Long Range Acquisition Tool

Problem Statement

United States Code Title 15, Section 637(A)(12)(C), requires the Department of the Navy (DoN) to prepare a forecast of expected contract opportunities for the next several fiscal years and make the forecast available to small businesses. This expected volume should be based on historical data. Prior to this report, there was no established method of providing this projection of out year business volumes.

Data

The data available while building the projection tool was limited to FYs 2006, 2007, 2008 and 2009. The data was provided in the form of excel workbooks, one workbook per FY. The Small Business Program Office (SBPO) requested that output be in an excel format.

Methodology

In order to maintain simplicity and flexibility for use of the projection tool in future years, the data provided was used to establish visual basic for applications (VBA) modules that could be used by the SBPO. The VBA modules employ excel simple linear regression functions (LinEst) as the basis for predictions. The user defines number of out-years and the number of standard errors they wish to use to build the prediction interval.

LinEst is used for the premise of the prediction. Other statistical regression analysis could be employed. However, the data provided has a single X and a single Y variable. X is defined as the year a certain NAIC was contracted for, while Y is the total value spent towards that NAIC during the year. There are other variables that could impact the amount spent any year, but no other data is collected.

The user is responsible for providing the number of years they are interested in predicting. The number of years is simply how many out years the user would like to advertise an estimate for future business volume. This number will be collected by the VBA code and a prediction will be made using the linear regression model of the data.

NAICs are up to six digits in length. If fewer than six digits are used, the resulting population is still related but denser. The user has the opportunity to provide any NAICs data set prior to executing the VBA code. The NAICS worksheet is used to define the NAICs and their descriptions.

Finally, the user must decide how much error he would like to include in published predictions. The standard error is a measure of the amount of error in the prediction of Y for an individual X. If, for example, the next years' actual volumes were $10, $50, $20, 36.67 for the years 2006, 2007, 2008 and 2009 the standard error for the Y-estimate would be 28.5. If a linear regression is done on these numbers, a prediction for the next year, in this case 2009 would be $36.67. If the user chose to use 3 standard errors, the prediction interval would be $36.67 – 3*$28.50 to $36.67 + 3*$28.50 or (-$48.83 to $122.17). The VBA script will automatically change negative values to zero making the prediction interval ($0 –

$122.17). This allows the published prediction to account for errors that are in the prediction as a result of the myriad of unknown and uncontrolled variables affecting the overall business volume.

The user should understand the use of the standard error of the estimate. Figure 2 graphically shows the effects of adding standard error to your estimate. Simply put, the more standard errors used, the greater the chance the actual transactions for the coming year will fall within your prediction interval. However, the range of your interval grows and may become less useful. The same logic applies to fewer standard errors used: as the number of standard errors decreases, the smaller your prediction interval and the larger the possibility your actual numbers will not fall within the interval.

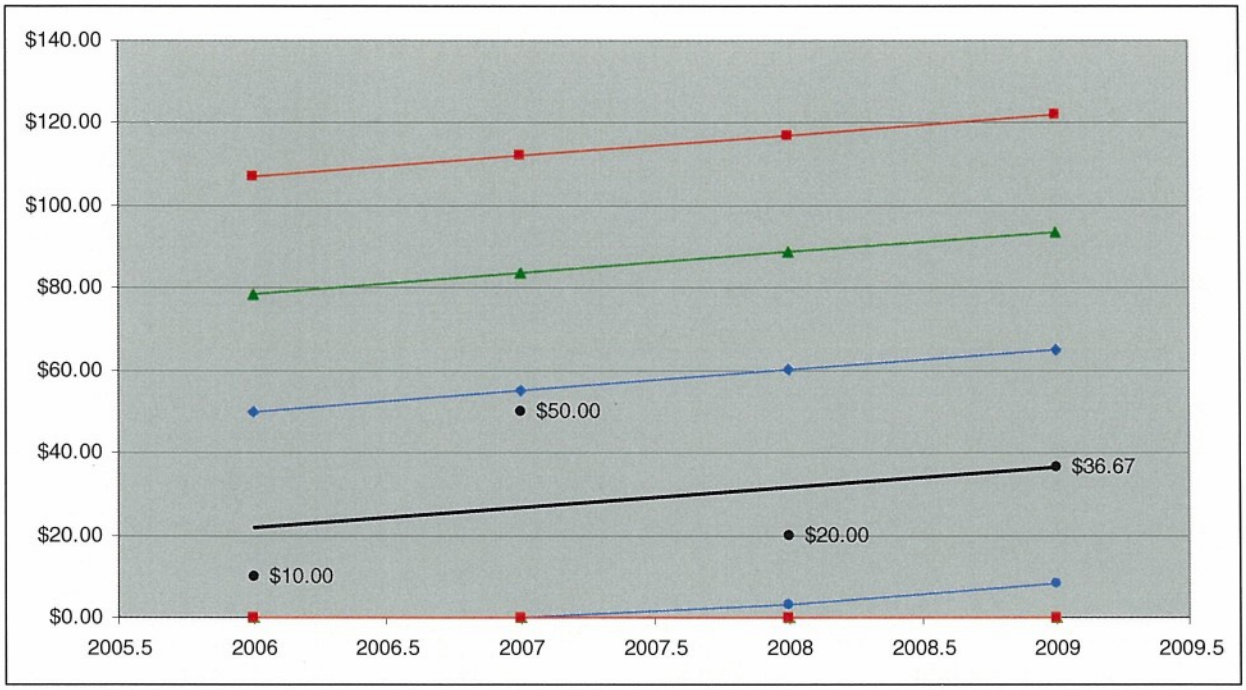

Figure 1: The black dots represent actual values as described in the example. The black line is a trend line with the FY09 out year captured. Finally, the blue, green and red lines are +/- 1, 2, and 3 Standard Errors for the Y estimate.

Assumptions

- The prediction made is simply a planning factor and being off will not cause penalty.
- The data provided to the model is accurate and correctly formatted.
    - Each historical year is saved in a folder named 'Data' within the same folder as the Long Range Acquisitions Forecast (LRAF) tool.
    - Each historical year is contained in a single excel workbook saved in a 'yyyy' format

- Each row of each workbook contains data
- The following column headers exist: 'NAICS Code', 'Action Obligation', and 'Actions'

- The user will want each NAIC treated in the same manner.
- Standard Error is assumed rather than Standard Deviation because some gaps in data were discovered during the development of the tool.

Conclusions

The prediction tool produced provides a reasonable estimate based on the available data using simple linear regression techniques resident in Microsoft Excel. The prediction will become more accurate as more data is added over time.