THE NRO CAAG CER ANALYSIS TOOL

DONALD MACKENZIE

MacKenzie Consulting, Inc.
320 Mockingbird Circle, Santa Rosa CA, 95409
(707) 538-3179 dwmackenz@comcast.net

ABSTRACT

The NRO Cost and Acquisition Assessment Group (NRO CAAG) uses its CER Analysis Tool, CERAT, to support CER development. CAAG CERs are based on actual costs and technical data from NRO, Air Force and NASA space programs. The CAAG uses Ordinary Least Squares with log-transformed variables (LOLS) and Zero Bias, Minimum Percent Error (ZMPE) methods in fitting CER equations to statistical data.

The primary purpose of CERAT is to analyze CERs for sensitivity to influential data points (IDPs). An "important" target data point (TDP) in the CER data set is established first – usually the point with the largest estimated cost. Next CERAT systematically removes data points from the CER data set, one at a time, and refits the remaining data with each best-fit method.

The impacts (percent changes) of data point removals on estimated TDP costs are used to identify IDPs. Comparisons of impacts from previous CER developments help decide when to remove a data point. The changes in recorded values of CER constants when data points are removed help in assessing overall CER "stability."

CERAT provides several other aids to CER developers for each best-fit method: XY graphics for each continuous independent variable (IV), an actual cost vs. estimated cost graph, errors plotted against each IV (linear and log domains), linear and log error histograms, and Cook's Distance for each data point, modified for constant percent error models. As such, CERAT has become a CAAG standard for assessing candidate CERs.

Summary

The National Reconnaissance Organization (NRO) Cost and Acquisition Assessment Group (CAAG), among other things, is responsible for developing cost estimates for NRO systems. It collects actual costs and technical data for space systems and develops Cost Estimating Relationships (CERs) from these data to support its estimating process. The CERs relate costs for space hardware; space and ground and software; system engineering, integration and test and program management (SEIT/PM); and support

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

equipment to system technical attributes such as hardware weight, software source lines of code and amount of new design. Some CERs also relate system engineering and program management (SEPM) to lower-level cost totals, primarily those for hardware and software.

Most CAAG CERs are developed using regression analysis methods applied to sets of data for similar types of hardware boxes, software CSCIs, etc. The CAAG uses a tool known as the CER Analysis Tool, CERAT, to support regression-based CER developments. The primary purpose of CERAT is to analyze candidate CERs for influential data points (IDPs) within their data sets. However, CERAT also produces other useful information for evaluating candidate CERS, some of which is unique to the CAAG.

The origin of CERAT is described in the first section below. Next, the CERAT process for identifying IDPs and quantifying the magnitudes of their impacts is described, including example results for a hypothetical CER data set. This is followed by descriptions of the other types of CER assessment information generated by CERAT. The last section summarizes CAAG experience with CERAT.

Background

CERAT stems from a 2011 CAAG study focused on influential data points for CERs with one independent variable (cost driver).[1] CER data sets consisting of N data points, each with a dependent variable, Y, and an independent variable, X, were created using Monte Carlo sampling methods. The Monte Carlo sampling was designed to produce sets of X-Y data representative of actual CAAG CER data sets.

The underlying sampling equation in the IDP study had the form $Y = AX^B$, where both X and Y had lognormal sampling distributions. The CAAG has substantial statistical support for the use of lognormal distributions in simulating both X and Y CER variables.

A wide range of "analysis cases" were analyzed in the study. A given case had 200 sampled X-Y data sets, each with N data points. Each case also had specific values of the exponent B and the level of Y dispersion, as controlled by the log standard error (LSE) used in the sampling process.

The regression best-fit methods used in the study were:

- Log-transformed Ordinary Least Squares (LOLS)
- Minimum Unbiased Percent Error (MUPE -- also known as IRLS, Iteratively Reweighted Least Squares)
- Zero-Bias Minimum Percent Error (ZMPE), and
- Minimum Average Absolute Percent Error (AAPE)

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

The first three methods are widely used in the cost estimating community, while AAPE is not. However, its potential advantage over MUPE and ZMPE is that the percent error is not squared when forming the objective function. This would presumably reduce AAPE's sensitivity to data points with large errors, as does the LOLS method.

IDP Impact Definitions

The way of expressing the impact of an influential data point that was used in the study is shown in Figure 1 on the next page. The solid green line in the figure is the regression line derived from all data points in the X-Y regression data set. The gold line is the "exact" CER equation from which the X-Y samples were generated

When a point was removed from the data set, the remaining data points were refitted, yielding a new regression line. The dashed green line in the figure represents the case where the most influential data point (1st IDP) was removed from the data set.

In this study, the impact of removing a data point is expressed as the percent movement in the Y estimate for the data point with the largest X value in the CER data set (the target data point, or TDP). The movement is indicated by the ΔY_1 in Figure 1, where "1" denotes the most influential data point (1st IDP). It is the difference between the CER estimate for the TDP with a single data point removed, YE_1, and the baseline CER estimate, YE_{BL}, with all data points used to derive the CER equation.

The impact, DY_1, is simply the ratio of ΔY_1 to YE_{BL}:

$$DY_1 = (YE_1 - YE_{BL}) / YE_{BL}$$

A negative value for DY represents a downward movement of the estimate.

The DY impacts of IDPs are typically greater for CERs with few data points and high dispersion. "Normalized" IDP impacts (NDYs) were calculated by dividing actual DY values by the CER Standard Percenr Error, SPE:

$$NDY_1 = DY_1/SPE,\ NDY_2 = DY_2/SPE, \text{ etc.}$$

Normalized impacts for LOLS and MUPE are almost contant in SPE (i.e., independent of SPE). For ZMPE and AAPE, normalized impacts incease mildly with increasing SPE.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

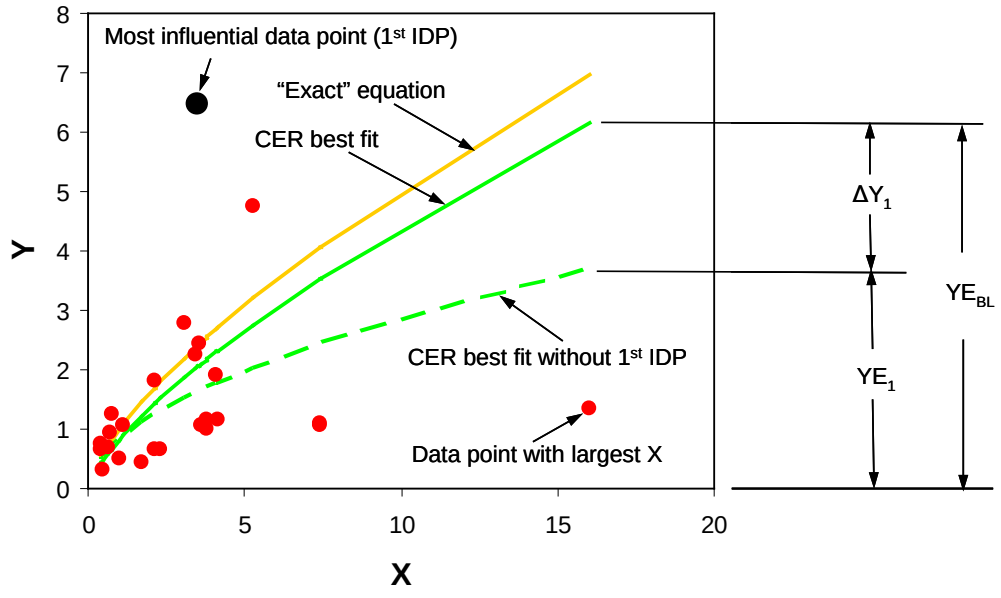

Figure 1. IDP Impact Definition

Normalized impacts are the primary indicator of how sensitive a given CER is to influential data points, although they are still influenced strongly by the number of CER data points. Thus NDY values should only be compared across CERs with about the same number of data points.

Influential Data Point General Behavior

The following summarize the findings of the influential data point study. They are repeated here to serve as a reference for CER developers.

- LOLS and MUPE have about the same average IDP impact
- LOLS and MUPE are less sensitive to IDPs than ZMPE and AAPE
  - ZMPE NDY impacts average 33% higher than LOLS and MUPE over all 26 analysis cases in the study (17% minimum, 78% maximum)
  - AAPE NDY impacts average 55% higher than LOLS and MUPE
- Impacts decrease dramatically with increasing number of data points
- Impacts increase moderately with SPE
- IDP Impacts are not sensitive to exponent B
- LOLS and MUPE have the same IDP 60-80% of the time
- All methods have the same IDP 15-30% of the time

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

The histograms in Figure 2 reveal the locations of IDPs were along the normalized X axis (normalized X = actual X divided by the maximum X value in the CER data set). While a lot of the IDPs for all methods are the data points with the largest X values, ZMPE and AAPE are also sensitive to data points with large positive errors near the low end of the X range. Although this my seem conterintuitive, it makes sense since both ZMPE and AAPE try to minimize large percent errors.

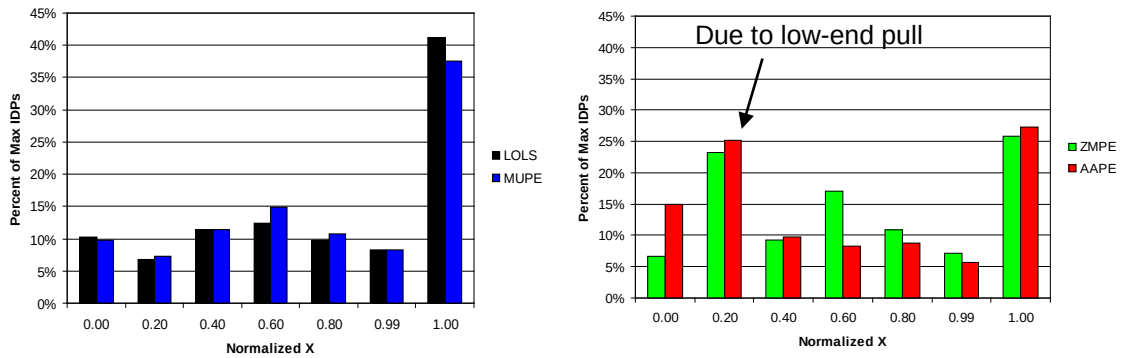

Figure 2. Frequency Histograms of IDPs vs. Normalized X

When a data point with a <u>low</u> value of X has a large percent error, ZMPE (or AAPE) works real hard to reduce it by adjusting the CER constant A and exponent B. The result is A increases and B decreases, significantly lowering the CER estimates at <u>high</u> X values. Thus large negative DY values result as the "tail [low end data point] wags the dog [estimate for a high value of X]." This behavior is characterized as "low-end pull." This is illustrated in Figure 3.

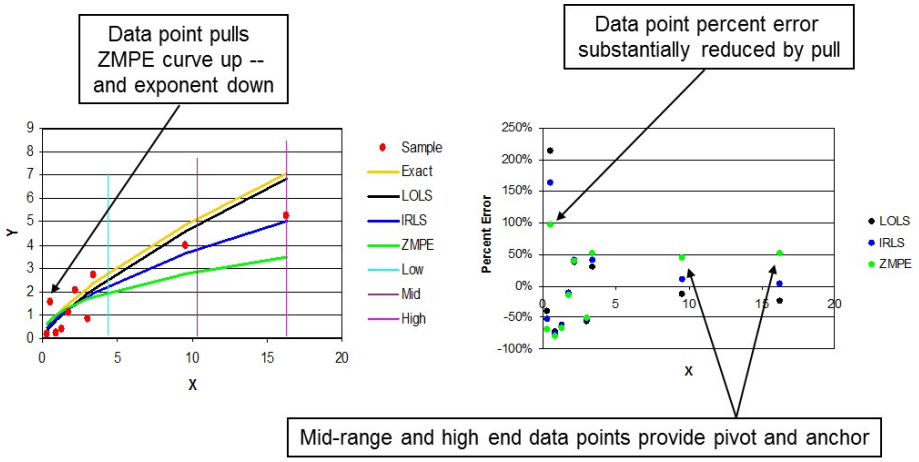

Figure 3. ZMPE Low-end Pull Example

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Even though the MUPE method also minimizes the sum of the percent errors, the reweighting process apparently diminishes the impact of low-end data points with large positive errors.

IDP Analysis Process

This section describes how CERAT is used to perform influential data point analyses. The process is the same as that used in the IDP study described above – with one significant difference. Instead of determining the impact of removing a data point on the estimate at the highest value of X (as was done in the study), CERAT determines the impact on the estimate for a "target" data point within the CER data set.

The CER developer selects the target data point. It is usually the data point with the largest estimated cost. If the CER is stratified into subgroups using binary stratifier variables, IDP analyses can be performed for target data points within each stratum.

Once the target data point is selected, individual data points are removed one at a time, and the resulting impacts on the baseline estimates for the target data point are determined. This process is repeated four times – once for each of the best fit methods. However, MUPE IDP analyses take much longer than ZMPE, LOLS and AAPE analyses. So, in practice, MUPE IDPs are usually suppressed by the CERAT user.

<u>Types of CERs Handled by CERAT</u>

CERs accommodated by CERAT have the general form $Y=AX^BY^CD^Z\ldots$ where X, Y and Z are variables. A, B, C and D are constants determined in the best fit optimization using the Solver add-in to Excel. CERAT is designed internally to handle up to 100 variables, although a practical limit is about 10. Higher numbers of variables would result in output display problems.

A term such as D^Z may be used for stratification within the CER data set. In this case, Z is a binary stratification variable with values of either 0 or 1. D is a factor determined by the best fit optimization. It represents the relative magnitude of Y values for the stratum of data points with Z = 1, as compared to the rest of the data points in the CER data set with Z = 0.

In some CAAG CERs, one or more of the exponents (B, C, and Z above) may be fixed, as opposed constants optimized by the best-fit process. Accordingly, CERAT allows the analyst to fix CER constants so that the origin (baseline) CER can be replicated exactly.

The following apply only to ZMPE and AAPE methods, because they can handle more general types of regression problems.

1. Estimating bias (average percent error) can be constrained to zero for any stratum (data subgroup)

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

2. The CER equation may have more than one additive term like that shown above

3. Exponents may have a compound form:

$$B = S_B * B'$$

where

S_B is a binary stratifier variable, and

B' is the exponent for the data points with $S_B = 1$, determined by the best-fit optimization

4. Fixed factors may be applied to data points:

$$Y_i = (AX_i^B Y_i^C D_i^Z)*F_i$$

where

i is the data point number (index), and

F_i is the fixed factor applied the ith data point

Compound exponents allow for different exponents for the same independent variable, depending on the data subgroup.

<u>Modified Cook's Distance</u>

One of the statistics CERAT produces as it removes data points one-by-one is a modified version of Cook's Distance. The standard definition of Cook's Distance, appropriate for Ordinary Least Squares (OLS) regressions, is:

$$Di = \sum_{j=1}^{n} \left(\hat{Y}_j - \hat{Y}_{j(i)} \right)^2 \Big/ p \cdot MSE$$

where

$\hat{Y}_j$ is the prediction from the full regression model for observation *j;*

$\hat{Y}_{j(i)}$ is the prediction for observation j from a refitted regression model in which observation i has been omitted;

MSE is the mean square error of the regression model; and

p is the number of fitted parameters in the model

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Since MUPE, ZMPE and AAPE are not OLS methods, a Modified Cook's Distance is used for all methods:

$$MCDi = \sum_{j=1}^{n} \left[\left(\hat{Y}_j - \hat{Y}_{j(i)} \right) \Big/ \hat{Y}_j \right]^2 \Big/ p \cdot MSPE$$

where

MSPE is the mean square percentage error of the regression model

Example Primary CERAT Statistics

Figure 4 shows example IDP analysis results for a notional CER with 25 data points. The baseline CER estimates for the target data point (the 25th in this example) are shown in the first data column ("B/L Y Est") of Figure 3. These are just for reference purposes.

	Influential Data Points (IDPs) and Impacts (DYs) on Estimate for Selected Data Point No. 25, Data Point 25							Results Over All Data Point Removals	
	B/L Y Est	1st IDP	1st DY	2nd IDP	2nd DY	3rd IDP	3rd DY	Min SPE	Max MCD
ZMPE	4.881	6	9.3%	4	8.9%	12	8.3%	37.3%	0.544
MUPE	4.880	12	12.3%	14	7.1%	14	6.3%	38.0%	0.384
LOLS	4.970	12	13.1%	14	7.5%	5	6.2%	39.1%	0.540
AAPE	5.291	6	12.7%	25	8.3%	1	6.3%	38.6%	0.839

Figure 4. Example CERAT Primary Statistics, Part A

CERAT identifies the 1st, 2nd and 3rd most influential data points for each best fit method. The values in the DY columns are the impacts of the IDP removals on the baseline estimates for the target data point.

The next to last column contains the minimum Standard Percent Error (SPE) over all data point removals. These values can be compared with baseline SPEs (not shown) to assess the amount of additional dispersion attributable to the data point with the largest impact on SPE (not necessarily one of the top three IDPs).

The last column in Figure 4 contains the maximum Modified Cook's Distance over all data point removals.

Figure 5 shows the rest of the primary IDP analysis statistics for the example.

The DY values in Figure 4 are "normalized" in Figure 5 as follows:

NDY = DY/SPE

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

	Normalized IDP Impacts (NDYs) and NDY Percentiles (NDPs)							
	1st NDY	1st NDP	2nd NDY	2nd NDP	3rd NDY	3rd NDP	Max SPE	Max GRSQ
ZMPE	0.23	17.9	0.22	15.2	0.20	12.0	42.2%	0.903
MUPE	0.29	54.3	0.17	14.0	0.15	10.5	44.1%	0.866
LOLS	0.30	60.8	0.17	10.2	0.14	1.9	46.4%	0.902
AAPE	0.31	40.3	0.20	18.7	0.15	7.8	45.1%	0.903

Figure 5. Example CERAT Primary Statistics, Part B

The normalization removes much of the influence of CER data point dispersion on the IDP impacts. This is shown by typical results from the 2011 Influential Data Point study in Figure 6.

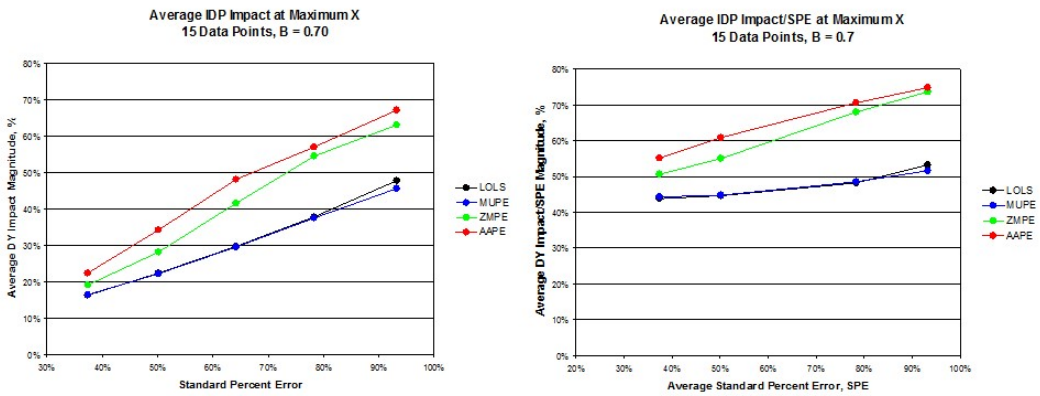

Figure 6. Raw and Normalized IDP Impacts vs. Standard Percent Error

The LOLS and MUPE NDY averages are almost flat over a wide range of SPEs, while the ZMPE and AAPE averages increase about 50% over the same SPE range.

Normalized DY Percentiles

The NDP values in Figure 5 are approximate NDY percentiles, derived from data generated in the 2011 IDP study. Monte Carlo sampled data points were used to generate reference percentiles as functions of the number of data points and the SPE – for target data points with the largest X values in the sampled data sets.

Figure 7 on the next page shows some of the reference percentiles for ZMPE CERs. The left-hand graph shows NDY percentiles vs. number of CER data points, N, for sampled data sets with LSE = 0.55, resulting in an average SPE of 64%. Additional reference percentiles vs. N were generated for LSE values of 0.35, 0.65 and 0.75 – resulting in average SPE values of 37%, 78% and 92%, respectively.

The right-hand graph in Figure 7 shows NDY percentiles vs. SPE for CERs with 25 data points. Additional reference percentiles vs. SPE were generated for N = 10, 15 and 50 data points.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

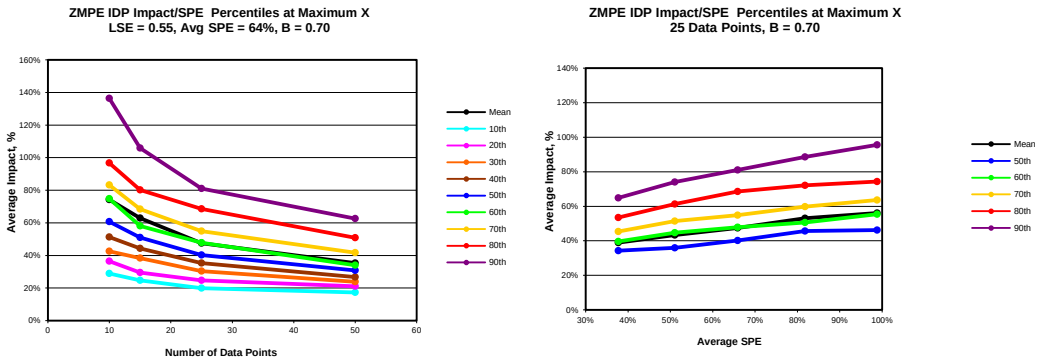

Figure 7. ZMPE NDY Percentiles vs. Number of Data Points and Standard Percent Error

CERAT establishes a given NDY percentile using two-way interpolation (or extrapolation). First it interpolates in the number-of-data point dimension, and then in the SPE dimension.

Figure 8 shows the 1st IDP NDY percentiles from the example IDP analysis. The lines represent interpolated reference percentiles to match the number of data points (25) and SPE values (41%-44%, depending on the best-fit method).

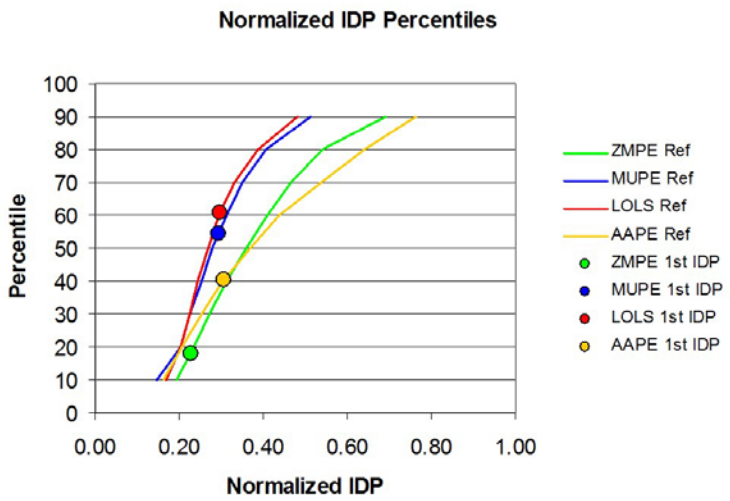

Figure 8. Example Normalized IDP Percentiles

In this case the ZMPE and AAPE percentiles are lower than the LOLS and MUPE percentiles. Thus, the ZMPE CER, in particular, might be favored over the other CERs, on the grounds that its 1st NDY (0.23) and associated percentile (17.9%) are significantly lower than the NDYs and percentiles for the other methods (see Figure 5).

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

However, other criteria may dominate a final selection of which best-fit method is best. For example, consider the regression coefficients in the baseline CERs shown in Figure 9. The exponents B and C vary a lot from one method to another (reminder: the data sets values are purely hypothetical).

Baseline Best-Fit Regressions -- Example Case

Constant	**A**	**B**	**C**	**D**	**E**
Original ZMPE Best Fit	1.036	0.250	0.365	0.926	1.016
CER Worksheet Column	N/A	D	E	F	G
CER Worksheet Header	LC	Weight	Freq	Type A	Type B
Used in Original Best-Fit?	Yes	Yes	Yes	Yes	Yes
ZMPE Baseline Fit	1.036	0.250	0.365	0.926	1.016
MUPE Baseline Fit	0.777	0.103	0.621	0.978	1.036
LOLS Baseline Fit	-0.203	0.044	0.733	0.010	0.020
LOLS Linear Values	0.674	0.044	0.733	1.023	1.047
AAPE Baseline Fit	1.092	0.317	0.339	0.842	0.916

Figure 9. Example Baseline Regression Constants

So what if frequency is considered a stronger cost driver than weight? This would favor the MUPE and LOLS best-fit methods over ZMPE, as their frequency exponents are much higher than the ZMPE exponent. Although the Type A and Type B stratification factors are not large, engineering analysis might favor one of the other methods over ZMPE. The LOLS weight exponent is fairly low. However, this can be misleading, as the discussion of variable "Swing Factors" below will reveal.

The drawback with the interpolated percentiles is that they are derived from simple $Y=AX^B$ CERs with only one independent variable, X, and a TDP at the maximum value of X. So applying these percentiles to CERs with multiple independent variables – some of which might be binary stratifiers – is a bit of a "stretch." However, developing percentiles for more complex, multivariate CERs would be challenging to design and very time-consuming to execute.

Percentiles for stratified CERs are difficult to deal with because the "effective" number of CER data points might be much less than the total if the 1st IDP is in a fairly small stratum (data subgroup). For this reason, and because the percentiles are derived from single-variate data sets, the CAAG takes the percentiles with "a grain of salt." They are intended to indicate how likely a given NDP would be for a CER with the same number of data points and level of dispersion (SPE), But except for simple $Y=AX^B$ CERs, the validity of the percentiles is not established.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Potential improvements have been developed by adjusting the effective number of data points used in the interpolations to reflect possible impacts of small to medium sized strata. In particular, minimum and maximum effective number of data points are developed for calculating percentiles. The minimum number is the smaller size of (a) the stratum containing the influential data point and (b) the stratum containing the target data point. The maximum number is the average of the total number of data points and the size of the largest stratum. These improvements seem to yield reasonable results, but they have not been thoroughly evaluated.

Considering the issues with the percentiles, the best way to judge the relative seriousness of influential data points is by comparison of NDY values with those from IDP analyses of other CERs with about the same number of data points.

Supplementary IDP Statistics

For each data point removal, the maximum impact over all other data point estimates (including the target data point) is determined. Figure 10 shows examples of these results for the ZMPE method.

ZMPE IDP Analysis

CER Data Set		Impacts on Selected Data Point Estimate			Max % Impact Over All Data Pts	
Data Pt.	Description	New Est Y	New Est - B/L Est	% Diff	Max Y % Diff	Data Point
	Baseline Values	4.881				
1	Data Point 1	4.951	0.070	1.4%	22.3%	1
2	Data Point 2	4.916	0.035	0.7%	-23.3%	12
3	Data Point 3	4.860	-0.021	-0.4%	-1.2%	1
4	Data Point 4	5.313	0.432	8.9%	-30.7%	1
5	Data Point 5	4.831	-0.050	-1.0%	8.2%	5
6	Data Point 6	4.427	-0.454	-9.3%	-9.3%	25
7	Data Point 7	4.831	-0.050	-1.0%	7.1%	10
8	Data Point 8	4.837	-0.044	-0.9%	-1.7%	1
9	Data Point 9	4.815	-0.066	-1.4%	2.6%	9
10	Data Point 10	4.784	-0.097	-2.0%	5.0%	10
11	Data Point 11	4.917	0.037	0.7%	-3.8%	11
12	Data Point 12	5.287	0.406	8.3%	16.8%	12
13	Data Point 13	4.862	-0.019	-0.4%	-0.8%	11
14	Data Point 14	5.170	0.289	5.9%	9.7%	1

Es t Y for D.P. 12 moves the most (-23.3%) when D.P. 2 is removed

2nd IDP

1st IDP

3rd IDP

Figure 10. Example ZMPE IDP Statistics for Individual Data Points

The calculation of the target point DY is shown in the first three data columns. The first three most influential data points are identified in the figure.

The last two columns in Figure 10 show the largest impact – on any data point – of each data point removal. For example, removal of Data Point 2 causes a downward movement of 23.3% in the CER estimate for Data Point 12. Thus, even though Data Point 12 is not the target data point, the large DY might be a cause for concern.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Figure 11 shows a CERAT output display that records the impact of each data point removal on the ZMPE CER regression constants. These displays help assess the overall "stability" of the CER by comparing the movement of the regression constants relative to the baseline constants shown in the first data row of the figure.

		ZMPE IDP Statistics						
Data Pt.	Description	A	B	C	D	E	SPE	MCD
	Baseline Values	1.036	0.250	0.365	0.926	1.016	41.0%	
1	Data Point 1	1.163	0.276	0.273	0.980	1.000	40.1%	0.171
2	Data Point 2	0.753	0.191	0.508	1.057	1.099	37.3%	0.382
3	Data Point 3	1.023	0.251	0.367	0.929	1.017	42.2%	0.002
4	Data Point 4	0.938	0.271	0.477	0.789	0.843	39.8%	0.544
5	Data Point 5	1.079	0.247	0.360	0.900	1.061	39.0%	0.058
6	Data Point 6	1.048	0.246	0.352	0.868	1.098	39.9%	0.144
7	Data Point 7	1.007	0.193	0.416	0.971	1.008	40.9%	0.043
8	Data Point 8	1.028	0.251	0.366	0.920	1.021	42.2%	0.003
9	Data Point 9	1.040	0.249	0.365	0.912	1.038	42.1%	0.008
10	Data Point 10	0.986	0.193	0.427	0.957	1.014	41.9%	0.022
11	Data Point 11	1.052	0.283	0.321	0.937	1.017	42.1%	0.012
12	Data Point 12	1.226	0.362	0.222	0.887	0.983	38.0%	0.125
13	Data Point 13	1.029	0.251	0.364	0.929	1.027	42.1%	0.001
14	Data Point 14	1.133	0.301	0.287	0.944	0.988	38.8%	0.067
15	Data Point 15	1.203	0.285	0.261	0.909	0.914	40.5%	0.239
16	Data Point 16	0.992	0.202	0.424	0.957	1.004	41.0%	0.028
17	Data Point 17	1.029	0.250	0.364	0.927	1.017	42.2%	0.001
18	Data Point 18	1.006	0.217	0.423	0.885	1.070	40.5%	0.087

Lowest SPE over all D.P. removals

Highest Modified Cook's Distance

Highest SPE

Figure 11. ZMPE CER Equation Constants by Data Point Removal

More IDP Impact Assessment Tools

CERAT produces specialized XY graphics to help visualize the impacts of IDPs on the regression equation. In particular, both a linear and a log-log graph is created for each method and each continuous independent variable.

Figure 12 on the next page shows LOLS graphs from the example CER, which has two continuous IVs – weight and frequency [Note: the example CER data set is totally artificial, not related to any real CER data set].

The Y values of the data points in Figure 12 are not the actual Y values, but rather adjusted values to enhance comparison of the data points with the regression lines in the figure. The adjustment process moves each data point along the regression surface to a specific value for each out-of-plane variable. The CER regression lines in the figures also use the same specific values for out-of plane variables.

CERAT uses the average for each out-of-plane continuous variable in generating the adjustments and the CER regression lines. The user can specify the values of each stratifier variable (either 0 or 1).

As an example of adjusting actual Y values for plotting against weight, consider the

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

baseline ZMPE regression from our example CER:

$$Y = A \cdot Wt^{B} \cdot Freq^{C} \cdot D^{TypeA} \cdot E^{TypeB}$$

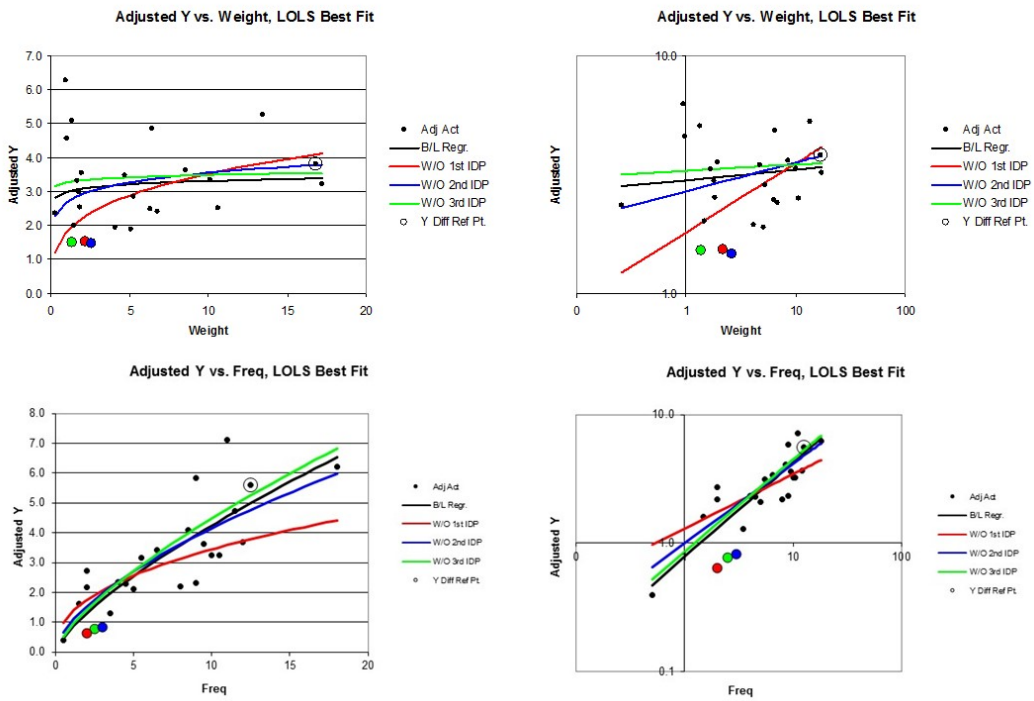

Figure 12. Example LOLS XY Graphs of Adjusted Y vs. Weight and Frequency

Weight (Wt) and frequency (Freq) are continuous variables. The variables TypeA and TypeB are binary stratifiers.

The ZMPE baseline CER regression constants, shown in Figure 9, are A = 1.036, B = 0.250, C = 0.365, D = 0.926 and E = 1.016

The adjusted value of the first data point is calculated below. This data point has actual Y = 0.329, frequency = 0.50 TypeA = 1 and TypeB = 0. The out-of-plane graphing values are the average frequency, 6.880 and the user-specified stratifier values TypeA = 1 and TypeB = 1.

The adjusted Y value for the data point is:

$$\text{Adj Y} = \text{Act Y} \cdot (6.880/0.50)^{0.365} \cdot (0.926^{1}/0/926^{1}) \cdot (1.016^{1}/1.016^{0})$$

$$= 0.329 \cdot 2.602 \cdot 1 \cdot 1.016 = 0.870$$

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Note that the CER leading constant A and the weight exponent B do not enter into the adjustment.

After the Y adjustments are made and used in a graph, the data points and the baseline CER are directly comparable because both are based on the same out-of-plane variable values.

CER regression lines where the first, second or third most influential IDP is removed are also shown in Figure 12 (red, blue and green, respectively). These lines can be compared with the black baseline regression line. However, these lines cannot be compared directly with the adjusted data points. To do so would require three separate graphs, one for each IDP. For a given IDP graph, each Y value would have to be adjusted using the CER equation constants with that IDP removed from the data set.

The movement of regression lines in graphs such as those in Figure 12 can seem counter-intuitive. For example the CER regression line without the first IDP actually moves closer to the adjusted Y value of the first IDP (red dot). However, in the frequency graph, the CER line moves away from the 1st IDP.

Figure 13 shows the maximum DY impact (over all data points) that results from removal of a data point plotted against the Modified Cook's Distance for the data point.

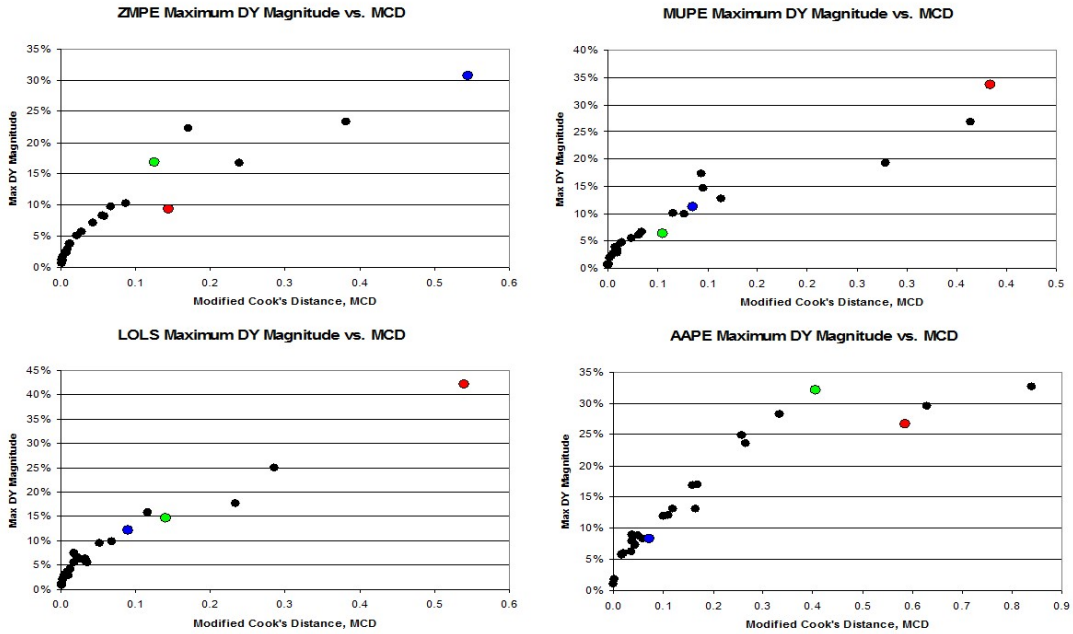

Figure 13. Data Point Maximum DY Impact vs. Modified Cook's Distance

The maximum DY and MCD are clearly correlated, although the maximum DY is not monotonic in MCD.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

The most influential data points are not necessarily those with the largest DY impacts (over all data points) or those with the largest MCD values. Data points with high maximum DY and MCD values should be reviewed for validity and reasonableness, even if they are not in the top three most influential with regard to impacts on the target data point.

Other CER Assessment Tools

The tools described in this section are not directly related to IDP analysis, but rather helpful to CER development in a general way. They have been included in CERAT to automatically create a robust set of information at the CER developer's fingertips.

Residuals vs. Continuous IV Graphs

Figure 14 shows graphs of CER residuals (percent errors) plotted against the two independent variables in our example CER, weight and frequency. The trend lines in the figure can reveal behavior which may be considered undesirable.

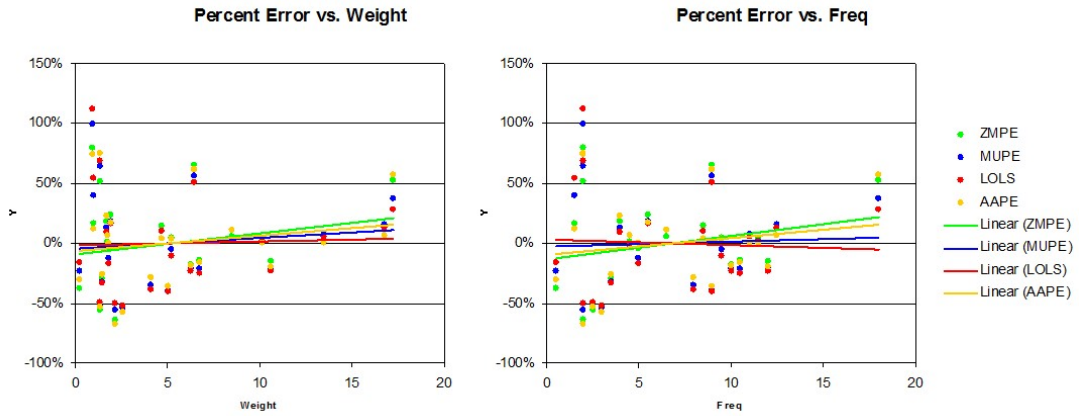

Figure 14. Example Residual Graphs

These types of graphs are always created at the CAAG to evaluate candidate CER heteroskasticity (absence of trends in the independent variables). If a trend line is too steep, the CER developer may take remedial action such as adding new variables or removing data points to reduce the magnitude of the trend line slope. In this case, the ZMPE and AAPE methods have steeper slopes than the LOLS and MUPE methods. However, considering the level of data scatter, the slopes are probably not steep enough to stimulate remedial action.

CAAG CER developers routinely plot residuals vs. cost mid year (CMY) to look for indications of time-trends. Fortunately, there are relatively few cases where the trend line is steep enough to cause concern.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

CERAT also produces Log-Log versions of the residual graphs. For LOLS, the errors in the log domain can be plotted against the corresponding logarithms of the variable values. However, for ZMPE, MUPE and AAPE, another approach must be taken. The percent errors in these methods have both positive and negative values. Since we can't take a logarithm of a negative number, the value 1.0 is added to each percent error, expressed as a fraction. Since the smallest error is larger than -1.0, logarithms can be taken for every data point and then plotted against the logarithms of the independent variable values.

The advantage of log-log plots is that they "spread out" the data points both horizontally and vertically. This helps visualize cause and effect. In many CER cases, there is a lot of data with both small cost and small IV values. This causes the data to be "bunched up" near the origin of linear XY graphs.

For LOLS CERs, the log-log plots help identify low-end data points that are often influential if they have large or small errors. For the other methods, the impact of individual data points is better assessed by looking at linear graphs. However, log-log graphs can help identify which data points are likely to be causing problems. For ZMPE and AAPE, data points exerting low-end pull may not be too obvious in linear graphs.

Residual Histograms

Residual histograms like those in Figure 15 can identify unusual CER data sets. The left-hand graph contains histograms of percent errors for each best-fit method. The right-hand graph has histograms of the logarithms of actual Y divided by estimated Y. Almost always, the percent error histograms reveal skewed behavior, typical of lognormal distributions.

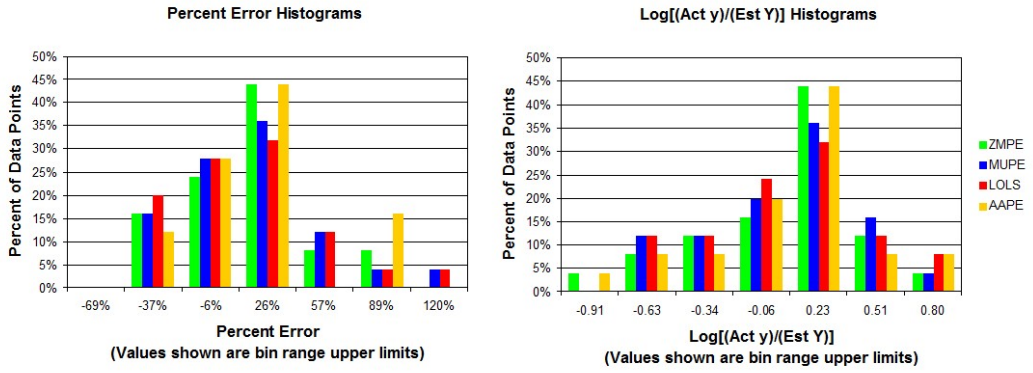

Figure 15. Example Error Histograms

The CAAG has a large volume of percent error data for both hardware development (NR) CERS and production (R, or Rec) CERs. Scaling the percent errors from a given CER by the inverse of SPE for the CER allows the errors from all NR CERs to be combined. The same can be done for all the R CERs. The distributions of the scaled, combined errors are very close to exact lognormal distributions – except for the right-side tails, which are not

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

as pronounced. This was first demonstrated in 2002[2] and has been reinforced with additional data since then.

The log histogram in Figure 15 illustrates what seems to be underlying behavior for space hardware. The distribution appears normal, except for truncated right-hand tails. This is probably caused by the inherent restrictions in space hardware development and production: when costs look like they are going to "go through the roof", actions are taken to control them. One action is relaxing requirements in development to keep the cost down. Performance requirements might also be waived during production to get on with system integration.

Correlation matrices and Variable Swing Factors

CERAT automatically creates correlation matrices for the CER data set as part of the baseline regression analysis process. The top portion of Figure 16 contains the correlation matrix for the example CER. This CER uses factors applied to each data point, hence the "Factor" line at the bottom of the matrix and in the last column.

Variable	Y	Weight	Freq	Type A	Type B	Factor
Y	1.000					
Weight	0.870	1.000				
Freq	0.844	0.896	1.000			
Type A	0.234	0.306	0.223	1.000		
Type B	0.083	0.159	0.177	-0.315	1.000	
Factor	0.234	0.082	-0.060	0.073	-0.254	1.000
ZMPE SWFs	8.552	2.867	3.695	1.080	1.016	2.000
MUPE SWFs	9.536	1.545	9.254	1.022	1.036	2.000
LOLS SWFs	10.206	1.203	13.841	1.023	1.047	2.000
AAPE SWFs	8.294	3.801	3.366	1.187	1.092	2.000

Note: Y values for SWFs are estimates with each variable at its maximum value.

Figure 16. Example Correlation Matrix and Swing Factors

It should be noted that CAAG CER developers usually create correlation matrices that contain a wide range of potential CER independent variables. On the other hand, CERAT only includes those variables that are used in the CER it is analyzing.

The bottom portion of Figure 16 contains Swing Factors for each best-fit method. A Swing Factor (SWF) defines how much IV values in the CER data set can impact (swing) a CER estimate. In particular, the SWF is the ratio of the largest estimate to the smallest

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

estimate resulting from movement of the IV from its minimum to its maximum in the data set.

The SWF for a continuous variable X is either:

(1) $SWF_X = (X_{max}/X_{min})$^B or (2) $SWF_X = (X_{min}/X_{max})$^B = 1/$(X_{max}/X_{min})$^B

where B is the exponent for variable X

The SWF takes the first form if B is positive, and the second form if B is negative.

For a stratifier variable Z, the SWF is:

(1) D_Z if $D_Z > 1.0$ or (2) 1/ D_Z if $D_Z < 1.0$

where D_Z is the stratification factor for variable Z

These SWF definitions are for a one-term CER. SWF values for CERs with multiple terms are more complex and not described here.

Swing Factors for factors applied to CER estimates are taken as the ratio of the largest factor to the smallest factor within the CER data set.

As an example of how Swing Factors can contribute to CER development, consider the MUPE and LOLS weight exponents in Figure 9 – 0.103 and 0.044, respectively. These results might seem too low to justify including weight as a variable in a CER. However, the weight Swing Factors in Figure 16 are 1.545 for the MUPE CER and 1.203 for the LOLS CER. These represent fairly significant swings in estimated cost over the range of weights in the CER data set. It would be difficult to justify excluding weight from either CER.

The lesson illustrated here is that the CER developer should not discard potential CER variables if they have low exponents *without looking at their Swing Factors*. A variable with a wide range of values within a CER data set can have a large impact on estimates even if its exponent is fairly low.

Stratifier variables are easier to assess for significance since their Swing Factors are either the regression-derived factors for the strata or their inverses. In Figure 9, four of the stratification factors are less than 1.0, so their inverses appear in Figure 16. The MUPE and LOLS SWFs for the TypeA stratum are 1.022 and 1.023, respectively. These are so small that it would not make sense to bother including the TypeA stratifier in either CER. However, the TypeA factors for ZMPE and AAPE are 1.080 and 1.187, respectively. Anything over 1.05 might be considered significant enough to embrace, so the ZMPE factor of 1.080 is probably large enough to justify including the TypeA stratifier in the CER. The AAPE SWF is well above a reasonable threshold and would clearly justify employing the TypeA stratifier.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

All but the AAPE TypeB SWFs in Figure 16 are less than 1.05, so this stratifier would be difficult to justify in the ZMPE, MUPE and LOLS CERs. The AAPE SWF for TypeB is 1.09, so it would probably make sense to include this stratifier in the CER.

The CAAG has a stratifier SWF of 1.07 in one of its CERs. Thirty-five of 57 CAAG box NR and R cost CERs are stratified, with a total of 57 stratifier variables. Three stratifier SWFs are less than 1.10 and seven are less than 1.20.

Swing factors are a common-sense alternative to the traditional T statistics used in OLS to assess statistical significance of CER independent variables. The T statistic is not appropriate for ZMPE, MUPE and AAPE. Moreover, it would not necessarily be the "last word" on whether a variable should be included in a CER or not even if it were appropriate. It is the author's opinion that a variable with a Swing Factor upwards of 1.05 can be justified -- in spite of any statistical significance test – as long as the variable and its SWF make engineering sense.

CER Percent Error Skew

Another way to compare CER data with lognormal distributions is by calculating the skew in the CER percent error data:

$$skew = \frac{n}{(n-1)\cdot(n-2)} \cdot \sum \left(\frac{(x-\bar{x})}{s} \right)^3$$

Where

x is the percent error,

$\bar{x}$ is the mean error (zero in this case),

s is the standard percent error, and

n is the number of data points

Figure 17 on the next page shows skew values for each of the best-fit methods in the example CER. Almost all CAAG CERs have positive skew values and a few have skew values above the exact lognormal line in the Figure.

All of the skew values are positive, as are the overwhelming majority of CAAG CER values. Since CERs seldom have extended right-hand tails, data set skew is usually less than that for exact lognormal distributions (solid line in Figure 17). Figure 15 reveals why the ZMPE and AAPE skew values in Figure 17 are relatively low. Neither has a data point in the last bin of the percent error histogram.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

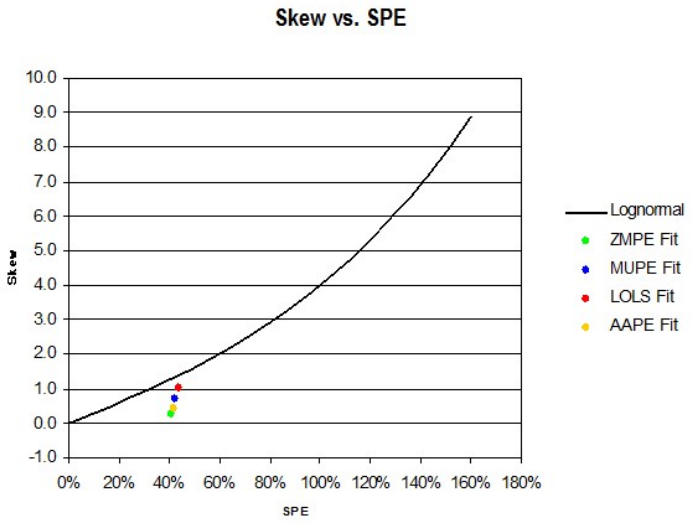

Figure 17. Skew vs. SPE

Generalized R^2 vs. SPE/B Graphs

Figure 18 shows a graph of the generalized R^2 values for each of the example CER best-fit methods, compared with a curve derived from sets of Monte Carlo XY data that were fit with LOLS regressions. The "generalized" R^2 is the Pearson correlation between the estimates and the actual values of Y in a CER.

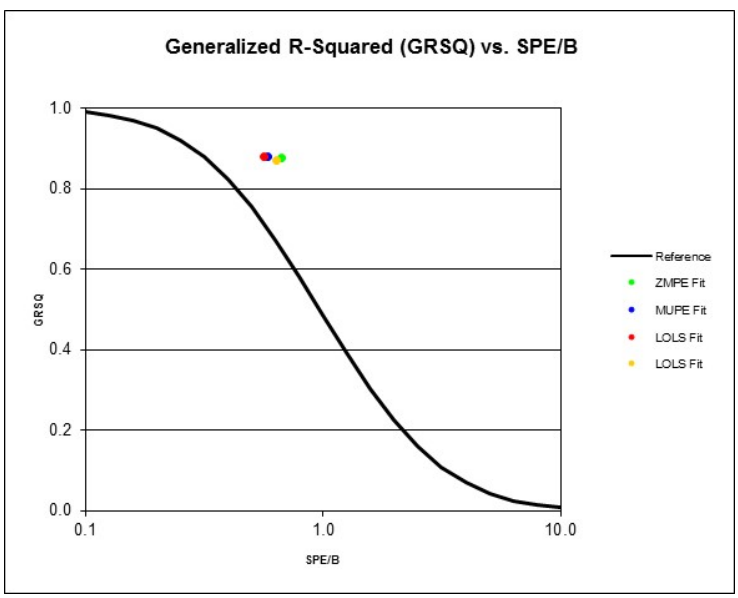

Figure 18. Generalized R2 vs. SPE/B

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

In 2005[3] the author first observed that LOLS CERs of the form $Y=AX^B$, where both X and Y are lognormally distributed, tend to follow the line in Figure 18. This line is the average over a large number of simulated CERs with different values of SPE and exponent B.

The basis for the curve in Figure 18 is shown in Figure 19, from Reference [3]. In the left-hand graph, R^2 values decrease with increasing SPE. In the right-hand graph they increase with increasing B. This suggested that dividing B into SPE and plotting R^2 against the result might be interesting. Surprisingly, all of the curves in Figure 19 collapsed into the single line shown in Figure 18!

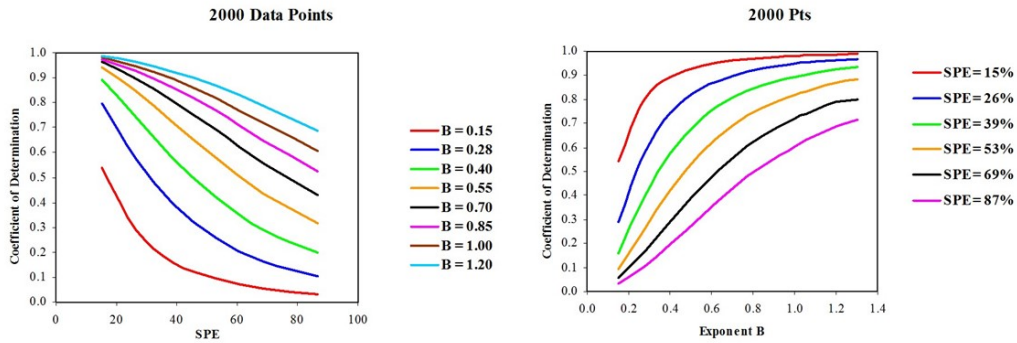

Figure 19. Generalized R^2 vs. SPE and Exponent B

For purposes of assessing the example CER, the weight and frequency exponents for a given CER were added together and the result divided into the SPE for that CER. Presumably this approach is reasonable even though the curve is based on a single IV with exponent B.

Any R^2 on or above the reference line in Figure 18 may be viewed as "good" (in the author's opinion). Results below the line might indicate the data set is not "well behaved."

Summary

This section summarizes CAAG experience analyzing CERs for influential data points with CERAT. CERAT has been used in 37 CER developments from 2012 to the present. These developments included CERs for estimating nonrecurring and recurring costs of space hardware boxes; software development and maintenance costs; and SEIT/PM costs for payloads and ground systems.

There are about 45 additional CAAG CERs that were developed before CERAT was available for use. Most of these will eventually be updated and CERAT will be used to support those efforts.

Overall Statistics

Table 1 on the next page shows statistics from the 37 CER developments where CERAT was used. Box recurring costs are most plentiful in the CAAG database and the average

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

DOF of box recurring CERs, 44, reflects that. On the other end of the spectrum, software costs are the least plentiful and hence the average software CER DOF is only 14.

Table 1. CAAG IDP Analysis Statistics

Description	Best-Fit Method	Box NR CERs	Box R CERS	S/W CERs	SEIT/PM CERs	All CERs
Number of CERs	**All**	**8**	**21**	**3**	**5**	**37**
Average Degrees of Freedom, DOF	**All**	**73**	**44**	**14**	**36**	**47**
Standard Percent Error, SPE	ZMPE	61%	50%	28%	52%	51%
	LOLS	65%	53%	28%	56%	55%
	AAPE	64%	52%	29%	52%	52%
Average 1st IDP NDY Magnitude	ZMPE	0.31	0.24	0.42	0.28	0.28
	LOLS	0.29	0.21	0.46	0.24	0.25
	AAPE	0.32	0.25	0.52	0.36	0.30
Average 2nd IDP NDY Magnitude	ZMPE	0.08	0.14	0.26	0.17	0.14
	LOLS	0.13	0.15	0.35	0.15	0.16
	AAPE	0.15	0.19	0.39	0.34	0.22
Average 1st IDP NDY Percentile	ZMPE	50	27	54	39	36
	LOLS	71	38	69	60	51
	AAPE	60	32	38	35	35
Average 1st IDP NDY Magnitude	Selected CER (ZMPE or LOLS)	0.28	0.22	0.42	0.25	0.25
Average 2nd IDP NDY Magnitude	Selected CER	0.11	0.15	0.26	0.18	0.15
Average 1st IDP NDY Percentile	Selected CER	56	33	54	37	41

The CAAG typically fits each candidate CER with both ZMPE and LOLS methods, selecting one candidate as the "best" CER. The selected CER is then used in all future estimates requiring that particular type of CER, such as a CER for estimating the recurring costs of structures. The ZMPE version was selected as best for 22 of the CERs analyzed with CERAT. The LOLS version was selected for the remaining 15 CERs.

The average Standard Percent Error for the selected CERs (not shown in the table) is 51%. The SPE values in the table show how ZMPE produces SPE values that are somewhat lower than those produced by LOLS and AAPE. It should be noted that the CAAG generally selects the CER that fits best with engineering expectations, as reflected by the CER exponents, stratification factors and, in a few cases, leading constants. The SPE with the lowest SPE, invariably a ZMPE CER, is often not chosen because the LOLS version fits better with engineering expectations.

The three software CERs have much lower SPE values than the other CER types. This is very likely a consequence of their small CER data sets. Figure 20 shows how CERs with small data sets tend to have low SPEs. This behavior was revealed in Reference [2]. It

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

can be thought of "over-achieving", because all best-fit methods tend to do a good job of fitting small data sets, yielding low SPEs but not producing realistic CER constants (i.e., the constants do not properly reflect underlying behavior). Also, the smaller the data set, the more likely extreme Y values will not be present, thus resulting in a "lucky" SPE. It would seem this type of behavior has occurred with the software CERs – as opposed to the (unlikely) case that software efforts have inherently less randomness than hardware and SEIT/PM efforts.

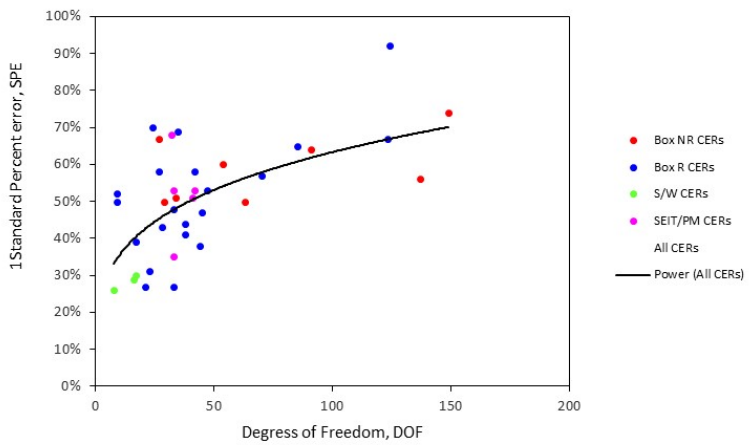

Figure 20 CER Standard Percent Error vs. Degrees of Freedom

Normalized IDP Impacts

The magnitudes of the 1st IDP normalized impacts, NDYs, for the selected CERs average 0.25, and the 2nd IDP NDY magnitudes average 0.15 (see Table 1). The average software NDY magnitudes are much higher than those for the other types of CERs. This too is a consequence of their small data sets. The 2011 IDP study showed that NDY magnitudes are much higher at low DOF values than at high DOFs.

The average 1st IDP NDY percentile for the selected CERs is 41. This is somewhat lower than what would be expected (50, or the median outcome). The LOLS percentiles are significantly higher than those for LOLS and AAPE. This seems paradoxical, because the LOLS average 1st NDY magnitudes are lower than the ZMPE and LOLS magnitudes for all CER types except software. This could be expected since LOLS NDY magnitudes averaged much less than ZMPE and AAPE magnitudes in the 2011 IDP study, thereby producing "tight" reference percentile data.

Figure 21 shows the distribution of 1st IDP NDY magnitudes and percentiles for the selected CERs. The NDY magnitudes range from a low of 0.09 to a high of 0.48. The NDY percentiles range from about 1% to 85%.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

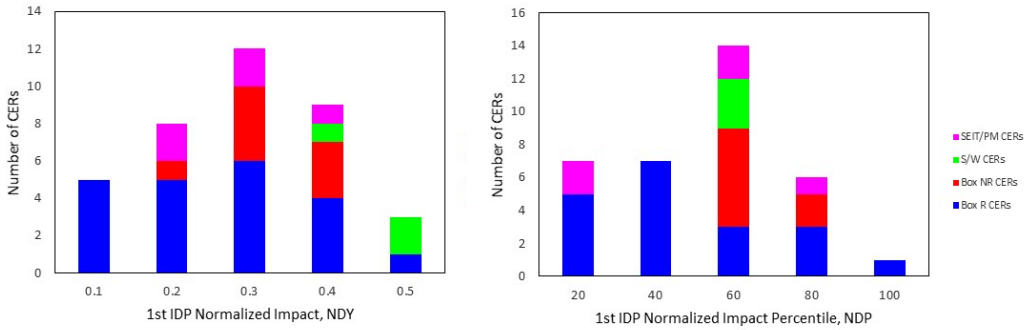

Figure 21. First IDP NDY Magnitude and Percentile Distributions

First IDP NDY magnitudes are plotted against DOF and SPE in Figure 22 on the next page. The upper graphs show actual results from the 39 CAAG CERs. The lower graphs show typical simulated CER results from the 2011 IDP study. The actual results are in good agreement with "theory", as it evolved from the study. The NDY magnitudes decease significantly with CER DOF (right side graphs). The actual magnitudes are virtually constant with SPE, while the ZMPE study results increase moderately with SPE. However, LOLS NDY values and much flatter than ZMPE values. So it's not surprising that the actual NDY magnitudes show no increase with SPE.

One of the CERAT features was used to create the upper graphs. The regression lines are from a LOLS fit to the data where both DOF and SPE were independent variables – along with stratifiers for box NR, software and SEIT/PM CERs. The data points in each graph were adjusted (as described in the "More IDP Assessment Tools" section above) to move them into the plane of the graph. For the DOF graph, each data points was adjusted to the average SPE value of 51.2%. Each data point was adjusted to the average DOF of 46.5 for the SPE graph.

NDY Percentiles

Figure 23 shows 1st NDY percentiles plotted against DOF and SPE. We might expect percentiles to more or less independent of both DOF and SPE. That seems to be the case for DOF, as the trend line is fairly flat. But the SPE trend line is not (it has essentially the same without the data point with the highest SPE). Attempts to create stratified regressions did not improve matters much. A regression with SPE and stratifiers, as well as one with both SPE and DOF plus stratifiers were created. They both had steep slopes in SPE. Maybe there is an underlying reason for the NDP vs. SPE slope. However, the pattern of data points in Figure 23 does not seem to reveal a cause-and effect behavior.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

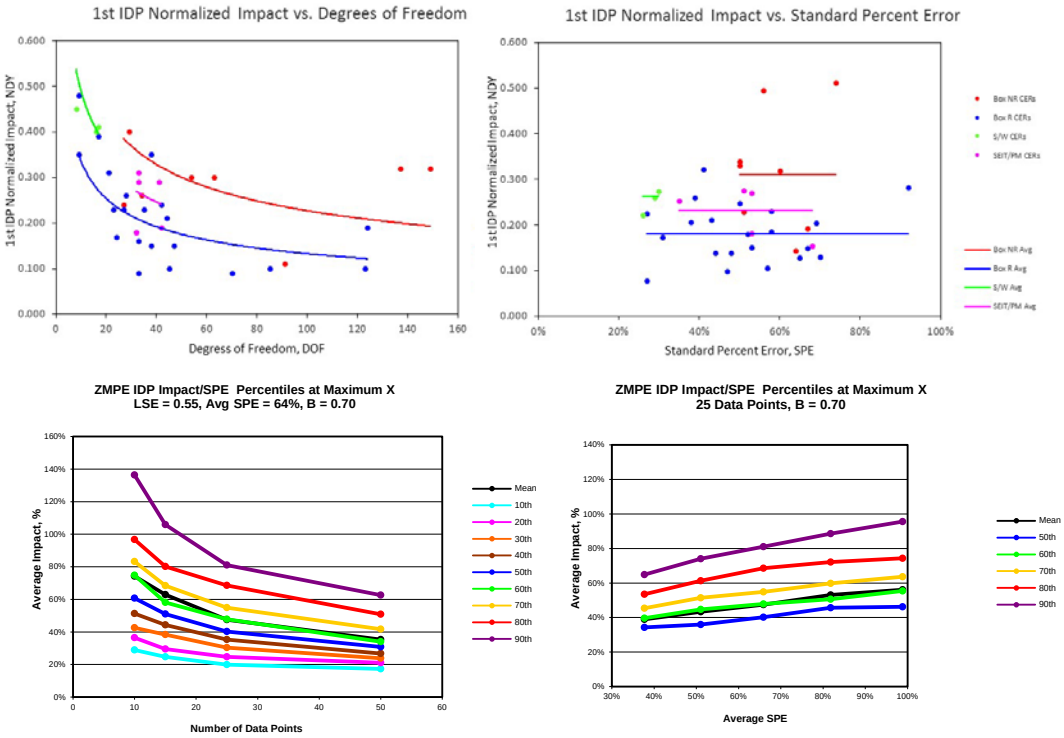

Figure 22 First IDP NDY Magnitude vs, DOF and SPE

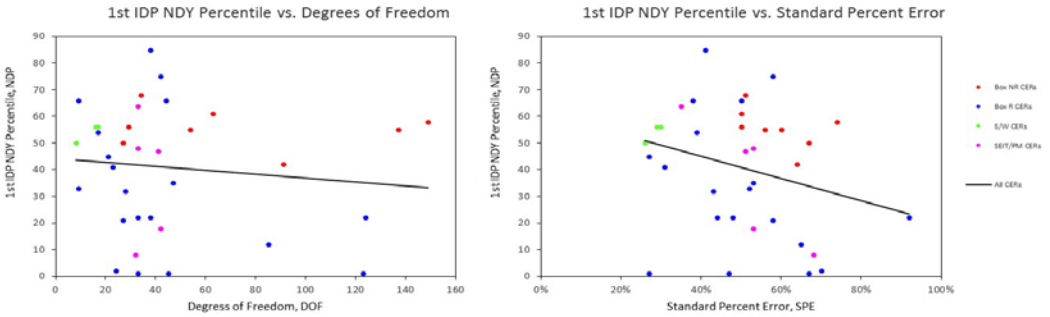

Figure 23. First IDP NDY Percentile vs. DOF and SPE

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Conclusions and Recommendations

CERAT has become a valuable tool for helping CAAG CER developments. In several cases, one or more highly influential data points were removed from CER data sets because they were identified by CERAT. Our experience with ZMPE low-end pull has taught us that the most influential data point may not be obvious from looking at a graph.

The CAAG practice of focusing on the IDP NDY magnitudes, as opposed to NDY percentiles, seems prudent. This is primarily because the percentiles are based on simulated CERs with only a single independent variable, while most CAAG CERs have multiple variables. That said, more attention should be paid to percentiles for CERs with a single IV.

NDY magnitudes are best assessed by comparison with those from other CER developments with about the same degrees of freedom and, secondarily, about the same Standard Percent Error.

CERs with small data sets require extra attention. Often their SPE values are artificially low and their NDY magnitudes are relatively high. Unfortunately, we are reluctant to remove a very influential data point from a small data set. However, the "proof of the pudding" is in the independent variable constants. If they are more in line with engineering expectations after removal of a date point, then common sense dictates removal. If they are not, then the CER is better off with the influential data point included in the data set – even if it has a relatively high NDY magnitude.

Impacts of data point removals on (a) CER constants and (b) DY values for data points other than the target data point should be reviewed for undesirable behavior – particularly "unstable" CER constants. Data points with high maximum DY impacts or high Modified Cook's Distances should be reviewed to be sure they are valid members of the CER data set and do not have atypical technical or cost characteristics.

Independent variable Swing Factors should be reviewed before eliminating an independent variable from a CER based on a low exponent magnitude. Of course, exponents should always have a sign that is consistent with engineering judgement. Likewise, stratification factors should also be consistent with engineering expectations.

Residual histograms, skew graphics and Generalized R^2 graphics should be examined to identify unusual data sets. For example, left-skewed residuals and lower than average R^2 values considering SPE and CER exponents.

CAAG CERs seem to behave in a way that is consistent with the results of the 2011 IDP Study. Future CAAG CER developments will shed more light on IDP behavior.

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

References

1. MacKenzie, Donald, ***Influential Data Points in Regression Analyses,*** 2012 ISPA/SCEA Joint International Conference and Workshop, Brussels, Belgium, May 14-16, 2012

2. MacKenzie, Donald, ***Cost Estimating Relationship Regression Variance Study,*** 2002 ISPA 24th Annual Conference, San Diego, CA, May 21-24, 2002

3. MacKenzie, Donald, ***A New Look At The Regression Statistic R-Squared,*** Eleventh U.S. Army Conference On Applied Statistics, Monterey, CA, October 19-21, 2005

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.