CUMAV OR UNIT:
IS CUMULATIVE AVERAGE VS. UNIT THEORY A FAIR FIGHT?

B. L. Cullis[1], R. L. Coleman[2], P. J. Braxton[3], J. T. McQueston[4].
[1,3,4] Northrop Grumman, IT TASC, 15036 Conference Center Drive, Chantilly, VA 20151, USA.
[2]Northrop Grumman, IT, 15036 Conference Center Drive, Chantilly, VA 20151, USA.
bethia.cullis@ngc.com

Introduction:

When considering learning curve in a production cost estimate, several factors influence the projected impact of learning. The learning curve slope, theoretical first unit cost, and position on the learning curve are typically discussed and researched; the choice of learning curve theory, however, tends to be studied less closely. To determine the theory to use, practitioners look to the statistics of regression to judge significance and "best fit"; problematically, these statistics are inherently biased towards Cumulative Average, as the Cumulative Average metric is smoothed by the very fact that it is cumulative. This paper will use generated data with and without error terms in an attempt to show the degree of this bias.

In addition, it is generally true that most practitioners in some commodities lean more toward one theory than the other. This is a puzzling phenomenon; industry preferences seem to indicate that smaller objects (like electronics) follow a cumulative average curve while larger objects (like ships) follow a unit curve. To decide whether one model or the other is best, either globally or locally, is beyond the scope of this paper; in lieu of that project, the authors hope that by shedding light on this problem they can enable all practitioners to decide whether their preference is rational and, in doing so, perhaps bring some rigor to this problem.

Background:

Two learning curve theories exist: Cumulative Average (CUMAV) and Unit Theory. Cumulative Average Theory states that the *cumulative average* cost of production decreases at a constant rate when production volume doubles. CUMAV curves are often called Wright Curves, after T. P. Wright, who first published this theory in 1936 as part of the article "Factors Affecting the Cost of Airplanes" (*Journal of Aeronautical Sciences).*[1] On the other hand, Unit Theory states that the *unit* cost of production decreases at a constant rate when production volume doubles. Unit theory was developed as part of a Government study performed by the Stanford Research Institute; the purpose of the study was to assess the validity of the improvement curve concept. Unit curves are also referred to as Crawford Curves.[2] Because the two theories have the same functional form but a different dependent variable, the actual cost improvement represented by a nominal learning curve slope (e.g., "90%") is different between the two; specifically, CUMAV represents steeper learning that Unit for the same nominal LCS.

In the investigation that follows, the two theories will be compared. This comparison will be executed using perfect data and injected error terms. In practice, perfect data does not exist. Before determining a learning curve theory, it is imperative that considerable care is exercised in the treatment of data. As demonstrated in an earlier paper[3], large errors result when data is not normalized and effects other than learning are not removed. These other effects can easily overwhelm learning effects and, as demonstrated in the aforementioned paper, can create learning curves with slopes as much as 6%-18% off. Such effects as change orders (COs) or engineering change proposals (ECPs), labor force composition, and percent overlap between successive units exert considerable influence on data. A learning curve derived from data that contains other effects beyond simply learning will herein be referred to as a "pseudo learning curve". If other (non-learning) effects are present, but not in sufficient force to overwhelm learning, pseudo learning curves often demonstrate statistical significance; statistical significance does not make this pseudo learning curve a valid curve from which to estimate.

Pseudo learning curves are most easily discovered graphically. Other effects may be present when:

1. The data demonstrate patterns of rising or falling above and below the regressed curve, in which case other long-term effects are present and must be considered. This consideration is easily tested using conventional tests such as non-parametric tests of "runs" and the like.
2. The learning curve substantially shifts or changes slope. This can occur for a number of reasons: i.e., a shift in percent overlap of consecutive units, shifts in labor force composition, or shifts in manufacturing location or process.
3. The curve fails to pass through or "near" more than a small fraction of the data points. "Sufficient nearness" is difficult to determine, though the sheer size of the object, in manhours, is a reliable indicator. Very large objects, such as ships, or objects requiring many hours to complete, which may be smaller but more complex, are likely to exhibit "law of large numbers" effects (perhaps more correctly, Chebyshev's inequality comes into play) and the expectation of nearness is greater. This consideration is not easily quantifiable, but as a single sense of the scale, large ships have been shown[4] to exhibit error terms as low as 2% after normalization.

Further, statistical significance must, of course, be tested in any determination of learning curve. An analyst should have no difficulty rejecting a regression of either form when it is not statistically significant. The focus here will be on the goodness of fit of the model; this is the issue that presents difficulty to the practitioner. As will be shown, goodness of fit metrics implicitly favor the CUMAV theory; the degree of the distortion is surprising.

Investigation:

The study for this paper was done using generated data; this allows the advantage of knowing the underlying theory before examining the statistics. CUMAV and Unit data were created; the data were then regressed twice. First, the data were regressed assuming the known underlying theory (i.e., CUMAV-as-CUMAV and Unit-as-Unit), then assuming the other theory (i.e., CUMAV-as-Unit and Unit-as-CUMAV). The statistics of the regression (standard error) were compared, as they would be in practice, to determine which theory best fits the data set. The perfect data were examined first, followed by data with injected error.

The generation of the data was problematic, specifically the error term. In particular, if there is an incorrect formulation in our analysis, it would most likely be in the CUMAV error term. The form of the error term was carefully determined; a simple normal error term was selected for both models, and this error term was imposed upon the individual units of data in each case. For the CUMAV model, an imposition of error on each subsequent unit may seem at odds with the formation of the model; to impose a cumulative average error, however, seemed to suggest that the history had changed. For this reason, the error was imposed on each subsequent unit, and the unit error was allowed to perturb the Cumulative Average. Though more straightforward, the Unit model also presents a difficulty. The same absolute error was imposed on each unit in turn; it may also be reasonable to assume that error terms diminish with each unit. This case could be modeled as a specific percentage. Bearing in mind these concerns, no claim is made as to correctness of the precise form of the error term. The form selected was reasonable for the purposes of the exploration of this issue.

The regression fits were determined using the traditional approach of log-log transformation and applying ordinary least squares (OLS) regression to the transformed data, which is also inherent in the Excel power-curve "Trend Lines" on the graphs below. As a topic for further research, another regression technique such as minimum unbiased percent error (MUPE) could be used in conjunction with an injected multiplicative error term, though such an excursion would not be expected to alter the fundamental findings of this paper.

Perfect Data:

Perfect data was generated with the following parameters:

1. The Theoretical First Unit: 100 (for both CUMAV and Unit)
2. The Learning Curve Slope: 80% (for both CUMAV and Unit)

As expected, when regressing with the known theories, the r^2 is 1 and the standard error is 0. When regressing assuming the other theory, CUMAV-as-Unit results in a markedly higher standard error (0.06243) than does Unit-as-CUMAV (0.015174). See Figure 1 and Figure 2 below, respectively.

LC	80%	
T1	100	
Unit	Cost (CAC)	Cost (Unit)
1	100.00	100.00
2	80.00	60.00
3	70.21	50.63
4	64.00	45.37
5	59.56	41.82
6	56.17	39.19
7	53.45	37.13
8	51.20	35.46
9	49.29	34.05
10	47.65	32.86
Standard Error	0	0.062431

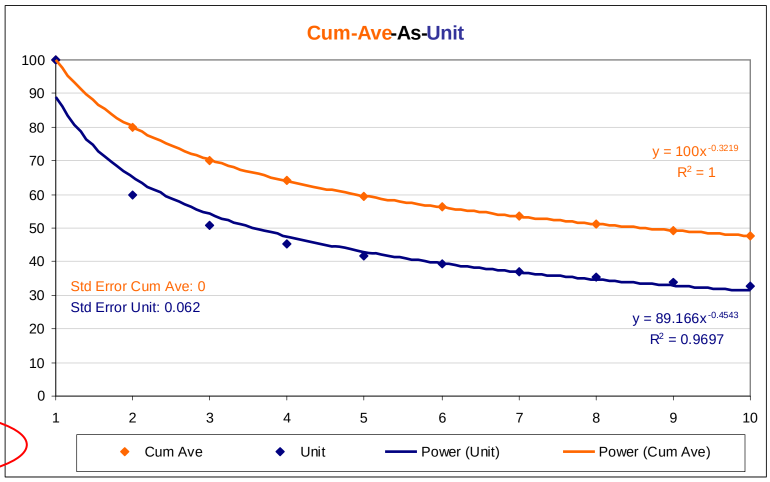

Figure 1.

LC	80%	
T1	100	
Unit	Cost (Unit)	Cost (CAC)
1	100.00	100.00
2	80.00	89.65
3	70.21	83.87
4	64.00	78.67
5	59.56	74.90
6	56.17	71.87
7	53.45	69.31
8	51.20	67.34
9	49.29	65.50
10	47.65	63.53
Standard Error	0	0.015174

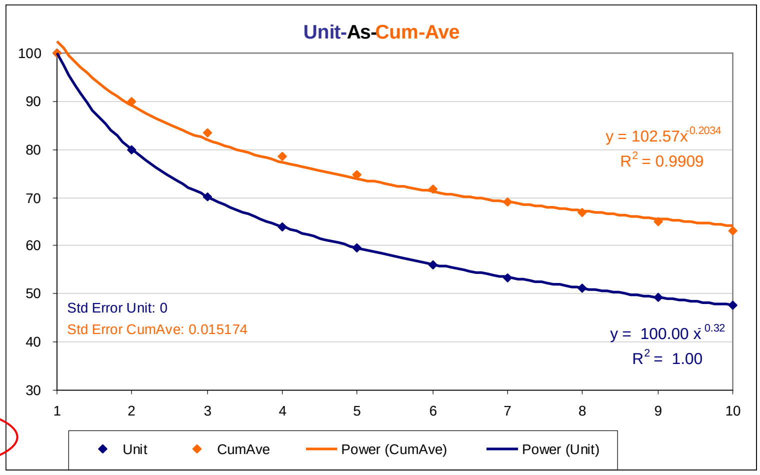

Figure 2.

Imperfect Data:

With perfect data, statistics reliably direct an analyst to the correct model. In practice, however, data is never perfect. To generate imperfect data that has the characteristics of real data, random error was injected into each model; the size of the random error was based on observed error. In a learning study[4] produced previously, the authors observed a coefficient of variation (CV) at the 8th Unit (of a 17 Unit series) of 0.0187. Using this information, the curves with error were generated as follows:

<u>Unit</u>

1. A perfect curve was generated using the same parameters as before:
   a. T1 = 100
   b. LC = 80%
2. 17 units of data were projected.
3. A random normal error term was generated for each unit with a
   a. Mean of 0
   b. Standard deviation computed using the CV of 0.0187 multiplied by the cost of the eighth “perfect” unit
4. The error term was added to each “perfect” unit cost
5. 500 imperfect data sets were found (with a new random-error-term set for each run)

<u>CUMAV</u>

1. A perfect curve was generated using the same parameters as before:
   a. T1 = 100
   b. LC = 80%
2. 17 cumulative average data points were projected.
3. The unit metric for each point was found; for example:
   a. If X=1, Y=100 (cumulative average cost)
   b. If X=2, Y=80 (cumulative average cost)
   c. (100 + 80)/2 = 60 (unit metric of unit 2)
4. The error term was generated as described above.
5. The error term was added to each unit metric derived from the “perfect” CUMAV data
6. 500 imperfect data sets were found (with a new random-error-term set for each run)

As before, the data was regressed using the known theory and compared to a regression using the other theory. A sample output (one of the 500 runs) is graphed below (Figures 4 and 5). This output demonstrates how little variation was added to perfect data; if confronted with data resembling the below in “real life”, an analyst would not hesitate to derive a learning curve slope. A simulation of 500 trials shows that, with this error term, an analyst should never mistake CUMAV data for Unit data; on the other hand, *59% of the time* the standard error indicates that CUMAV is a better fit for Unit data than Unit Theory is. On average, the standard error for CUMAV Theory as applied to Unit data is *lower* than the standard error of Unit Theory applied to the same data. 500 runs of each trial yields the following:

CV: .0187	Average SE	# of "false positives"	% of "false positives"
Unit-as-Unit	0.018963337		
Unit-as-Cum-Ave	0.017925980	295	59%
Cum-Ave-as-Cum-Ave	0.002934640		
Cum-Ave-as-Unit	0.061955773	0	0%

Figure 3.

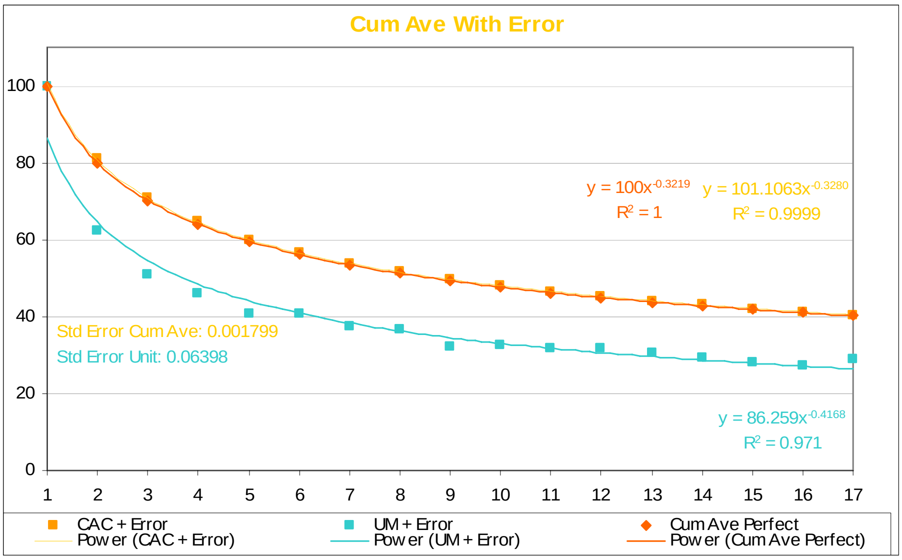

Figure 4.

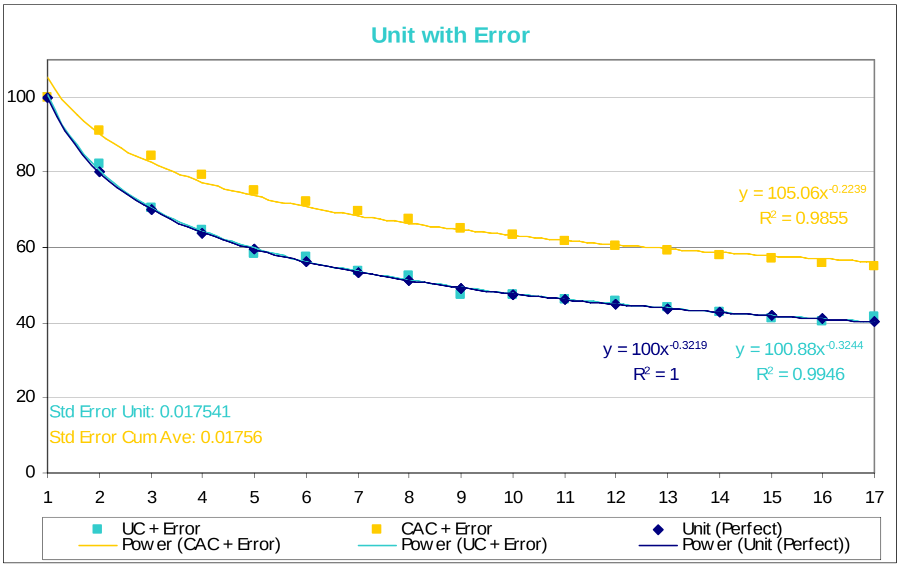

Figure 5.

The error term injected to generate imperfect data was relatively low compared to residual error in "real" data; in order to determine how low the CV would have to be in Unit data in order to test as Unit data (and not CUMAV), the above process was repeated with an error term ranging from a standard deviation of 0.005 times the 8^{th} unit of cost up to a standard deviation of 0.04 times the 8^{th} unit of cost. The average standard errors from each set of trials are plotted below. Figure 6 shows the average standard error for CUMAV data when tested for CUMAV Theory vs. Unit Theory; as expected, CUMAV data *never tests* as though it followed Unit theory.

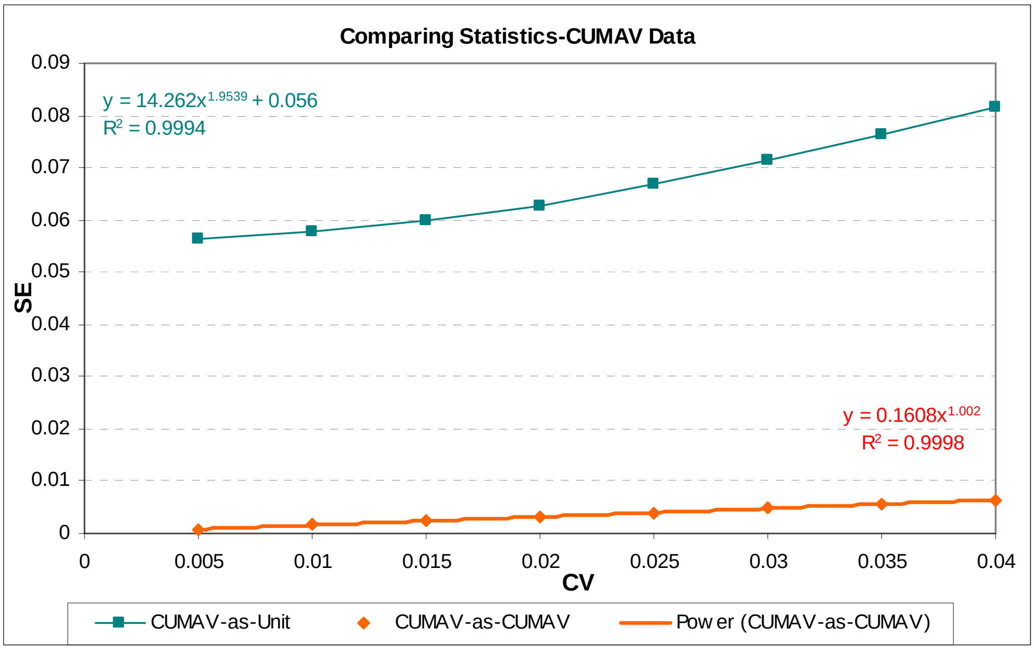

Figure 6.

On the other hand, Figure 6 shows the average standard error for Unit data when tested for Unit Theory vs. CUMAV Theory; here, when the CV in the underlying data is greater than 1.7858% on the eighth of seventeen units, Unit data begins to test *on average* as though it were CUMAV. It is hypothesized that the precise value of this "cross-over point" is linked to the underlying LCS (in this case, 80%).

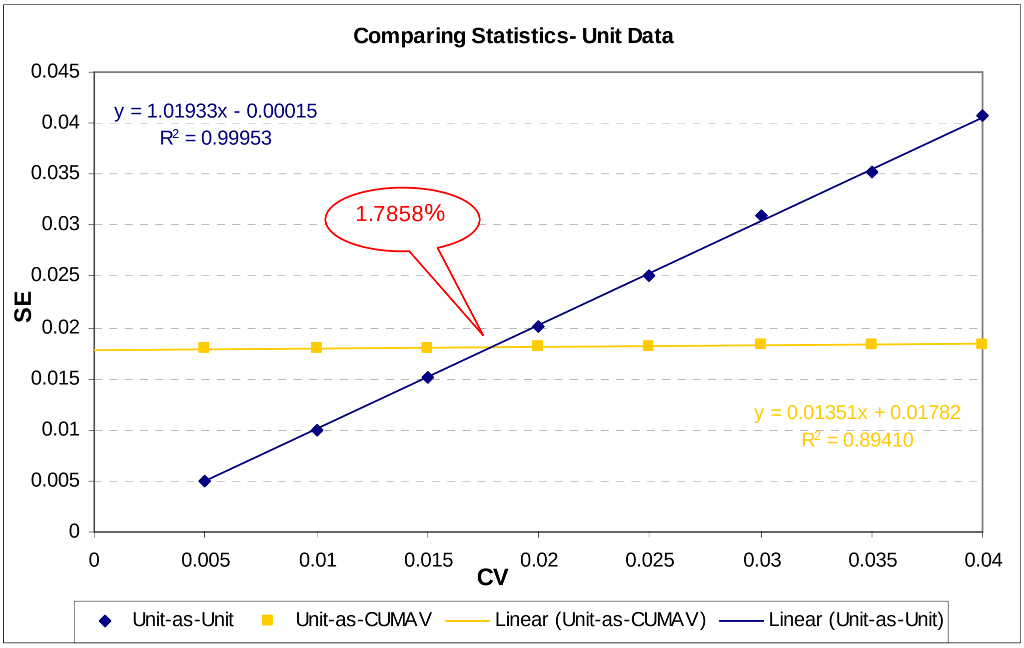

Figure 7.

The above graphs were generated using the average of the Standard Error from each set of data; this smoothes the results. The trend, however, is clear; while the Standard Error for Unit increases one-for-one as CV increases, the increase in Standard Error for CUMAV on the same data is smaller by a factor of about one hundred.

Finally, the percent of "false positives" when fitting CUMAV to Unit data for each set of trials graphed against the CV clearly shows the probability that an analyst will be mislead by statistics as the CV increases. Some additional data points were generated in order to flesh out the curve in Figure 8. From the below, when the CV of the underlying data is ~0.018, the probability that an analyst will be mislead by the statistics is 50%; if the CV is ~0.03, that probability increases to 98%.

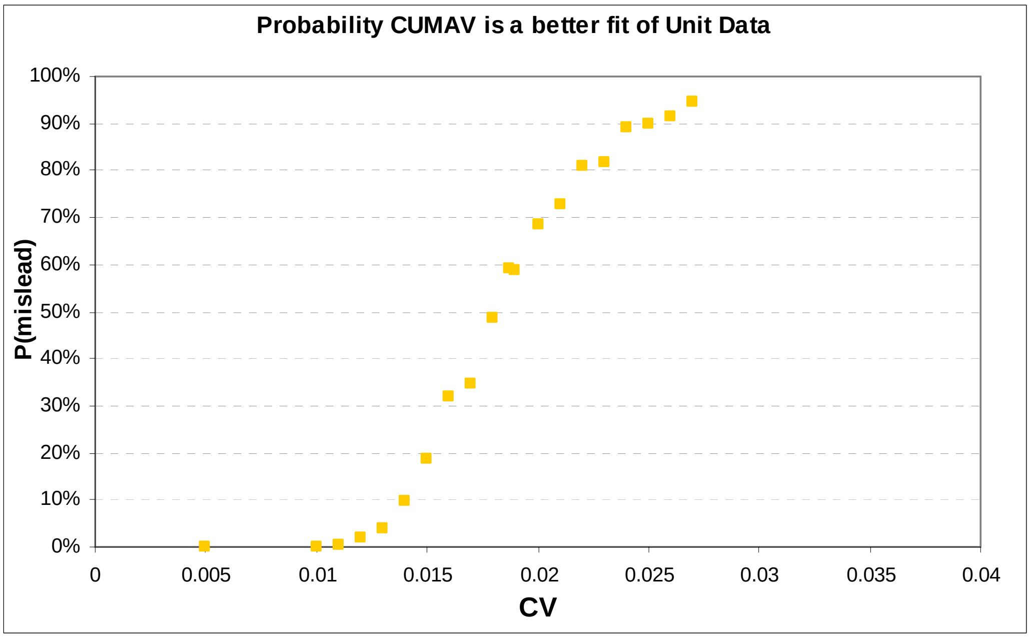

Figure 8.

Conclusions

The analysis in this paper is clearly summarized in Figures 6, 7 and 8. These graphs demonstrate that, using statistics, an analyst should never accidentally select Unit Theory when the underlying data follow a Cumulative Average curve; on the other hand, when the underlying data have a CV of 1.786% or greater, applying CUMAV Theory to Unit data returns, on average, a lower standard error than does Unit Theory. This is problematic, particularly since there is no better way to determine which theory the data truly follow.

Recognizing that the form of the error term used affected the specifics of this conclusion, no claim is made that this answer is precise. Nevertheless, the general finding is irrefutable: *the form of the CUMAV model implicitly favors the selection of this model and it does so for data of remarkably little error.* The rhetorical question put forward in the title of this paper is thus answered: the fight between the two models is decidedly not fair. The following are conclusions based upon the results of this analysis:

1. If indeed both models are valid, there must exist many cases of mistaken identity where the CUMAV model appeared to be superior on the basis of goodness of fit where the Unit model was in fact the underlying mechanism. Accordingly, in any learning curve analysis, this subject must be given careful consideration.
2. There exists a possibility that the CUMAV model may be wholly invalid. Metrics favor the CUMAV model so strongly; the model could easily appear to exist when the data follows a completely different trend. The authors were biased with this opinion at the outset of this investigation. The reason for this initial feeling was purely aesthetic: it was based on the notion that the model, purporting that each subsequent production *unit* exerts some effect only through the *cumulative average*, seems to defy logic. One would expect, in other words, that learning would be asymptotic, as both models are, and that learning would follow from the number of units produced by a constant work force. The difficulty of this paradigm affected the ability to conceptualize the base model as well the ability to conceptualize the error term for the CUMAV model. It may also be that the CUMAV model has been chosen in the past because it is much easier and more convenient to use with lot data; now that sophisticated lot midpoint (LMP) techniques and high-powered computing are readily available, choice of model should be based on merit, not convenience. Finally, the CUMAV model runs afoul of the tests proposed by both Occam and Dirac, who can be paraphrased as saying that the simpler model is preferable and that the most beautiful model is more likely to be true, respectively.

As is always true when assessing learning curves, an analyst must be sure to normalize data before arriving at any conclusions regarding learning curve (theory *or* slope). Failure to normalize will leave an analyst with a pseudo learning curve. It should be clear by this point that the errors introduced by failure to normalize will favor the CUMAV theory: assuming CUMAV will, almost invariably, increase the coefficient of variation; in turn, the goodness of fit metric will favor CUMAV, regardless of the true underlying model.

In summary, analysts should entertain significantly more skepticism about the CUMAV Theory as a learning curve model than they have heretofore. The results of this analysis should be carefully considered in the selection of a model form for both backward-looking analyses of actuals and forward-looking projections. Further, the estimating community would benefit from more penetrating investigations into the two models; specifically studies focused on ascertaining whether both models exist in practice.

References:

[1] http://www.acq.osd.mil/dpap/contractpricing/vol2chap7.htm.
[2] *Ibid.*
[3] "An Enterprise Model of Rising Ship Costs: Loss of Learning Due to Time between Ships and Labor Force Instability," R. L. Coleman, J. R. Towers (neé Summerville), B. L. Cullis, E. R. Druker, G. B. Rutledge, P. J. Braxton, ISPA/SCEA 2007, 4th Annual Acquisition Research Symposium, 2007.
[4] The study contains proprietary data and thus cannot be cited.