes in the other seasonal levels. Therefore, if the first monthly indicator, $X_{2t}$ is excluded (i.e.,
Table 3
Monthly Seasonal Coding Structure

<table>
<thead>
<tr>
<th>Monthly Period</th>
<th>Value of Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_{2t}$</td>
</tr>
<tr>
<td>1-Jan</td>
<td>1</td>
</tr>
<tr>
<td>2-Feb</td>
<td>0</td>
</tr>
<tr>
<td>3-Mar</td>
<td>0</td>
</tr>
<tr>
<td>4-Apr</td>
<td>0</td>
</tr>
<tr>
<td>5-May</td>
<td>0</td>
</tr>
<tr>
<td>6-Jun</td>
<td>0</td>
</tr>
<tr>
<td>7-Jul</td>
<td>0</td>
</tr>
<tr>
<td>8-Aug</td>
<td>0</td>
</tr>
<tr>
<td>9-Sept</td>
<td>0</td>
</tr>
<tr>
<td>10-Oct</td>
<td>0</td>
</tr>
<tr>
<td>11-Nov</td>
<td>0</td>
</tr>
<tr>
<td>12-Dec</td>
<td>0</td>
</tr>
</tbody>
</table>

$X_{3t}$ will always be equal to 0), the seasonal regression model can then be developed as follows:

$$\hat{Y} = b_0 + b_1 X_{1t} + b_2 X_{2t} + b_3 X_{3t} + b_4 X_{4t} + b_5 X_{5t} + b_6 X_{6t} + b_7 X_{7t} + b_8 X_{8t} + b_9 X_{9t} + b_{10t} X_{10t} + b_{11t} X_{11t} + b_{12t} X_{12t} + b_{13t} X_{13t}$$

(25)

The coefficients $b_{3t}$, $b_{4t}$, $b_{5t}$, $b_{6t}$, $b_{7t}$, $b_{8t}$, $b_{9t}$, $b_{10t}$, $b_{11t}$, $b_{12t}$, and $b_{13t}$ indicate the expected effects of seasonality in the second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, eleventh, and twelfth monthly periods, respectively, relative to the first monthly period.

To better understand how the regression model behaves, notice that for observations occurring in the first monthly period (i.e., $X_{3t} = 0, X_{4t} = 0, X_{5t} = 0, X_{6t} = 0, X_{7t} = 0, X_{8t} = 0, X_{9t} = 0, X_{10t} = 0, X_{11t} = 0, X_{12t} = 0$, and $X_{13t} = 0$) the regression model reduces to:

$$\hat{Y} = b_0 + b_1 X_{1t}$$

(26)

This model represents the predicted values of import beer sales in the 1st monthly period indexed on time. This analogy applies for the quarterly or bimonthly seasonal models also. In the following examples, the first seasonal indicator is omitted for each of the three models studied.

Import Beer Sales Example

Figure 1.1 displays three views of the partial data set for United States Import Beer sales in millions of barrels (Brls) from January 2000 to September 2005 in an Excel spreadsheet. The views are displayed to illustrate the differing setups or data structure that accompanies each seasonal regression model. The data structure or setup in Excel is particularly important when using Excel’s regression procedure. In other words, how the data is input and in which order it is featured (e.g., what data is in the first column versus the second column, etc.).

View (a) in Figure 1.1 displays a quarterly model, view (b) the bimonthly model, and view (c) the monthly model. Take note that (1) the data in the last column in each view is obtained after the Regression procedure is performed in Excel and is explained later in this paper, (2) the first seasonal indicator is omitted in each model’s setup (see discussion in The Seasonal Model section), and (3) the independent variables (time and seasonal indicators) are all contiguous (i.e., side by side with no gaps).

To perform regression analysis use Excel, the Data Analysis Regression procedure must be invoked. The following steps illustrate how this is done for the quarterly seasonal model depicted in Figure 1(a):

Step 1: In an open Excel worksheet, select the Tools menu and within Tools select Data Analysis (refer to Figure 2). If the Data Analysis option does not appear, click Excel’s Help Menu and do a search on installing Excel’s Data Analysis package. The Data Analysis package comes standard with all licensed versions of MS Office.

Step 2: A menu screen resembling Figure 3 will appear, scroll down locating the Regression option and click the OK button.

Step 3: In the Regression dialog box (refer to Figure 4), enter the dependent variable’s worksheet location in the Input Y Range.
Figure 1
Partial US Imported Beer Sales Data and Seasonal Regression Setup Spreadsheets

(a) Partial US Imported Beer Sales Data and Seasonal Regression Setup Spreadsheets

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Quarterly Period</th>
<th>Time</th>
<th>Indicator for Quarterly Period</th>
<th>Quarterly Period</th>
<th>Actual Beer Import Sales</th>
<th>Predicted Beer Import Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Jan</td>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>Q1</td>
<td>1.406</td>
<td>1.406</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>Q1</td>
<td>2</td>
<td>1</td>
<td>Q1</td>
<td>1.498</td>
<td>1.498</td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>Q1</td>
<td>3</td>
<td>1</td>
<td>Q1</td>
<td>1.585</td>
<td>1.585</td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>Q1</td>
<td>4</td>
<td>1</td>
<td>Q1</td>
<td>1.672</td>
<td>1.672</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>Q2</td>
<td>2.183</td>
<td>2.183</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>Q2</td>
<td>2</td>
<td>1</td>
<td>Q2</td>
<td>2.264</td>
<td>2.264</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>Q2</td>
<td>3</td>
<td>1</td>
<td>Q2</td>
<td>2.152</td>
<td>2.152</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>Q2</td>
<td>4</td>
<td>1</td>
<td>Q2</td>
<td>2.045</td>
<td>2.045</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>Q3</td>
<td>1.952</td>
<td>1.952</td>
</tr>
<tr>
<td></td>
<td>Oct</td>
<td>Q3</td>
<td>2</td>
<td>1</td>
<td>Q3</td>
<td>1.809</td>
<td>1.809</td>
</tr>
<tr>
<td></td>
<td>Nov</td>
<td>Q3</td>
<td>3</td>
<td>1</td>
<td>Q3</td>
<td>1.671</td>
<td>1.671</td>
</tr>
<tr>
<td></td>
<td>Dec</td>
<td>Q4</td>
<td>4</td>
<td>1</td>
<td>Q4</td>
<td>1.539</td>
<td>1.539</td>
</tr>
</tbody>
</table>

(b) Partial US Imported Beer Sales Data and Seasonal Regression Setup Spreadsheets

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Quarterly Period</th>
<th>Time</th>
<th>Indicator for Quarterly Period</th>
<th>Quarterly Period</th>
<th>Actual Beer Import Sales</th>
<th>Predicted Beer Import Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Jan</td>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>Q1</td>
<td>1.306</td>
<td>1.306</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>Q1</td>
<td>2</td>
<td>1</td>
<td>Q1</td>
<td>1.390</td>
<td>1.390</td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>Q1</td>
<td>3</td>
<td>1</td>
<td>Q1</td>
<td>1.473</td>
<td>1.473</td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>Q1</td>
<td>4</td>
<td>1</td>
<td>Q1</td>
<td>1.551</td>
<td>1.551</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>Q2</td>
<td>2.051</td>
<td>2.051</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>Q2</td>
<td>2</td>
<td>1</td>
<td>Q2</td>
<td>2.132</td>
<td>2.132</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>Q2</td>
<td>3</td>
<td>1</td>
<td>Q2</td>
<td>2.018</td>
<td>2.018</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>Q2</td>
<td>4</td>
<td>1</td>
<td>Q2</td>
<td>1.905</td>
<td>1.905</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>Q3</td>
<td>1.790</td>
<td>1.790</td>
</tr>
<tr>
<td></td>
<td>Oct</td>
<td>Q3</td>
<td>2</td>
<td>1</td>
<td>Q3</td>
<td>1.663</td>
<td>1.663</td>
</tr>
<tr>
<td></td>
<td>Nov</td>
<td>Q3</td>
<td>3</td>
<td>1</td>
<td>Q3</td>
<td>1.539</td>
<td>1.539</td>
</tr>
<tr>
<td></td>
<td>Dec</td>
<td>Q4</td>
<td>4</td>
<td>1</td>
<td>Q4</td>
<td>1.416</td>
<td>1.416</td>
</tr>
</tbody>
</table>

(c) Partial US Imported Beer Sales Data and Seasonal Regression Setup Spreadsheets

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Quarterly Period</th>
<th>Time</th>
<th>Indicator for Quarterly Period</th>
<th>Quarterly Period</th>
<th>Actual Beer Import Sales</th>
<th>Predicted Beer Import Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Jan</td>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>Q1</td>
<td>1.216</td>
<td>1.216</td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td>Q1</td>
<td>2</td>
<td>1</td>
<td>Q1</td>
<td>1.295</td>
<td>1.295</td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td>Q1</td>
<td>3</td>
<td>1</td>
<td>Q1</td>
<td>1.373</td>
<td>1.373</td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>Q1</td>
<td>4</td>
<td>1</td>
<td>Q1</td>
<td>1.448</td>
<td>1.448</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>Q2</td>
<td>1.914</td>
<td>1.914</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>Q2</td>
<td>2</td>
<td>1</td>
<td>Q2</td>
<td>1.981</td>
<td>1.981</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>Q2</td>
<td>3</td>
<td>1</td>
<td>Q2</td>
<td>1.942</td>
<td>1.942</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>Q2</td>
<td>4</td>
<td>1</td>
<td>Q2</td>
<td>1.894</td>
<td>1.894</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>Q3</td>
<td>1.774</td>
<td>1.774</td>
</tr>
<tr>
<td></td>
<td>Oct</td>
<td>Q3</td>
<td>2</td>
<td>1</td>
<td>Q3</td>
<td>1.652</td>
<td>1.652</td>
</tr>
<tr>
<td></td>
<td>Nov</td>
<td>Q3</td>
<td>3</td>
<td>1</td>
<td>Q3</td>
<td>1.528</td>
<td>1.528</td>
</tr>
<tr>
<td></td>
<td>Dec</td>
<td>Q4</td>
<td>4</td>
<td>1</td>
<td>Q4</td>
<td>1.404</td>
<td>1.404</td>
</tr>
</tbody>
</table>
edit box—for our example that location is $H3:H72$. Take note that this range changes depending on which model you are employing.

Step 4: In the Regression dialog box (refer to Figure 4), enter the independent variable’s worksheet location in the Input X Range edit box—for our example that location is $D3:G72$. Take note that this range changes depending on which model you are employing.

Step 5: Select the Labels check box—for our example we included labels so this action is appropriate.

Step 6: Select the New Worksheet Ply option button and check the Residuals box then click the OK button.

After clicking OK in the Regression dialog box, a worksheet resembling Figure 5 is generated. Figure 5 provides a partial screenshot from the Data Analysis Regression procedure output for the quarterly seasonal model. Due to space constraints a partial screenshot is displayed. The information that is omitted contains the residuals for the regression. A discussion on residuals appears later in the section on calculating error. The next section discusses the output obtained for the quarterly seasonal model, Figure 5, the bimonthly seasonal model, Figure 6, and the monthly seasonal model, Figure 7.

**Interpretation and Analysis of Regression Outputs**

Nash and Quon (1996) provide a valuable overview of issues in teaching statistical thinking with spreadsheets. They discuss numerous advantages as well as deficiencies regarding employing spreadsheets for statistical teaching. One very important point they make is the verification of statistical output from spreadsheets via traditional statistical packages (e.g., SPSS). It must be noted here that all Excel regression results presented herein were verified via SPSS 13.0.

One advantage to employing Excel’s regression procedure is the ease in interpreting the results of the regression analysis (refer to Figure 5, 6, or 7) and understanding its “statistical quality.” To interpret the regression analysis readers should begin by checking the values for three statistical measures: the R Square statistic, the F statistic, and the t statistic. The value for R Square (cell B5, Figure 5, 6, or 7) provides an assessment of the forecast model’s accuracy. Specifically, it
Figure 3
Data Analysis Menu

Figure 4
Regression Dialog Box

Figure 5
Quarterly Seasonal Model Regression Procedure Output
**Figure 6**
Bimonthly Seasonal Model Regression Procedure Output

**Figure 7**
Monthly Seasonal Model Regression Procedure Output

**Table 4**
R Square Statistics for Various Seasonal Regression Models

<table>
<thead>
<tr>
<th>Model Type</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly</td>
<td>0.6487</td>
</tr>
<tr>
<td>Bimonthly</td>
<td>0.8587</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.8820</td>
</tr>
</tbody>
</table>
can be interpreted as the proportion of the variance in $Y$ attributable to the variance in the $X$ variables. Generally, values above 0.7 provide a minimum threshold of accuracy for business models, while values above 0.8 are considered very good (Kros 2007). Table 4 presents the $R^2$ values for each of the seasonal regression models.

For the models presented here, it can be seen from Table 2 that the $R^2$ statistic for the quarterly model is 0.6487, for the bimonthly model is 0.8587, and for the monthly model is 0.8830. Therefore, it can be said that the quarterly seasonal regression model developed is explaining around 65% of the variation in U.S. Import Beer sales, while the bimonthly seasonal regression model is explaining around 86% of the variation in U.S. Import Beer sales, whereas the monthly seasonal regression is performing the best and explaining approximately 88% of the variation in U.S. Import Beer sales.

The second area to investigate is the value for the $F$ statistic. The $F$ statistic tests how well the coefficients for the independent $X$ variables predict the dependent $Y$ variable. The $F$ statistic can be found in cell E12 (Figure 5, 6, or 7). To properly interpret an $F$ statistic, degrees of freedom must be determined and an $F$ critical value obtained for a given level of significance (e.g., 5%). However, Excel delivers another option for interpreting the $F$ statistic that does not require obtaining an $F$ critical value. Excel calls this the Significance $F$. The Significance $F$ can be found in cell E12 (Figure 5, 6, or 7) and is compared to the level of significance determined for the regression model by the user (e.g., 5% or 0.05). If the Significance $F$ value is lower than the predetermined level of significance, then it is determined that the $F$ statistic is significant. Conversely, if the Significance $F$ value is higher than the predetermined level of significance, then it is determined that the $F$ statistic is not significant. It is highly desirable for the $F$ statistic to be deemed significant if a regression model is to be trusted in predicting the independent variable. Table 5 contains the $F$ statistics and Significance $F$ for the three seasonal regression models.

From Table 3 it is noted that all Significance $F$ values are much lower than the predetermined level of significance of 5%, or 0.05. (NOTE: The numbers in the Significance $F$ column of Table 3 are in Scientific Notation and are read as the first three numbers multiplied by 10 raised to the power shown, denoted by the E.) All three of these values are infinitesimally small; for example, the Significance $F$ value for the Quarterly model can be read as $6.27 \times 10^{-14}$ or a decimal point, 13 zeros, then 627. Therefore, it can be concluded that as a whole, for each of the models the coefficients for all the independent variables are doing a good job of explaining the dependent variable (i.e., overall the seasonal regression model is doing a good job of predicting U.S. Import Beer sales).

The next step in interpreting the regression analysis results requires the reader to determine the significance of the $X$ variables. Specifically, whether or not the results (the $X$ variable coefficients) generated by the seasonal regression model may be attributed to chance or random error. This is accomplished by analyzing the $t$ statistics (for example refer to cells D17:D21, Figure 5).

A good rule of thumb to follow is to verify that the absolute value of a $t$ statistic is above 2.0. Absolute values above 2.0 are considered statistically significant and indicate that the individual independent variable is significantly contributing to the overall regression model. Analogous to the $F$ statistic, users can also observe the $p$-values (cells E17:E21, Figure 5) associated with each independent variable and compare them to the predetermined level of significance (e.g., 5% or 0.05). If the $p$-value is lower than the predetermined level of significance, then it is determined that the $t$ statistic is significant. Conversely, if the $p$-value is higher than the predetermined level of significance, then it is determined that the $t$ statistic is not significant. Table 6 displays the $t$ statistics and $p$-values for each of the three seasonal models.

It is apparent from Table 4, the time, Q2, and Q3 independent variables are statistically significant but the Q4 independent variable is not statistically significant (i.e., $p$-value=0.5637 and is greater than 0.05) for the quarterly model. For the bimonthly model, each bimonthly variable as well as time is statistically significant. In addition, all independent variables are statistically significant for the monthly model. Overall, each of the seasonal models demonstrates that their unique independent variables are doing a “good” job of predicting U.S. Import Beer sales.
provide additional insight. These coefficients are used to construct the seasonal regression equation itself. Each of the model parameter coefficients is positive for all three of the seasonal models. However, the coefficients are relative to each other within each model. For the quarterly model, it is seen that the Q4 (Oct, Nov, Dec) coefficient (Figure 5, cell B21) is much lower than the coefficients for Q2 and Q3. This is also true for the bimonthly model where the S6 (Nov, Dec) coefficient (Figure 6, cell B23) is much lower than the other coefficients and likewise the Nov, Dec, and Feb coefficients (Figure 7, cells B28, B29, B19) in the monthly seasonal model are lower compared to the other monthly coefficients. Each of these parameters corresponds to the time frame of the Fall into Winter seasons. In turn, it could be tentatively concluded that some type of winter effect is influencing the model. The idea of a “winter” effect is discussed in detail later in this section.

The supposition of this effect comes from what the coefficients denote within the regression model itself. Within regression models, positive coefficients move in unison with changes in the model parameter. For example, if any parameter with a positive coefficient increases, there will be an increase in the predictor, \( \hat{Y}_t \), all things held constant. However, if that parameter’s coefficient is smaller relative to the other parameter’s coefficients, the increase in the predictor, \( \hat{Y}_t \), will be markedly less, all things held constant (i.e., it contributes less, hence the moniker “coefficients of contribution” being applied to the independent variable coefficients).

From our example, it should be noted that the coefficient of the Q4, S6, and the Nov, Dec, and Feb parameters are lower compared to the other coefficients in their respective models; therefore, they provide a damping effect on their predictors. The lower coefficients associated with the Q4, S6, and the Nov, Dec, and Feb parameters are most likely attributable to the holiday lag and temperature drop that occurs every year around November, December, January, and February in the United States. It is well known that beer sales in the United States lag during the colder months of each year and the holiday season (i.e., Thanksgiving, Christmas, and New Years, respectively) due to consumers drinking other beverages, both alcoholic, non-alcoholic, and non-chilled (e.g., think champagne or eggnog around the holidays and Irish Coffee or hot cider during the colder months)(http://www.foodandbeveragereports.com).

According to the analysis any of the proposed seasonal regression models are statistically significant and could be used to effectively develop forecasts for U.S. import beer sales. However, a good manager needs to know which model produces the “best” results. The next section illustrates how to construct the seasonal regression equation for the monthly model, defines and describes error for each model, and discusses model efficacy via error coupled with the regression statistics.

**Construction of the Seasonal Regression Equation and Error Calculation**

From a mathematical perspective the coefficient values are easily inserted into the original regression model to yield the seasonal regression forecasting equation. To illustrate the construction of the seasonal regression equation, the monthly seasonal model output is employed. The following table (Table 7) juxtaposes the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Quarterly Model t-statistic</th>
<th>p-value</th>
<th>Bimonthly Model t-statistic</th>
<th>p-value</th>
<th>Monthly Model t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>5.8654</td>
<td>1.72E-07</td>
<td>time</td>
<td>9.0629</td>
<td>5.80E-13</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>7.4558</td>
<td>2.94E-10</td>
<td>S2</td>
<td>9.7415</td>
<td>4.06E-14</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>5.2936</td>
<td>1.57E-06</td>
<td>S3</td>
<td>13.5102</td>
<td>3.90E-20</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>0.5804</td>
<td>0.5637</td>
<td>S4</td>
<td>11.8738</td>
<td>1.30E-17</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>5.4045</td>
<td>1.10E-06</td>
<td>S6</td>
<td>2.7668</td>
<td>0.0075</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>7.9715</td>
<td>8.61E-11</td>
<td>Mar</td>
<td>2.23E-12</td>
<td>1.30E-17</td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td>8.9429</td>
<td>2.23E-12</td>
<td>May</td>
<td>10.8681</td>
<td>2.05E-15</td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>11.5998</td>
<td>1.61E-16</td>
<td>Jun</td>
<td>10.8650</td>
<td>2.08E-15</td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>9.2102</td>
<td>8.26E-13</td>
<td>Jul</td>
<td>5.3188</td>
<td>1.89E-06</td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td>5.1533</td>
<td>3.45E-06</td>
<td>Sep</td>
<td>3.4771</td>
<td>9.87E-04</td>
<td></td>
</tr>
<tr>
<td>Sep</td>
<td>11.5998</td>
<td>1.61E-16</td>
<td>Nov</td>
<td>3.0587</td>
<td>3.41E-03</td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td>Oct</td>
<td>10.8650</td>
<td>2.08E-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td>5.3188</td>
<td>1.89E-06</td>
<td>Dec</td>
<td>3.0587</td>
<td>3.41E-03</td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td>9.2102</td>
<td>8.26E-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6**

**t Statistics and p-values for Each of the Three Seasonal Models**
independent variables with their corresponding regression coefficients for the monthly seasonal model.

Therefore, the monthly seasonal regression equation can be written as:

\[
\hat{Y} = 1.2093 + 0.0068X_{1t} + 0.1771X_{2t} + 0.5461X_{3t} + 0.6128X_{4t} + 0.7450X_{5t} + 0.7956X_{6t} + 0.7457X_{7t} + 0.6325X_{8t} + 0.3656X_{9t} + 0.3704X_{10t} + 0.2500X_{11t} + 0.2200X_{12t} + 0.3656X_{13t}
\]  

(27)

To demonstrate the seasonal regression equation a forecast for October 2005 will be created. Since the forecasting time period is October of 2005, \(T = 70\) and in turn \(X_{1t} = 70\), \(X_{2t} = 0\), \(X_{3t} = 0\), \(X_{4t} = 0\), \(X_{5t} = 0\), \(X_{6t} = 0\), \(X_{7t} = 0\), \(X_{8t} = 0\), \(X_{9t} = 0\), \(X_{10t} = 0\), \(X_{11t} = 0\), \(X_{12t} = 0\), and \(X_{13t} = 0\). Therefore, the seasonal regression model produces the following result.

\[
Y = 1.2093 + 0.0068 * (70) + 0.3704 * (1) = 2.0543
\]  

(28)

The forecasted U.S. import beer sales for October 2005 are 2.0543 million Brls.

Calculating the error associated with a forecasting model is cornerstone to determining a model’s performance. Coupling the error measurement with the regression statistics allows a forecaster to comment on a model’s overall efficacy. A model that performs well on all statistical measures and carries relatively low error can be deemed adequate. Error, also referred to as deviation, is defined as the actual less the predicted value. Excel can generate error terms for a regression model when prompted.

Referring back to Figure 4, the Regression Dialogue Box, take note of the section labeled Residuals. This section of the Regression Dialogue Box contains check boxes regarding the calculation of what Excel calls a residual, which is synonymous with the aforementioned term error. A forecaster can instruct Excel to automatically calculate the error terms or the residuals of a regression model by checking the Residuals box in the Regression Dialogue Box (refer to Figure 4). When the regression model is computed via the Regression Dialogue Box, the residual output is generated and returned below the regression output. Figure 8 displays a partial example of the residual output for the monthly seasonal regression model within Excel.

A standard measure of error in forecasting is mean absolute deviation or MAD and is calculated as follows:

\[
MAD = \frac{\sum_{i=1}^{n} |actual_{i} - forecast_{i}|}{n}
\]  

(29)

Using the residual terminology from Excel’s regression tool, the formula would resemble the following:

\[
MAD = \frac{\sum_{i=1}^{n} |residual_{i}|}{n}
\]  

(30)

To obtain MAD for the residuals contained in Figure 8, a user would first take the absolute value of each residual in the residual column (e.g., column C, Figure 8) and then take an average of those absolute residuals. Forecasters generally compute error for multiple forecasting models and then compare across each model where the model with the lowest error is considered superior. The MAD for each of the seasonal models is presented in Table 8.

From Table 3 the monthly seasonal regression model has the lowest MAD of the three models. This further reinforces the notion that the monthly seasonal model does the best job of describing the data. One could say that when taking into consideration both the statistical and error measures that the monthly seasonal model is the best performing model of the three.
Another means to illustrate how the seasonal regression model fits the historical data is a plot of the predicted sales juxtaposed with the historic sales. This is accomplished using Excel’s Chart Wizard button or by accessing the Chart option in the Insert menu. For the data set, the most efficient manner in developing the plot is using the Chart option in the Insert Menu. The following steps can be employed to create the plot in Figure 6, a graph of the monthly seasonal model predicted sales and historical sales.

Step 1: Highlight cells D4:D73 and cells P4:Q73, Figure 1.1 (c) (this can be done by highlighting the data in column D first then pressing and holding the Ctrl key down while highlighting the data in column J and K)

Step 2: Choose Chart from the Insert menu in Excel

Step 3: From the Chart Wizard Step 1 of 4 – Chart Type screen choose the Standard Types menu tab and then choose the XY (Scatter) option from the scroll box

From here it is up to the user’s own preferences how they proceed through Steps 2, 3, and 4 of the Chart Wizard option screens. The authors have added a chart title, x and y axis monikers, renamed and moved the legend to the bottom of the plot, and adjusted the x axis to reflect a three month scale.

It is readily apparent from Figure 8 that the seasonal regression model fits the actual data very well by accounting for the seasonal trends in import beer sales over time. The figure also visually reinforces the calculated value of $R^2$ (.88) presented earlier. Based on the regression statistics, MAD, and the Predicted versus Actual graph (Figure 9), it can be determined that the monthly seasonal model is the superior model of the three models presented and overall is doing a good job of forecasting U.S. Import Beer Sales.

### Conclusions, Recommendations, and Limitations

Seasonality is found in almost every facet of marketing, from products to placement to promotion to price. Virtually every product and service has a seasonal component whether it is caused by weather, holidays, or supply and demand. Common examples of seasonal impacts include the demand for heating oil in the winter, fresh fruits and vegetables in season, or the need for additional inventories to meet increased consumer de-

<table>
<thead>
<tr>
<th>Seasonal Model</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly</td>
<td>0.1469</td>
</tr>
<tr>
<td>Bimonthly</td>
<td>0.0981</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.0872</td>
</tr>
</tbody>
</table>

| Table 8 Seasonal Regression Model MAD Comparisons |

<table>
<thead>
<tr>
<th>Seasonal Model</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly</td>
<td>0.1469</td>
</tr>
<tr>
<td>Bimonthly</td>
<td>0.0981</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.0872</td>
</tr>
</tbody>
</table>
mand during the holiday season. As a result businesses must be prepared to adjust their forecasts and ultimately their operations in order to meet changing demand.

It can be seen from the example and subsequent model that Excel can accommodate seasonal data fairly easily. Although effective in the development of seasonal models, it must be noted that Excel does lack certain sophisticated statistical methods common to specialized forecasting software packages. However, this does not diminish the fact that Excel can and should be used to model seasonal patterns in data. This work illustrated that Excel is ubiquitous within industry as the forecasting standard and across PC software suites. In addition, it is widely employed by business managers for their everyday forecasting needs. These facts coupled with the fact that the vast majority of educators employ Excel within the classroom as the standard for their forecasting curriculum are substantial enough reasons to lend credence to it use.

The authors believe that critical thinking is paramount in the forecasting context and especially when seasonality is present. It has been said that when students develop a forecast with easy-to-use software, they do not take the time to assess the reasonableness of the forecast or the evaluate what the underlying model is doing or even question if the model was appropriate to begin with. Still others believe that students have a tendency to blindly accept the results of a computer forecasting model without evaluating it (also see Hunt, Eagle, and Kitchen 2004). However, as this work has shown, it is not difficult for a spreadsheet user to overcome the mentality of “whatever comes out of the model must be correct” and that students can be taught critical thinking regarding forecasting models.

The authors suggest to marketing educators the following areas to focus on when imparting upon students the intricacies of seasonal forecasting models:

- Is the forecast performing well based on the historical pattern?
- Is the model being used appropriate for the forecasting problem?
- Is there supporting information that should accompany the forecast to help the user better understand the forecast?
- Are the final results from the model not only realistic but are they believable and not “black box” type of work?

Not only is this a call for a “reality check,” but it also raises the “black box” issue in forecasting. Many business people have expressed confusion about how forecasts are derived (i.e., they come from the magical “black box”) and need reassurance that the numbers they are being given aren’t just something spit out of a complicated yet unseen device (i.e., the “black box”). The authors believe...
that by providing the model in a common venue, a spreadsheet, seasonal forecasting done in Excel can put clients more at ease with the model as well as the results.

Limitations of this research could include a faculty member’s own lack of computer skills which could inhibit the quality its presentation. There is also the natural reluctance to add yet another assignment to an already full schedule, especially when it requires more time and effort in terms of preparation and grading. Another limitation arises from the issue of time. Most faculty feel pressed to cover all of the material in a given text and may be reluctant to spend more time on a topic necessitating the removal of another.

Ultimately, the decision of whether or not to adopt active learning technologies such as teaching students how to develop seasonal forecasting models with Excel is one that requires careful consideration. On one hand, instructors want their students to be prepared for the life in the business world, and on the other as a member of a larger College of Business we must comply with accreditation mandates. However, the authors sincerely believe that most faculty within Colleges and Schools of Business possess the needed computer skills and drive to create and deliver seasonal regression models in Excel.

References


Data Source
