A Novel Non-Recurring Production CER Methodology

Lisa Hackbarth and Raymond Covert
MCR LLC Covarus LLC
lhackbarth@mcri.com rpcovert@covarus.com

ABSTRACT

Production costs are generally categorized as either non-recurring or recurring. Typically non-recurring costs include tooling and pre-production activities, among others. Cost estimating relationships (CERs) are generally developed first for recurring hardware costs, and then non-recurring CERs are developed as a function of recurring hardware costs.

Non-recurring hardware costs are notoriously difficult to estimate for two reasons. First, a particular contractor may not break-out non-recurring costs while others may break-out non-recurring costs using definitions inconsistent with other contractors. Second, non-recurring costs may only be relevant for certain productions lots such as very early lots, lots which represent a significant increase in production quantity, and lots which represent a change in design. Therefore, actual production cost data show a mix of production lots with zero and non-zero non-recurring productions costs.

A common way to develop a non-recurring CER is to calculate a factor as the ratio of non-recurring hardware cost to recurring hardware cost. Another way is to calculate the non-recurring hardware cost as a percent of the recurring hardware cost for each production lot and then regress these percentages as a function of the lot cumulative quantity. The non-recurring costs tend to be higher for early lots and are often zero for later lots, so a CER based on quantity will result in a steeply declining estimate as the quantity increases which better fits the data. If a large number of lots have zero non-recurring costs, both estimating methods will have large statistical errors, but the CER method will have smaller errors.

This paper provides a novel, alternate methodology to reduce statistical errors. This methodology combines (a) a method to predict whether the non-recurring costs will be zero or non-zero and (b) the CER regression method mentioned above, but with only the non-zero values of non-recurring hardware cost as a percent of the recurring hardware cost for each production lot.

The method to predict whether the non-recurring costs will be zero or non-zero is known as logistic regression. It is a technique used to find relationships between an independent variable that can take on multinomial, categorical values (i.e. binary or other multinomial value) and a series of dependent variables. Logistic regression has been used in studies of cost growth (Lucas

& White, 2009) (White, Sipple, & Greiner, 2004) and failure analysis modeling to model bimodal (i.e. zero and non-zero) behavior of data.[1,2]

In this paper we will (a) provide example data for weapon systems from which we will derive non-recurring CERs, (b) walk through the methods typically used to estimate non-recurring production costs and their weaknesses, (c) discuss the logistic regression technique, (d) show the logistic-regression-enhanced CER, and then (e) show the error analyses for the different methods of estimating non-recurring costs.

INTRODUCTION

MCR developed a cost model to assist in estimating the cost of a weapons system using cost estimating relationships (CERs). We regressed recurring (R) production CERs against various technical parameters (i.e. weight, range, etc.) in order to find the best way to estimate R production hardware costs. We used the zero percent bias, minimum percent error (ZMPE) unit-as-an-independent variable (UAIV) regression technique to develop these R CERs (Covert & Wright, 2012).[3]

For completeness, we require estimates of the total production costs which are the sum of non-recurring (NR) and R hardware production costs. Typically, NR production costs were estimated using a simple NR/R factor, but we wondered whether there was a better way to estimate NR production costs.

Weapon systems NR lot production costs are notoriously difficult to estimate for several reasons. First, a particular contractor may not break-out NR production costs at all while others may have differing definitions of what constitutes NR cost. Second, the NR data may be combined with the missile R production lot costs and may be inseparable (for our data set we excluded programs where the contractor did not break-out NR and R production costs). Third, the data often show a mix of production lots with zero and non-zero NR production costs. Fourth, typically there is a decrease in NR production costs in later lots. Fifth, and finally, the cost drivers of NR production costs are unknown, unreported, and not necessarily the same drivers as those of R production costs and therefore cannot be used to develop the CERs.

The problem is how to create an estimating method with a binary data set with zero and non-zero NR production cost data. How can we fit a curve given all of the zero NR production

---

[1] Lucas, B. and White E., "A Macro Approach to Estimate Engineering and Manufacturing Development Cost Growth", Cost Engineering, 2009, 51:6, pp. 30-34.

[2] White, E.D., Sipple, V.P., and Greiner, M.A. Using Logistic and Multiple Regression to Estimate Engineering Cost Risk, Journal of Cost Analysis and Management (Summer), 2004, 67-79.

[3] Covert, R. and Wright, N., "Estimating Relationship Development Spreadsheet and Unit-as-an-Independent Variable Regressions", 2012 SCEA/ISPA Conference, Orlando, FL, June 26-29, 2012.

costs? Should we be trying to estimate NR production cost for the lots that had zero NR production costs, or do some production lots truly have no NR production costs?

For some production lots the NR production cost data show recognizable trends with the quantity produced (i.e., cost decreases as quantity increases), yet in other lots the NR production cost data are absent or have zero values. This creates a situation in which CER development is hindered by very poor fit statistics, as it is difficult to fit a line through both zero- and non-zero NR production cost data, and a lack of recognizable NR production cost drivers with which to regress the data.

In this paper we (a) provide example data for weapon systems from which we derive NR CERs, (b) discuss the methods typically used to estimate NR production costs and their weaknesses, (c) discuss the logistic regression technique, (d) show the logistic-regression-enhanced CER and methodology, and then (e) show the error analyses for the different methods of estimating NR costs.

DATA

The data consist of the NR production costs of low rate initial production (LRIP) and full scale production (FSP) lots in thousands of fiscal year 2013 dollars (FY13$K). There are 90 lots in total, 54 of which have actual (non-zero) NR production costs. Therefore, 60% of our data have non-zero NR production costs and 40% have zero NR production costs. We also include our estimate of R production lot costs (REC^ column in Tables 1-3) based upon previously developed CERs using the ZMPE UAIV regression method and various technical parameters. The first and last production units for each of these lots are gathered and an assumed lot midpoint (LMP_A) is calculated based on the pooled learning curve slope derived from all of the R production lots.

Table 1 Weapon Systems Data Table, Part 1 of 3

Task	First	Last	LMP_A	NR 2013$K	REC^
LRIP 1 of 1	1	15	6	$ 238	$ 40,273
Lot #1	1	15	6	$ -	$ 19,726
LRIP 1 of 2	1	66	22	$ 24	$ 80,370
Lot #1	1	67	22	$ -	$ 37,229
Lot #1	1	80	27	$ 56,680	$ 78,912
Lot #1	1	132	43	$ -	$ 266,255
Lot #1	16	85	44	$ 14,095	$ 50,188
Lot #1	16	90	46	$ 408	$ 108,438
Lot #2	31	138	76	$ 3,294	$ 183,361
Lot #1	1	240	76	$ 3,196	$ 370,935
LRIP 1 of 2	1	352	111	$ 16,083	$ 130,457
LRIP 2 of 2	67	170	113	$ 4,768	$ 77,785
Lot #1	1	390	123	$ 37,046	$ 206,763
Lot #2	68	251	147	$ 5,623	$ 58,365
Lot #2	86	295	177	$ 29,704	$ 99,756
Lot #2	91	290	178	$ 1,280	$ 193,524
Lot #2	81	316	182	$ 24,056	$ 131,098
LRIP 1 of 3	1	735	229	$ 29,126	$ 52,885
Lot #3	85	434	230	$ -	$ 153,683
LRIP 1 of 1	1	800	249	$ 29,635	$ 143,951
Lot #3	139	403	256	$ -	$ 313,396
Lot #1	171	488	312	$ 16,264	$ 175,895
Lot #3	241	524	371	$ -	$ 273,870
Lot #3	252	687	446	$ 5,079	$ 99,310
Lot #3	317	609	453	$ 24,969	$ 124,004
Lot #3	291	662	460	$ 23,333	$ 271,337
Lot #4	404	668	529	$ -	$ 252,485
Lot #2	489	788	631	$ 2,655	$ 134,516
Lot #5	525	974	735	$ 5,872	$ 353,932
Lot #4	435	1178	767	$ 163	$ 228,237

Table 2 Weapon Systems Data Table, Part 2 of 3

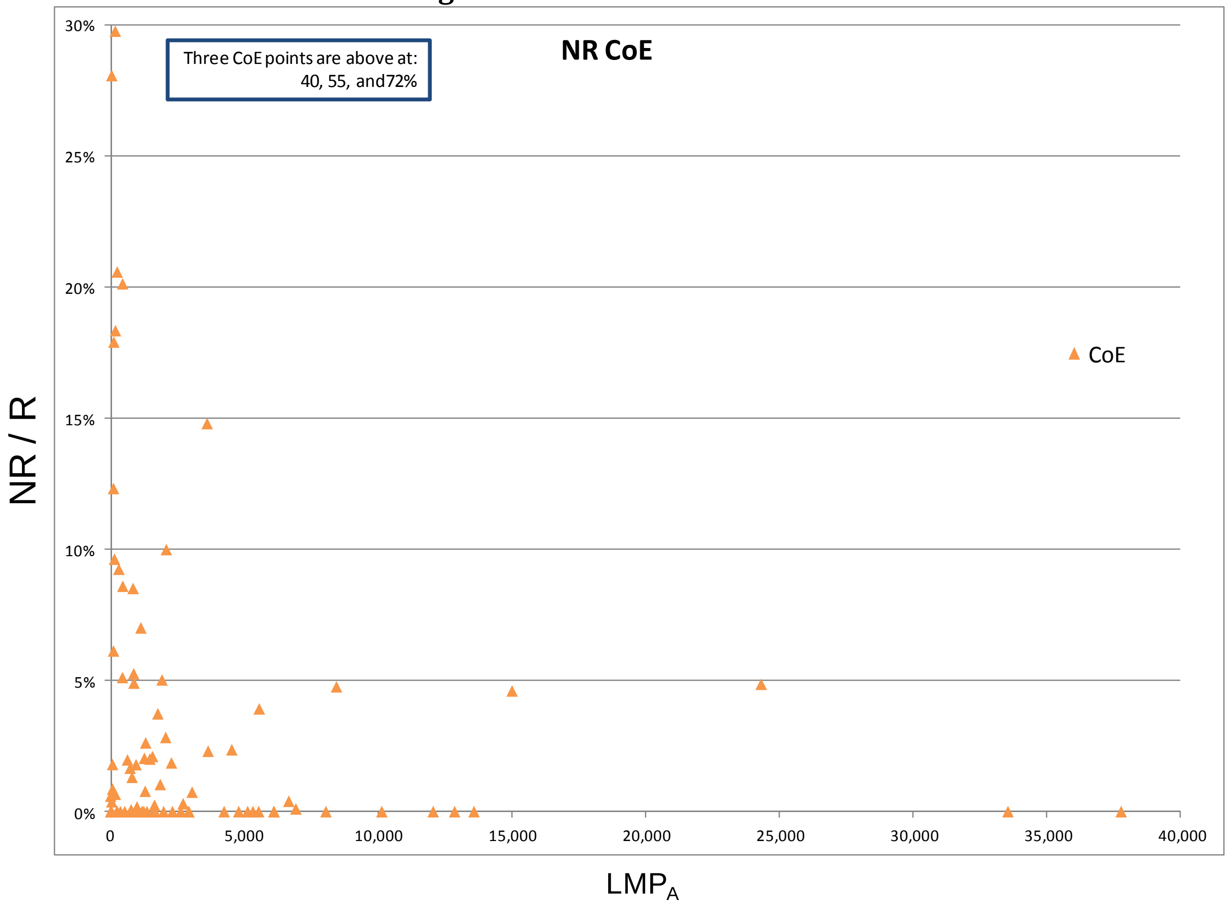
Task	First	Last	LMP_A	NR 2013$K	REC^
Lot #5	669	933	796	$ -	$ 223,493
LRIP 2 of 2	353	1408	805	$ 2,837	$ 216,673
Lot #4	506	1255	844	$ 19,042	$ 223,610
Lot #4	688	1061	866	$ 3,679	$ 69,924
Lot #4	663	1112	875	$ 13,286	$ 271,024
Lot #4	610	1259	909	$ 89,796	$ 223,468
Lot #3	789	1139	957	$ 2,493	$ 138,995
LRIP 2 of 3	874	1074	972	$ -	$ 36,498
Lot #1	1	3218	990	$ 657	$ 371,030
Lot #6	934	1086	1009	$ -	$ 120,251
Lot #2	391	1880	1016	$ -	$ 420,420
LRIP 2 of 3	736	1604	1134	$ 2,719	$ 38,786
Lot #7	1087	1266	1175	$ -	$ 135,183
Lot #6	975	1406	1182	$ -	$ 294,867
Lot #5	1113	1382	1244	$ -	$ 146,412
Lot #4	1140	1397	1266	$ 1,913	$ 94,002
Lot #5	1062	1555	1298	$ 638	$ 81,853
LRIP 2 of 3	1075	1572	1313	$ 2,170	$ 82,665
Lot #8	1267	1461	1362	$ -	$ 140,127
Lot #5	1398	1545	1471	$ 1,033	$ 51,564
Lot #6	1546	1595	1570	$ 359	$ 17,083
Lot #7	1407	1836	1615	$ -	$ 267,412
Lot #6	1383	1922	1643	$ 658	$ 269,552
Lot #9	1462	1881	1666	$ -	$ 284,266
Lot #6	1556	1996	1770	$ 2,488	$ 66,624
Lot #5	1256	2565	1861	$ 3,192	$ 308,631
Lot #1	1409	2508	1925	$ 8,744	$ 174,059
Lot #8	1837	2136	1984	$ -	$ 175,477
LRIP 3 of 3	1573	2610	2063	$ 4,264	$ 150,598
LRIP 3 of 3	1605	2620	2086	$ 3,781	$ 37,814

Table 3 Weapon Systems Data Table, Part 3 of 3

Task	First	Last	LMP_A	NR 2013$K	REC^
Lot #7	1997	2574	2278	$ 1,505	$ 81,001
Lot #9	2257	2376	2316	$ -	$ 67,027
Lot #3	1881	3473	2625	$ -	$ 338,724
Lot #6	2566	2865	2714	$ 191	$ 63,158
Lot #8	2537	3305	2910	$ -	$ 323,726
LRIP 3 of 3	2611	3264	2930	$ -	$ 85,471
Lot #2	2061	4197	3048	$ 1,353	$ 182,132
Lot #6	2954	4323	3610	$ 46,228	$ 312,283
Lot #1	3265	4052	3649	$ 2,224	$ 96,457
Lot #4	3474	5080	4244	$ -	$ 296,102
Lot #1	4053	5036	4533	$ 2,666	$ 112,912
Lot #12	4363	5239	4792	$ -	$ 318,187
Lot #3	4198	6127	5123	$ -	$ 140,901
Lot #2	3219	7861	5320	$ -	$ 324,151
Lot #13	5240	5837	5535	$ -	$ 207,843
Lot #1	2621	9395	5558	$ 7,380	$ 188,288
Lot #14	5838	6372	6102	$ -	$ 180,615
Lot #5	5081	7227	6113	$ -	$ 354,833
Lot #15	6373	6954	6661	$ 760	$ 191,423
Lot #2	2509	13131	6928	$ 1,190	$1,147,575
Lot #5	7228	8906	8048	$ -	$ 255,653
Lot #3	7041	9959	8446	$ 13,250	$ 278,251
Lot #6	8907	11433	10136	$ -	$ 359,212
Lot #7	11434	12700	12060	$ -	$ 171,012
Lot #8	12701	13022	12861	$ -	$ 42,636
Lot #4	11046	16343	13583	$ -	$ 464,430
Lot #2	9396	21714	15007	$ 11,719	$ 254,638
Lot #3	21715	27048	24318	$ 4,642	$ 95,480
Lot #5	30834	36347	33542	$ -	$ 222,375
Lot #6	36348	39210	37767	$ -	$ 111,450

Using these data, we calculate a cost-on-estimate (CoE) factor for each lot. A CoE factor uses the actual costs of the item to be estimated as the numerator and estimated costs for the denominator. In this case we use actual NR production costs divided by estimated R production costs (REC^), since we are trying to estimate the NR production cost based on our R production cost estimate. Using the CoE factor and the $LMP_{A,}$ the NR production cost data are plotted on a scatter plot (Figure 1) in order to analyze the data for apparent trends.

Figure 1 NR Production CoE Chart



The scatter plot in Figure 1 shows the large scatter of NR production CoE factors and the decreasing trend of NR production CoE factors as the LMP$_A$ increases. Figure 1 also shows a large number of lots that have NR production CoE factors equaling zero, which indicates the binary nature of weapon systems NR production costs.

NON-RECURRING ESTIMATING METHODS

Two estimating relationship forms, a factor and a log-unit-based CER, are created from the data. The factor is:

$y = a\varepsilon$, where (1)

a = the ratio of NR/R production costs, and

ε = the multiplicative error of the factor

The log-unit CER is:

$y = [a + b * \ln(Unit)]\varepsilon$, where (2)

a and b are coefficients of the regression,

$\ln$ = the natural logarithm,

$Unit$ = the unit number represented by LMP_A, and

ε = the multiplicative error of the CER

Neither of these methods estimates the NR production costs well, since they did not account for the binary nature of the NR data. One way to address the binary nature of the NR production cost data is through the use of logistic regression. Using logistic regression we can predict which production lots should have NR production costs and which production lots should not. After determining the best logistic regression for the weapon systems data – based on zero bias, lowest percent standard error (PSE), and highest hit rate – we combine the logistic regression with a CER regression to develop the NR production cost estimate.

We define the PSE as:

$$PSE = 100\% * \sqrt{\frac{1}{n-m}\sum_{i=1}^{n}\left[\frac{y_i - f(x_i)}{f(x_i)}\right]^2}, \text{ where} \quad (3)$$

n = number of data points

m = number of coefficients

y_i = "actual" data point i

$f(x_i)$ = estimate of data point i

and the percent bias as:

$$Percent\ Bias = 100\% * \frac{1}{n}\sum_{i=1}^{n}[y_i - f(x_i)], \text{ where} \quad (4)$$

n = number of data points

y_i = "actual" data point i

$f(x_i)$ = estimate of data point i

Factor Method

In the absence of other known, reported cost drivers, the factor method is typically used to estimate NR production costs from R production costs estimates. Since NR costs usually make up a very small percentage of total production costs, any errors in estimating NR production costs would most likely have a small impact on the total hardware cost estimate. We calculate the factor using two methods, both using the form in Equation 5:

$$y = ax \quad (5)$$

$a =$ The ratio of NR/R production costs

$\varepsilon =$ Multiplicative error

Two different factors are developed – one through with weighted average and the other through the ZMPE method. The weighted average is the more commonly used method when developing a cost factor, and is very simple to compute (the total of all NR production costs divided by the total of all R production costs). However, because the methods to follow were all created using the ZMPE technique, we chose to develop a factor using the same method in order to accurately compare the results. One reason is the weighted average factor developed using the weapon systems data is biased and does not provide an equal comparison with the unbiased relationships developed using ZMPE.

There are several drawbacks to using the factor method. This method predicts an NR cost for every production lot, regardless of whether there were actual NR costs for that lot or not. It also estimates the same ratio of NR production costs to R production costs for every lot regardless of quantity. For the weighted average factor method the results can be biased. The ZMPE factor method estimates a factor without any bias. However, both factor methods have poor PSE and R^2 statistics partially due to the zero-cost lots.

Figure 2 Factor Method Chart

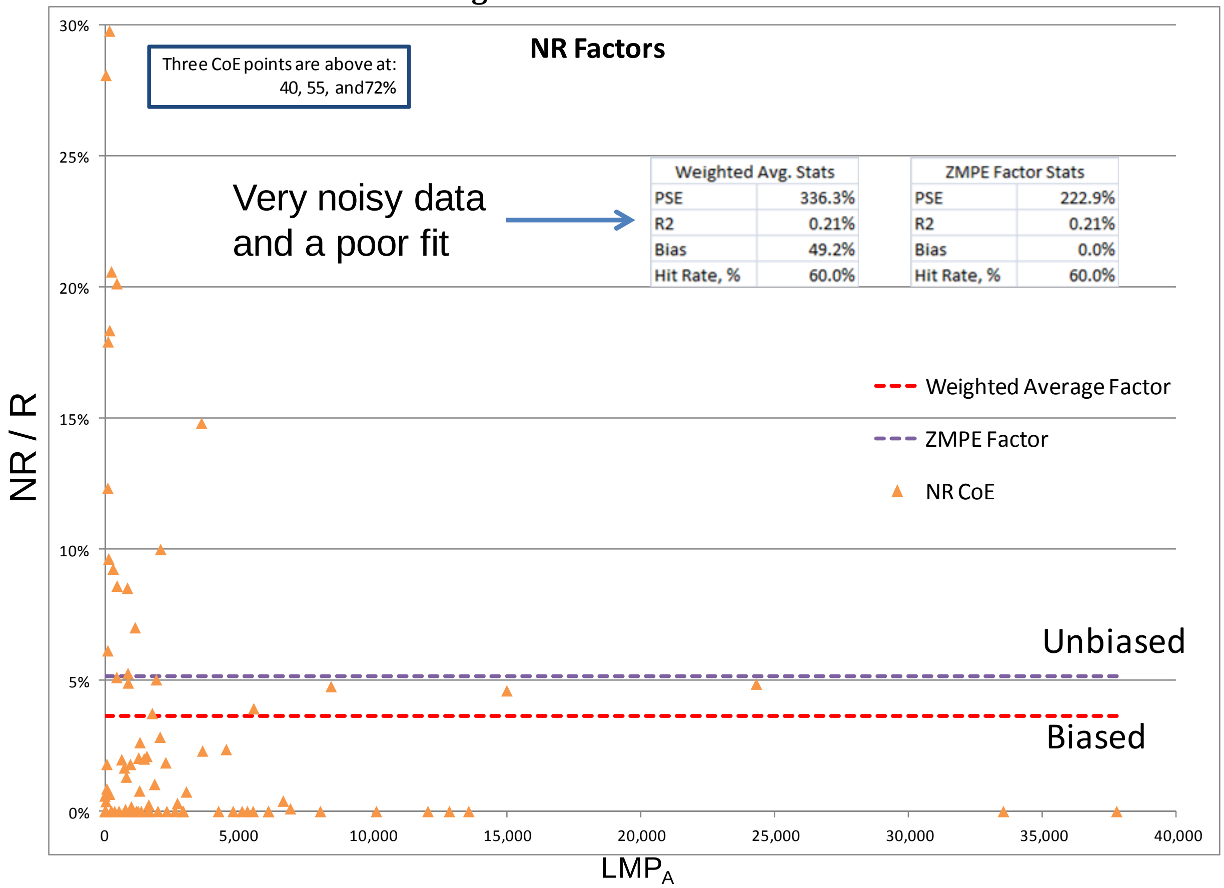


Figure 2 shows both factor methods do not provide a good estimate of NR production cost. The weighted average factor is biased for the data and the presence of bias makes it difficult to compare to the other, unbiased results. The weighted average factor has a high PSE (336%) and very low Pearson's Correlation Squared (R^2) of 0.2%. While the ZMPE factor is unbiased, it has a high PSE (223%) and a very low R^2 of 0.2%. The poor fit statistics were due to the large number of zero NR production costs as well as the inability to account for the quantity trend indicated by LMP_A.

CER Method

A CER can be an attractive alternative to a factor; however, NR production costs are often unknown, unreported, and not necessarily the same drivers as those of R production costs. The only cost driver considered, therefore, was quantity. The plot in Figure 1 shows a decrease in cost with an increase in quantity, which indicates that a natural logarithm (ln) of quantity would provide an acceptable solution. The equation we used for the CER method was:

$y = [a + b\ln(Unit)]\,\varepsilon$, where (6)

$y =$ the ratio of NR/R production costs,

a and b are coefficients of the regression

$Unit$ = The unit for which NR production cost factor is being determined

ε = multiplicative error

There are several drawbacks to this particular CER. This CER also predicts a NR production cost for every lot, regardless of whether there were actual NR costs for that lot or not. The error statistics for this CER were also poor due to the presence of zero NR cost lots, but the PSE and R^2 statistics are improved over the factor method since the CER responds to changes in NR production costs due to quantity. Also, the “ln” term in Equation 6 results in a negative value when the “unit” variable is large– which is impossible. This can be mitigated by estimating NR production costs as zero if the CER estimates a negative cost. The quantity data used to create the CER are not large enough for this to happen, so this concern applies only to quantity data that are outside the range of applicability of the data.

Figure 3 CER Method Chart

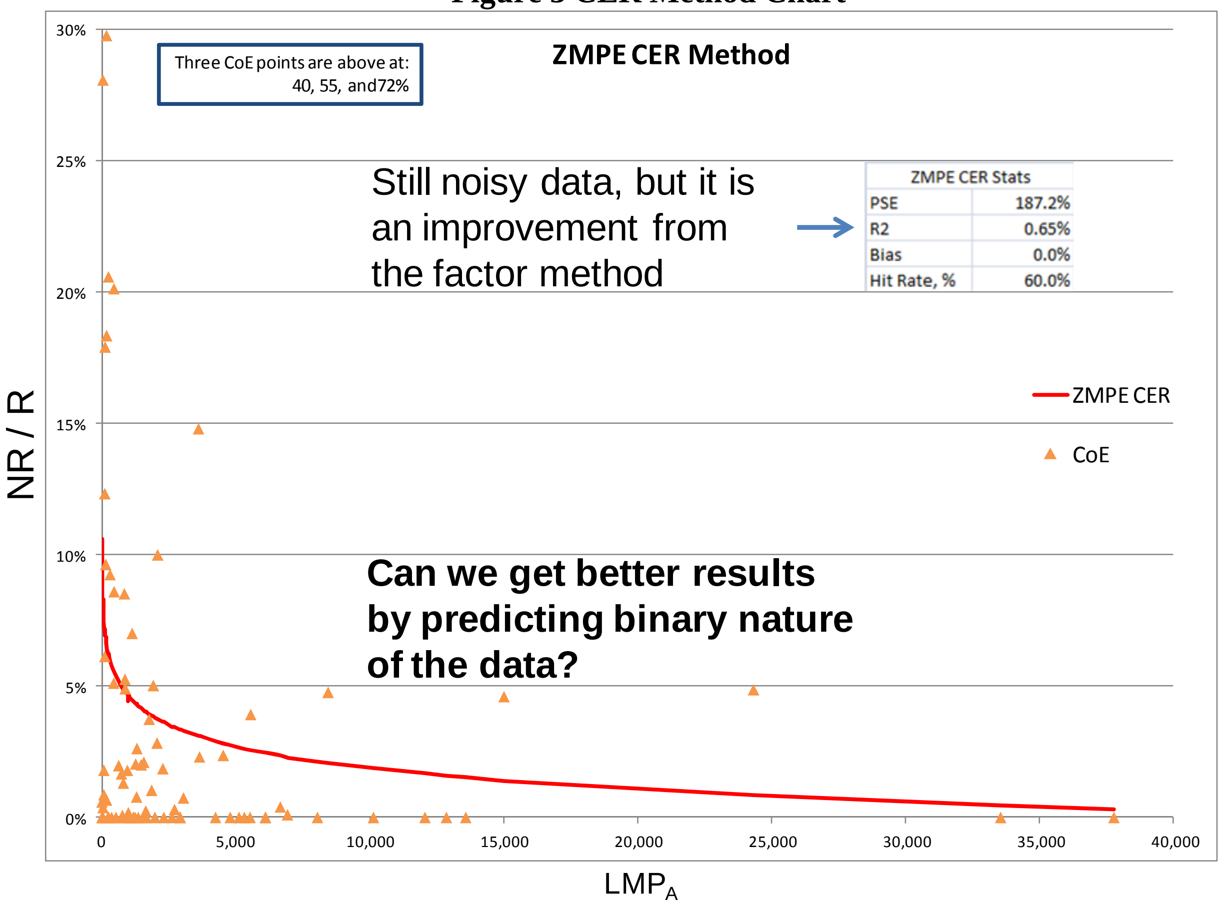


Figure 3 shows the fit of the CER to the data. The error statistics are better than those of the factor method with a PSE of 187% and an R^2 of 0.7%. The regression line fits the CoE data better since the equation is unit-driven, but the error statistics are still poor partially due to the zero-cost NR production lots for which this method does not account (i.e., it predicts a NR cost even for those lots that do not have NR cost). The ZMPE method was used to develop the CER so it is unbiased.

Logistic Regression

Logistic regression is a technique used to find relationships between a categorical independent variable and a series of dependent variables. It is used to predict the outcome based on the series of dependent variables. It is commonly used in social sciences (e.g. to determine whether one is disposed to diabetes or whether a given person will exhibit different types of behavior based on a series of factors), but very rarely used in cost analysis. Logistic regression has been used in studies of cost growth (Lucas & White, 2009) (White, Sipple, & Greiner, 2004), [1,2] and is used in this paper to model the binomial (i.e., zero and non-zero) behavior of NR production cost data.

Logistic regression uses the logistic function, $\pi(x)$, shown in Equation 7 to relate the dependent variables, x_i, to a range of values between zero and one (i.e. [0,1]). We use the logistic function to determine the likelihood of NR production costs.

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}; 0 \leq \pi(x) \leq 1 \text{ , where} \tag{7}$$

$g(x)$ is the logit function of x (typically, but not necessarily, a linear relationship)

The function $g(x)$ is called the logit function which represents the log-odds of an event taking place.

$$g(x) = a + bx_1 + cx_2 + dx_3 + ex_4 + \varepsilon \text{ , where} \tag{8}$$

$a, b, c, d,$ and e are coefficients of the regression

$\varepsilon =$ Additive error

The best coefficients and *a priori* variables are found by selecting those producing the highest 'hit rate' for the data. 'Hit rate' is the percentage that the logit function correctly predicts the presence or lack of NR production costs. The hit rate is calculated by taking the number of times the logit function correctly predicted the presence or lack of NR production costs divided by the total number of observations in the data. If the logistic function correctly predicted the presence of NR production costs, then it was considered a hit. If it did not correctly predict the presence of NR production costs, then it was considered a miss. The ZMPE regression method with an additive error was used to find the coefficients (and best choice of variables) of the logit function that provided the best hit rate. The best logistic regression found for the weapon systems data is:

$$g(x) = a + bx_1 + cx_2 + dx_3 + \varepsilon \text{ , where} \tag{9}$$

$a, b, c,$ and d are coefficients of the logistic regression
$x_1 =$ Natural log of the LMP_A, using the pooled learning curve slope
$x_2 =$ Lot number, integer value for FSP lots and a fraction for each LRIP lot based on total program LRIP lots (e.g. for a program with 2 LRIP lots: LRIP 1 = 0.33, LRIP 2 = 0.66)
$x_3 =$ First unit in the lot divided by 1000
$\varepsilon =$ Additive error

Table 4 shows the coefficients and fit statistics for the $g(x)$ used in this paper.

Table 4 Logit Function Statistics

$g(x)$	
a	2.7603
b	-0.1230
c	-0.3483
d	-0.0441
Measurement Data	
Observations	90
Coefficients	4
Degrees of Freedom	86
Error Statistics	
PSE	52.8%
Bias	(0.00)
Hit Rate, %	73.3%

The hit rate for the logit function using these data is 73.3%. The logit hit rate is an improvement over the naïve hit rate of the previous methods that apply a NR production cost to every lot, which was only 60%. The naïve hit rate for the other methods is the total number of production lots that have NR production costs divided by the number of total production lots in the data.

We use the logistic function in Equation 7 to determine the likelihood of NR production costs using the following parameters:

$$f_{\pi(x)} = \begin{cases} 1, & if\ \pi(x) \geq 0.5 \\ 0, & otherwise \end{cases}$$

(10)

Equation 10 can be interpreted as:

$$\begin{matrix} NR\ deemed\ present & if\ \pi(x) \geq 0.5 \\ NR\ deemed\ not\ present & if\ \pi(x) < 0.5 \end{matrix} \quad (11)$$

Using these results we are able to categorize the data into those lots where we predict the presence of NR production costs and those lots where we predict there are no NR production costs. A plot of the two data categories is shown in Figure 4.

Figure 4 NR CoE with Logistic Regression Results

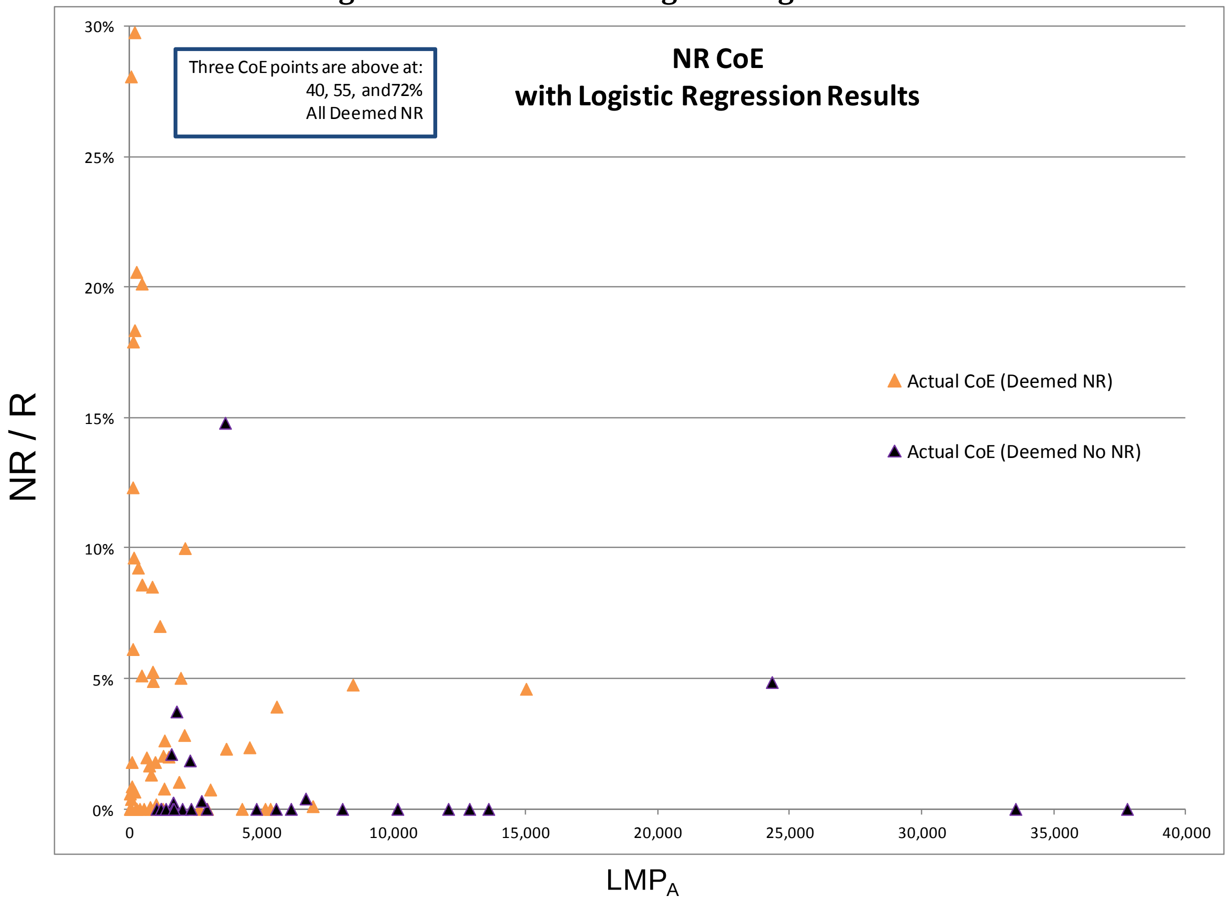


Figure 4 shows which lots Equation 10 correctly and incorrectly predicts the presence of NR production costs. The y-axis values of the triangles on the graph are the ratios of the NR production costs to the R production cost estimates, and the color of the triangles represents the results of the logistic function.

If the logistic function correctly predicts the presence of NR production costs, then it is considered a hit. If it does not correctly predict the presence of NR production costs, then it is considered a miss. Logistic regression correctly predicts NR production costs for the majority of lots that have actual NR production costs. Of the 54 weapon system lots that have actual NR production costs, the logistic function predicts that all but eight of those lots have NR production costs. The logistic function improves our ability to predict the presence of NR production costs from the naïve hit rate of 60% to 73.3%.

Logistic-Regression-Enhanced CER

The logistic function allows some predictive ability of the presence of NR production costs or the lack thereof. It can be used to develop a better NR production CER than the previously demonstrated methods since it allows us to 1) categorize the data into zero- and non-zero NR production cost data, and 2) with the non-zero NR production cost data, estimate those costs using a CER. The logistic function can be combined with a log-unit CER with the same form as Equation 6. The combined logistic-regression-enhanced CER is:

$$y = [f_{\pi(x)}(a + b\ln(Unit))]\varepsilon \quad \text{, where} \qquad (12)$$

y = NR/R factor

a and b are coefficients of the regression

$f_{\pi(x)}$ = Logistic function

$Unit$ = The unit for which NR production cost factor is being determined

ε = Multiplicative error

We regressed all the lots, including the lots that had zero actual NR production costs, using the ZMPE method with the goal of finding a logistic-regression-enhanced CER with the lowest PSE. Due to the logistic regression function, we had to change how PSE was calculated for two instances: 1) if the logistic regression produced a hit and there was no actual NR production cost, then the PSE for that lot is 0 and 2) if the logistic regression produced a miss and there was an actual NR production cost, then the PSE for that lot is -1. Otherwise PSE is calculated using Equation 3.

While all program lots were used in the complete regression and in the calculation of the error statistics, only those where the logistic function deemed NR production costs were technically used in calculating the log-unit portion of the equation (e.g. only the orange triangles in Figure 5). This is because the ZMPE regression method solves for the lowest PSE by adjusting the coefficients a and b until the lowest PSE is reached, yet the PSE for those lots where it was deemed no NR production costs remains static regardless of how a and b change.

The estimated NR production cost is needed in order to find the PSE, for the CER method this was simply the NR/R factor (the result of the log-unit CER) the CER produced multiplied by the estimated R production costs. The difference with the logistic-regression-enhanced CER is that the NR/R factor was calculated as the result of the log-unit CER multiplied by the logistic function, which produces a 0% factor for those lots which the logistic regression predicts there will be no NR production costs. Since the NR/R factor will always be zero for those lots, the PSE (for those lots) remains static regardless of the values of the coefficients of the log-unit CER (either 0 or -1) and therefore they do not influence the solution of coefficients a and b in the ZMPE regression.

Figure 5 Logistic-Regression-Enhanced CER

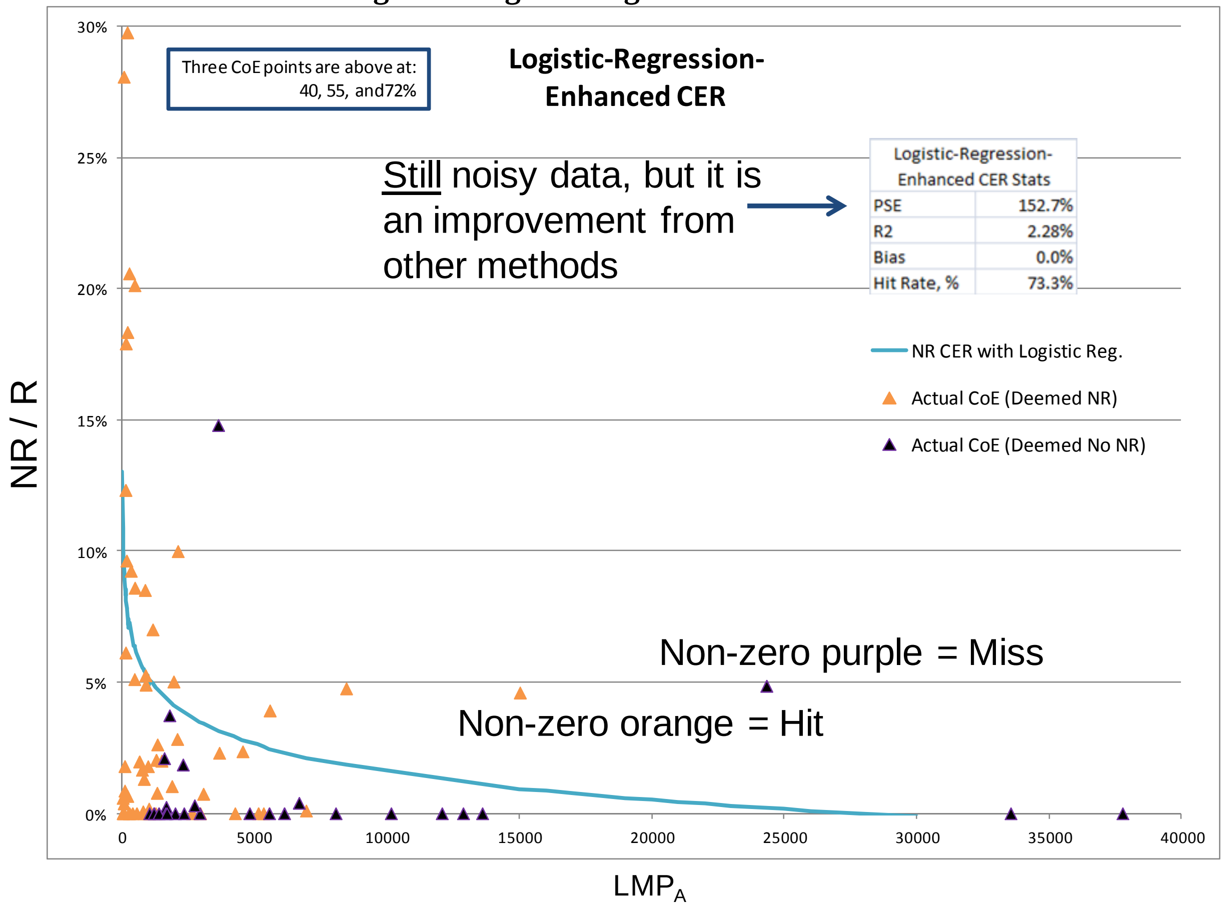


Figure 5 shows the logistic-regression-enhanced CER. The blue line shows the estimates of the log-unit portion of Equation 12. The estimates for the lots where the logistic function predicts there will be zero NR production costs are not plotted since they are all zeros. Even with this noisy data, the error statistics for the logistic-regression-enhanced CER have improved over the other methods with a PSE of 153% and an R^2 of 2.3%.

COMPARISON

After all three methods were completed we compared the three using error statistics and hit rate as shown in Table 5.

Table 5 Error Statistics

ZMPE Factor	
PSE	223%
R^2	0.2%
Pct Bias	0.0%
Hit Rate, %	60.0%

CER Method	
PSE	187%
R^2	0.7%
Pct Bias	0.0%
Hit Rate, %	60.0%

Logistic Reg + CER	
PSE	153%
R^2	2.3%
Pct Bias	0.0%
Hit Rate, %	73.3%

While the PSE and R^2 statistics are poor for all of these methods, we see improvements with each respective method. The unit-driven CER is an improvement over the simple factor method that does not take quantity into account. The logistic-regression-enhanced CER is an improvement over the unit-driven CER due to the ability of the logistic regression to account for the binary behavior of the NR production cost data.

The weighted average factor is not compared to the other models as it is biased and therefore no direct comparison can be drawn between it and other, unbiased methods.

Figure 6 All Methods

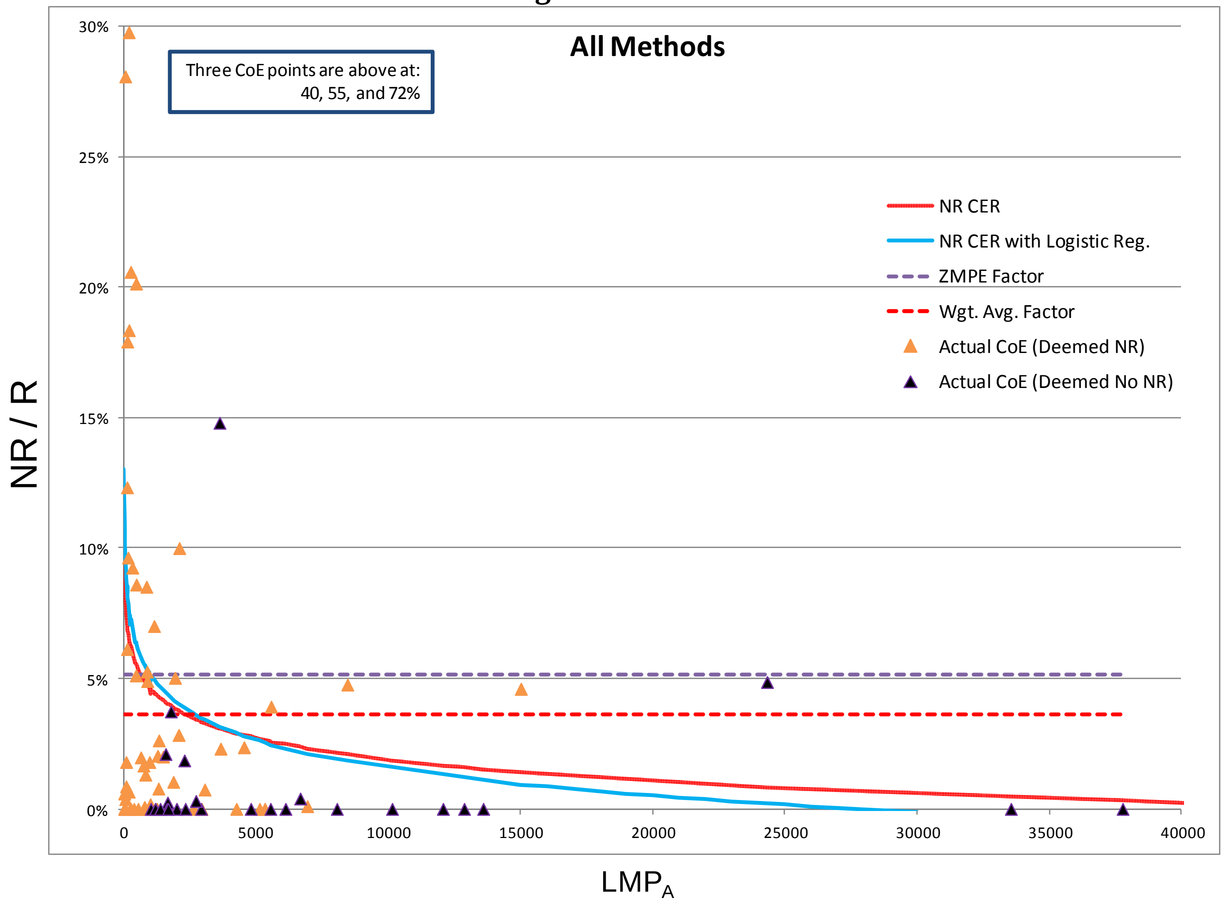


Figure 6 shows the different methods of calculating NR production costs on one graph and illustrates the different ways to calculate NR/R. The logistic-regression-enhanced CER has a higher peak and decays faster than the CER method. It crosses the x-axis before the unit-driven CER due partially to the fact that it predicts zero NR production costs for the later lots.

CONCLUSION

NR production costs are difficult to estimate due to the binary nature of the data, the amount of scatter of the data, and lack of established NR production cost drivers with which to perform regressions. This forced us to create NR factor-type CERs. To improve the ability to estimate these costs, we combined logistic regression with CER development techniques. Logistic regression provided us the ability to predict the binary nature of the data and improve the CER statistics.

The logistic-regression-enhanced CER has the best error statistics of the three methods we presented. The PSE of the logistic-regression-enhanced CER decreases 32% from the factor method and decreases 19% from the CER method. The logistic-regression-enhanced CER also models the binary nature of the data better, and improves the hit rate 22% over the other methods. The inherent correlation between the actual and the estimates is starting to improve (through increased R^2). Although the R^2 of the logistic-regression-enhanced CER is still quite low, it is much greater than the other methods. Given the improvements in fit statistics using the logistic-regression-enhanced CER, using logistic regression in CER development offers a promising approach to modeling cost data that are binary in nature (i.e., include both zero- and non-zero-costs).

ACRONYMS

a,b	Coefficients of the regression
CER	Cost Estimating Relationship
CoE	Cost-on-Estimate
ε	Error
FSP	Full-Scale-Production
FY13$K	Fiscal Year 2013 Dollars in Thousands
LMP_A	Assumed Lot Midpoint
ln	Natural Log
LRIP	Low-Rate-Initial-Production
PSE	Percent Standard Error
NR	Non-Recurring
R	Recurring
R^2	Pearson’s Correlation Squared
REC^	Estimated recurring production costs
UAIV	Unit-as-an-independent variable
ZMPE	Zero percent bias, minimum percent error

REFERENCES

Covert, R. & Wright, N. (2012). Estimating Relationship Development Spreadsheet and Unit-as-an-Independent Variable Regressions. *2012 ISPA/SCEA Conference,* Orlando, FL.

Lucas, B. & White, E. (2009). Macro Approach to Estimate Engineering and Manufacturing Development Cost Growth. *Cost Engineering , 51* (6), 30-34.

White, E., Sipple, V., & Greiner, M. (2004). Using Logistic and Multiple Regression to Estimate Engineering Cost Risk. *Journal of Cost Analysis and Management , Summer*, 67-79.