Forecasting Computer Usage

Julie M. Hays
University of St. Thomas

Journal of Statistics Education Volume 11, Number 1 (2003),
ww2.amstat.org/publications/jse/v11n1/datasets.hays.html

Copyright © 2003 by Julie M. Hays, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Causal forecasting; Model-building; Seasonal Variation; Simple linear regression; Time-series forecasting; Transformations.

Abstract

The dataset bestbuy.dat.txt contains actual monthly data on computer usage (Millions of Instructions Per Second, MIPS) and total number of stores from August 1996 to July 2000. Additionally, information on the planned number of stores through December 2001 is available. This dataset can be used to compare time-series forecasting with trend and seasonality components and causal forecasting based on simple linear regression. The simple linear regression model exhibits unequal error variances, suggesting a transformation of the dependent variable.

1. Introduction

One of the most prevalent uses of regression analysis in actual business settings is for forecasting. For a summary of some forecasting methods see Armstrong (2001) or Arsham (2002). The bestbuy.dat.txt dataset can be used to demonstrate and discuss both time-series and causal forecasting. Time constraints and the interests and needs of the students determine whether I supply the analyses or have the students perform the analyses.

I have used this dataset throughout the semester in an MBA Decision Analyses class. This class is a core requirement for all evening MBA students and covers a range of decision analysis and statistical topics, including regression analysis and forecasting. Most students are required to take an introductory business statistics course prior to this course, so they have had some exposure to statistical topics, but few students have any academic experience with forecasting.

Best Buy Co., Inc. (NYSE:BBY), headquartered in Eden Prairie, Minnesota, is the largest volume specialty retailer of consumer electronics, personal computers, entertainment software and appliances. In August of each year, Best Buy purchases mainframe MIPS (Millions of Instructions Per Second, a measure of computing resources) in anticipation of the coming holiday season. Computing resources are needed to track and analyze retail information needed for billing, inventory, and sales. For planning and
budgeting purposes they also wish to forecast the number of MIPS needed the following year. Best Buy Corporation actually used this dataset to predict computer usage in order to budget for and purchase an appropriate amount of computing power. However, prior to 2001, Best Buy did not do any statistical analysis of this data. Best Buy only looked at the numbers (they did not even graph the data) and then guessed at the amount of MIPS needed in the coming year.

Students are asked to forecast the MIPS needed for December 2000 and December 2001 using the bestbuy.dat.txt dataset. This dataset was obtained from the Best Buy Corporation and contains monthly data on computer usage (MIPS) and total number of stores from August 1996 to July 2000. Additionally, information on the planned number of stores through December 2001 is available.

Students can easily understand the seasonality that retail operations experience. Best Buy Corporation has experienced significant growth over the past few years and most students understand that as a firm grows, their need for computing power also increases. Therefore, this dataset can be used to demonstrate time-series forecasting with both a trend and seasonality.

This dataset can also be used to demonstrate causal forecasting based on simple linear regression of computer usage and number of stores. The simple linear regression model exhibits unequal error variances, suggesting a transformation of the dependent variable.

Finally, a comparison between the time-series model and causal model can be made and discussed with the students.

2. Time Series Forecasting

Before I allow the students to begin any numerical analyses, I have the students plot computer usage versus time. I have the students “forecast” the number of MIPS needed for December 2000 and December 2001 using only the plot of computer usage (MIPS) versus time, Figure 1. The plot clearly shows a trend in MIPS usage with time. Typically, students “eyeball” the graph and predict MIPS usage of 500 for December 2000 and 600 for December 2001.
Figure 1. MIPS vs Time.

Students who actually fit a line to the data forecast MIPS usage of 527 for December 2000 and 624 for December 2001 (Figure 2).
I introduce simple moving average, weighted moving average and exponential smoothing forecasting techniques to the students before they attempt to use these forecasting models to predict future MIPS usage. I also discuss the evaluation of forecasting models using MAD and CFE (explained below). The interested reader can find more detailed discussions of these topics in Stevenson (2002) or at Sparling (2002).

**Moving Average**
An \(n\)-period moving average is the average value over the previous \(n\) time periods. As you move forward in time, the oldest time period is dropped from the analysis.

\[
\text{Simple Moving Average} = \frac{\sum \text{value in the previous } n \text{ periods}}{n}
\]

**Weighted Moving Average**
An \(n\)-period weighted moving average allows you to place more weight on more recent time periods by weighting those time periods more heavily.

\[
\text{Weighted Moving Average} = \frac{\sum \text{weight for period} \times \text{value in period}}{\sum \text{weights}}
\]

**Exponential Smoothing**
The forecast for the next period using exponential smoothing is a smoothing constant, \(\alpha\) (\(0 \leq \alpha \leq 1\)), times the demand in the current period plus (1- smoothing constant) times the forecast for the current period.

\[
F_{t+1} = \alpha D_t + (1-\alpha)F_t
\]

where \(F_{t+1}\) is the forecast for the next time period, \(F_t\) is the forecast for the current time period, \(D_t\) is the demand in the current time period, and \(0 \leq \alpha \leq 1\) is the smoothing constant. To initiate the forecast, assume \(F_1 = D_1\). Higher values of \(\alpha\) place more weight on the more current time periods.

Because this model is less intuitive, I usually expand this equation to help the students understand that demand from time periods prior to the current period is included in this model.

\[
F_{t+1} = \alpha D_t + (1-\alpha)[\alpha D_{t-1} + (1-\alpha)F_{t-1}]
\]

and

\[
F_{t+1} = \alpha D_t + (1-\alpha)[\alpha (D_{t-1} + (1-\alpha)D_{t-2} + (1-\alpha)(F_{t-2})]
\]

where \(D_{t-1}\) is the demand in the previous time period, \(D_{t-2}\) is the demand in the time period before the previous time period, and \(F_{t-1}\) is the forecast in the previous time period, and \(F_{t-2}\) is the forecast in the time period before the previous time period.

Because the data storage requirements are considerably less than for the moving average model, this type of model was used extensively in the past. Now, although data storage is not usually an issue, it is typical of real-world business applications because of its historical usage.

**Mean Absolute Deviation (MAD)**
The evaluation of forecasting models is based on the desire to produce forecasts that are unbiased and accurate. The Mean Absolute Deviation (MAD) is one common measure of forecast accuracy.
Cumulative sum of Forecast Errors (CFE)
The Cumulative sum of Forecast Errors (CFE) is a common measure of forecast bias.  
\[
\text{CFE} = \sum (F_t - D_t)
\]

“Better” models would have lower MAD and CFE close to zero.

After explaining these techniques, I have the students work through the following simple example in class. I give the students the demand profile (Table 1) and have them calculate forecasts using a 3-period moving average and exponential smoothing with a smoothing constant of 0.2. I also have them calculate the MAD and CFE for both models. We discuss using the MAD and CFE to determine the “best” model.

I also point out to the students that I have arbitrarily chosen the number of periods for the moving average model and the smoothing constant for the exponential smoothed model. I discuss using MAD and CFE to determine the “best” choice for these variables.

---

Table 1. In-class forecasting example.

| Time Period | Demand | Forecast, 3-Period Moving Average | Forecast, Exponential Smoothed In-Class Forecasting Example $\alpha = 0.2$ | Deviation, 3-Period Moving Average | Deviation, Exponential Smoothed | Absolute Deviation, 3-Period Moving Average | Absolute Deviation, Exponential Smoothed |
|------------|--------|----------------------------------|--------------------------------------------------------------------------|-----------------------------------|--------------------------------------|---------------------------------------------|---------------------------------------------|
| 1          | 4.1    | 4.10                             |                                                                          |                                   |                                      |                                             |                                             |
| 2          | 3.3    | 4.10                             |                                                                          |                                   |                                      |                                             |                                             |
| 3          | 4.0    | 3.94                             |                                                                          |                                   |                                      |                                             |                                             |
| 4          | 3.8    | 3.80 = $(4.1 + 3.3 + 4.0) / 3$   | 3.95 = $(0.2*4.0) + (0.8*3.94)$                                         | 0.00 = 3.80 - 3.80                | 0.15 = 3.95 - 3.80                    | 0.00 = $|0.00|$                                 | 0.15 = $|0.15|$                                |
| 5          | 3.9    | 3.70                             | 3.92                                                                     | -0.20                            | 0.02                                 | 0.20                                       | 0.02                                       |
| 6          | 3.4    | 3.90                             | 3.92                                                                     | 0.50                             | 0.52                                 | 0.50                                       | 0.52                                       |
| 7          | 3.5    | 3.70                             | 3.81                                                                     | 0.20                             | 0.31                                 | 0.20                                       | 0.31                                       |
| 8          | 3.7    | 3.60                             | 3.75                                                                     | -0.10                            | 0.05                                 | 0.10                                       | 0.05                                       |
|            |        |                                  |                                                                          |                                   |                                      |                                             |                                             |
| CFE        |        | 0.40 = 0.00 - 0.20               |                                                                          |                                   |                                      |                                             | 1.06                                       |
Once the students are familiar with these techniques, I have them estimate MIPS for December 2000 and 2001 using a 3-period moving average and exponential smoothing with a smoothing constant of 0.2 (Figure 3). This can be done using Excel, Minitab or any statistics package. The forecast for the 3-period moving average is 463 MIPS and for the exponential smoothed is 450 MIPS.

![Figure 3. Actual and forecast MIPS.](image)

The students can easily see that there is a “problem” with their forecasts. Although I have told the students that exponential smoothing and moving average forecasting models are only appropriate for stationary data, they don’t really understand this until they try to use the technique. This exercise helps the students understand that moving average and exponential smoothing are really only averaging techniques and helps
them comprehend the need to account for trends in forecasting. I demonstrate adjusting for trends by using double exponential smoothing. Double exponential smoothing is a modification of simple exponential smoothing that effectively handles linear trends. Good explanations of this technique can be found in Wilson and Keating (2002) or at Group6 (2002).

**Double Exponential Smoothing**

\[ F_{t+1} = A_t + T_t \]

where \( F_{t+1} \) is the forecast for the next time period,

\[ A_t = \alpha (D_t) + (1 - \alpha) F_t \]

where \( F_t \) is the forecast for the current time period, \( D_t \) is the demand in the current time period, and \( 0 \leq \alpha \leq 1 \) is the smoothing constant and \( T_t \) is the exponentially smoothed trend component in the current period.

\[ T_t = \gamma (C_t) + (1 - \gamma) T_{t-1} \]

where \( 0 \leq \gamma \leq 1 \) is the smoothing constant for the trend, \( T_{t-1} \) is the trend in the previous period, and \( C_t \) is the current trend

\[ C_t = A_t - A_{t-1} \]

The forecast for \( n \) periods into the future is

\[ F_{t+n} = D_t + nT_t \]

To initiate the forecast, assume \( F_1 = D_1, C_1 = T_1 = D_2 - D_1, \) and \( A_1 = D_1 - T_1 \).

After I explain this model, I have the students go back and re-estimate their forecast using this model (Figure 4). Minitab has these functions built in and will compute the optimal smoothing parameters, \( \alpha \) and \( \gamma \), based on minimizing the sum of squared errors, but any statistics package could be used. Minitab will also compute Mean Absolute Prediction Error (MAPE), Mean Absolute Deviation (MAD), Mean Square Error (MSE) and provides 95% confidence prediction intervals (see Figure 4).

The forecasts obtained are essentially the same as the forecasts obtained from fitting a line to the data, MIPS usage of 527 for December 2000 and 624 for December 2001.
Figure 4. Optimal double exponential smoothed.

I ask the students if they are happy with their forecast now or if there is anything else they need to do. I supply a plot of errors versus time for the double exponential smoothed model with the December errors highlighted (Figure 5). Most students are aware that retail firms have their highest sales during the Christmas season (December). Therefore, students typically mention seasonality and we discuss the possible ways that we could account for seasonality.
The students usually mention both an additive and multiplicative adjustment for seasonality using all the past data or only some of the past data. Simple explanations of these two techniques can be found in Hanke and Reitsch (1998) or at Nau (2002). In other words, we could compare the forecast for December 1999 to the actual for December 1999 and for the additive model we would add this difference to our forecast for December 2000. Or, for the multiplicative model, we would multiply the forecast for December 2000 by the actual December 1999/forecast December 1999. They carry this further and discuss using the data from 1998, 1997, and 1996 to produce an average adjustment. I lead the discussion towards the smoothing techniques we have been discussing and how we could use these types of techniques to come up with seasonal adjustments for our forecasts. I explain that Winter developed just such a technique of triple exponential smoothing. Winter’s technique basically adds (or multiplies) a smoothed seasonal adjustment to the model, similar to the addition of a smoothed adjustment for a trend in the double exponential smoothed model. The interested reader can find the calculation formulas and explanations of triple exponential smoothing (or Winter’s method) in Minitab (1998b) or Prins (2002a).

I use Minitab to demonstrate Winter’s model (Figure 6) because the calculations for this method are fairly complex and most students only need to have a general understanding of this type of technique. Using Winter’s model the forecast for December 2000 is 521 MIPS and the forecast for December 2001 is 606 MIPS.

I also use this opportunity to mention ARIMA models and direct interested students to resources such as Minitab (1998a) for more information about ARIMA models.

Figure 5. Double exponential smoothed model errors.
3. Causal Forecasting

I supply the students with a plot of computer usage (MIPS) vs. number of stores (Figure 7) and again have them “forecast” computer usage for December 2000 and December 2001. Best Buy believes that they will have 394 stores in December of 2000 and 445 stores in December of 2001.
Figure 7. MIPS vs number of stores.

Again most students “eyeball” the graph and use graphical linear extrapolation to arrive at their forecast. They predict usage of 600 MIPS for December 2000 and 800 MIPS for December 2001.

I have the students perform a simple linear regression of MIPS on number of stores and produce the residual plot (Figures 8 and 9). I use this opportunity to emphasize the usefulness of the residual plot in evaluating the model. I highlight the “megaphone” shape of the residual plot (the residuals are increasing as the number of stores increases) and explain that this implies that a transformation of the dependent variable is indicated.
Although I used the Box-Cox procedure (Box and Cox 1964) to determine the appropriate transformation, this technique is beyond the scope of this class. Therefore, I just tell the students that the appropriate
transformation is square root of MIPS and mention that there are mathematical techniques that can be used to determine the appropriate transformation. I direct interested students to Neter, Kutner, Nachtsheim, and Wasserman (1996) or Prins (2002b) for descriptions of this technique.

I have the students re-estimate the regression equation and produce the residual plot for this regression (Figures 10 and 11). Although the $R^2$ is slightly lower, the residuals are now more evenly distributed.

![Figure 10. Square root MIPS vs number of stores.](image)

$y = 0.1064x - 16.130$

$R^2 = 0.9566$
I also have the students predict computer usage for December 2000 and December 2001 using the fitted equation. If students have difficulty predicting MIPS, because of the square root transformation of MIPS, I explain the calculations in class. The new predictions are 664 MIPS for December 2000 and 977 MIPS for December 2001.

Again, an adjustment for seasonality could be made. Although, any of the seasonality adjustments discussed in the previous section could be used here, I usually have the students use an average multiplicative adjustment. This could be done by calculating actual/predicted for all months, averaging these seasonal factors for each particular month and multiplying the resulting seasonal factor by the predicted value. If this is done, the new predictions for December 2000 and December 2001 are 700 MIPS and 1029 MIPS.

4. Comparison of Methods

After the students have used the various methods to predict MIPS usage, I have them discuss which method they have most confidence in and why they believe that that model is the best. Several important points can be made here.

First, I emphasize that forecasting is a very imperfect science and no technique can perfectly predict the future. The “best” technique will balance the accuracy needed with the complexity (or cost) of the model.

Second, I emphasize the value of “plotting the data.” One of the best (and easiest) methods to evaluate various models is a visual examination of the data and forecasts that would be produced by the method under consideration.

Third, I emphasize the need to account for trends and seasonality if those are present in the data. Moving averages and exponential smoothing are appropriate forecasting methods only if the data are stationary. If
there are trends and/or seasonalities present, more sophisticated methods should be used.

Finally, we discuss the difficulty inherent in finding a causal predictor for most values we wish to predict in business environments.

5. Conclusion

The dataset bestbuy.dat.txt can be used to demonstrate both time series and causal forecasting. Analysis of the dataset leads to a discussion and comparison of the positives and negatives of various forecasting methods.

6. Obtaining the Data

The file bestbuy.dat.txt contains the raw data. The file bestbuy.txt is a documentation file containing a brief description of the dataset.

Appendix - Key to the variables in bestbuy.dat.txt

| Columns | Description |
|---------|-------------|
| 1 - 11  | Date dd-mm-yyyy, August 1996 - July 2000 |
| 12 - 19 | MIPS usage (MIPS are a measure of computing resources) |
| 20 - 30 | Number of stores |
| 31 - 44 | Date dd-mm-yyyy, July 2000 - December 2001 |
| 45 - 53 | Planned Number of Stores through December 2001 |

Values are aligned and delimited by blanks.

References

Armstrong, J. (ed.) (2001), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Hingham, MA: Kluwer Academic Publishers.

Arsham, H. (2002), "Time Series Analysis and Forecasting Techniques" [Online]. (ubmail.ubalt.edu/~harsham/stat-data/opre330Forecast.htm)

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, 26, 211-243.

Group6 (2002), "Estimation and Forecasting Tutorials" [Online]. (www.cs.usask.ca/resources/tutorials/csconcepts/1999_6/Tutorial/index.html)

Hanke, J. E., and Reitsch, A. G. (1998), *Business Forecasting*, (6th ed.), Upper Saddle River, NJ: Prentice Hall.
Minitab (1998a), "Fitting an ARIMA model," in *Minitab Help*, State College, PA: Minitab, Inc.

Minitab (1998b), "Level, Trend, and Seasonal Components - Winters' Method", in *Minitab Help*, State College, PA: Minitab, Inc.

Nau, R. F. (2002), "Decision 411 Forecasting: Seasonal Adjustment" [Online].
(www.duke.edu/~rnau/411seas.htm)

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996), *Applied Linear Statistical Models*, (4th ed.), Chicago, IL: Times Mirror Higher Education Group, Inc.

Prins, J. (2002a), "6.4.3.5. Triple Exponential Smoothing," *NIST/SEMATECH e-Handbook of Statistical Methods* [Online]. www.itl.nist.gov/div898/handbook/pmc/section4/pmc435.htm

Prins, J. (2002b), "6.5.2. What do we do when the data are 'non-normal'"? *NIST/SEMATECH Engineering Statistics Internet Handbook* [Online]. www.itl.nist.gov/div898/handbook/pmc/section5/pmc52.htm

Sparling, D. (2002), "Forecasting" [Online]. (www.uoguelph.ca/~dsparlin/forecast.htm)

Stevenson, W. J. (2002), *Operations Management*, (7th ed.), New York: McGraw-Hill.

Wilson, J. H., and Keating, B. (2002), *Business Forecasting w/ForecastX*, (4th ed.), Boston: McGraw-Hill.

Julie M. Hays
Graduate School of Business
University of St. Thomas
Minneapolis, MN 55403-2005
USA
jmhays@stthomas.edu