An Analytical Framework for Cost Estimating Relationship Development

Christian Smart, Ph.D.
Senior Parametric Cost Analyst
Science Applications International Corporation
christian.b.smart@saic.com

George Culver
Senior Operations Research Analyst
Science Applications International Corporation
george.a.culver@saic.com

Abstract

Cost estimating relationship (CER) development, particularly for spacecraft hardware, has traditionally been method-centric. Much attention has been paid to parameter calculation and goodness-of-fit statistics for particular methods. However, parametric modeling involves other critical factors that are often overlooked, including model validation. These factors are discussed, and a complete analytical framework for developing CERs is presented. Three popular methods for spacecraft hardware CER development are compared in the context of this framework, namely Minimum Unbiased Percent Error (MUPE), Minimum Percent Error with Zero Percent Bias Constraint (MPE-ZPB), and log-transformed Ordinary Least Squares (LOLS). Optimality considerations, goodness-of-fit statistics, and bias, which have been thoroughly studied for each method, are summarized. Model validation and selection involves analysis to ensure that a model is consistent with the underlying data. Issues are raised regarding the conditional variance of the estimates produced by each method when applied to NASA data. In particular, NASA cost data are not symmetrically distributed, and the applicability of each technique for modeling skewed data is discussed. The relationships between all the methods and the method of maximum likelihood are pointed out, as this is important for understanding the validity of these techniques for skewed data. Robustness and consistency of each method for spacecraft data are addressed, and the importance of developing robust CERs is highlighted. Parametric modeling techniques are not unique to cost analysis of spacecraft and weapons systems, but are widely used in other industries as well. Comparisons of the three methods with current, modern techniques for parametrically modeling skewed data in other industries is presented, including insurance, health care, and labor economics.

Introduction

A great deal of thought and effort has been put into techniques for developing and applying cost-estimating relationships (CERs). These efforts, including those by Dr. Steve Book, Shu-Ping Hu, Don Mackenzie, Dr. Matt Goldberg, Pierre Foussier, and many others, have helped to advance the field of CER development in particular and cost estimating in general (Refs.1-8). However, many of the methods and techniques have been developed in isolation, or with a particular problem in mind. What has been lacking is an overall model development framework. In this paper, we present one such framework. Figure 1 displays the proposed model development framework. Other

frameworks maybe acceptable but this serves as one logical structure that ensures all applicable modeling issues are addressed. We have adapted this from a similar framework used in the insurance industry in modeling losses, a topic strikingly similar to cost analysis, since a loss can be considered as a cost from an insurance company's point of view.

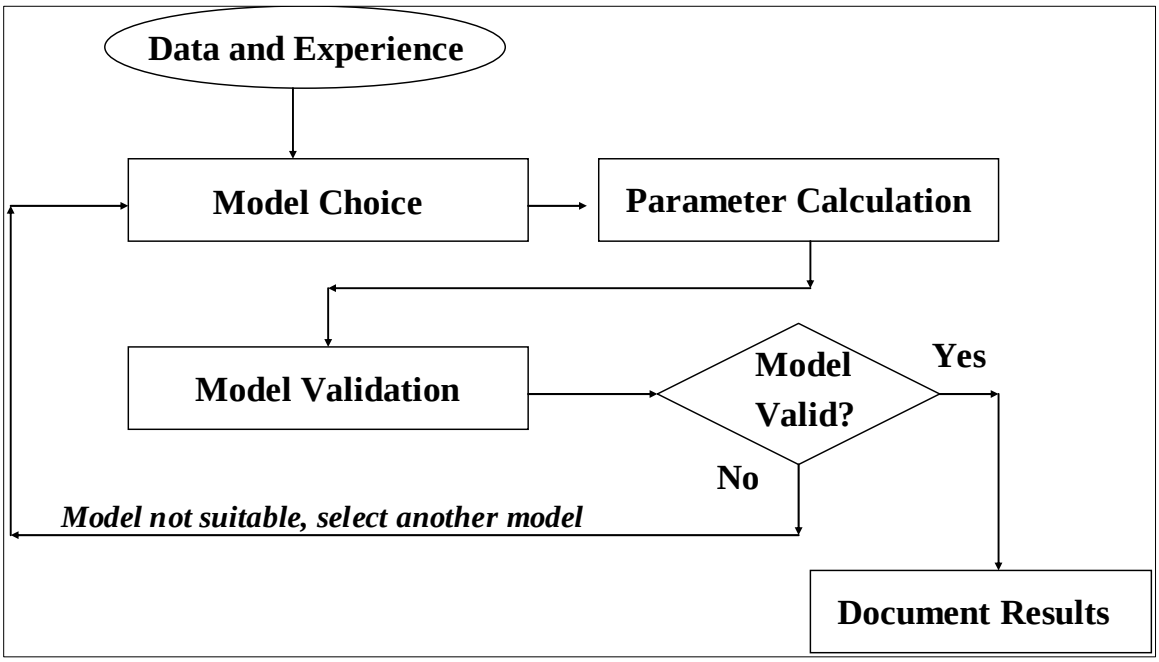

Figure 1. Model Development Framework (Adapted from *Loss Models*, Ref. 9).

In Figure 1, model development begins with the analyst's experience and with the data the analyst has on hand. Based on these two inputs, the analyst chooses a model. We limit our focus in this paper to cost-estimating relationships that follow power equations and which have multiplicative residuals. These are standard assumptions commonly used in CERs developed for spacecraft, boosters, launch vehicles, and other types of NASA and defense projects. Once a model has been chosen the analyst must choose a method for calculating the model's parameters. The parameter calculation method may be dictated by the model choice, or there may be many methods available. There are many ways to calculate the parameters of a model, from statistical methods, to neural networks, as well as bottoms-up types of approaches. There are numerous CER methods, but we focus our attention on three of the most commonly used: log-transformed OLS, Minimum Unbiased Percent Error (MUPE), and the Minimum Percent Error with Zero Percent Bias constraint (MPE-ZPB or ZMPE). We examine each by choosing maximum likelihood estimation (MLE) as our method for parameter calculation, and show the relationship of each of these three methods to this technique.

Once an analyst has chosen a model and calculated the parameters for a CER that is usually the end of the model development process for most analysts. The CER is then put into practice with little thought about how well the CER fits the data or if the model

hypotheses have been validated. This latter step, of prime importance, has often been neglected in the discussion of the merits of particular CER techniques. Model validity is an important part of the model development process. Models that violate the underlying assumptions cannot be relied upon to develop accurate estimates. Completing this final step helps to ensure that the estimates will enable us to make the best projections of future costs, given the limitations of the available data.

Data and Experience

Our goal is to use historical data to predict the cost of future programs and projects. It is important when developing models to limit our choices, since given enough models to choose from, there will be at least one model that appears to fit the data well, but will not help us effectively predict future cost. For example given ***n*** data points, we can perfectly predict the past by fitting ***n-1*** parameters. However, doing so will capture many idiosyncrasies in the historical data that are not likely to be repeated in the future, a phenomenon referred to as over fitting a model. Experience is a useful guide in limiting the universe of choices. Models are not typically developed in a vacuum, but rather as part of an evolving process. The seasoned analyst who has successfully used techniques that have worked well in the past will typically apply these same techniques again. The challenge for those of us who have successfully used models in the past is to not become so attached to one particular model or technique that we become over-reliant upon it and try to apply it to situations where it does not apply. In such cases we become like the carpenter who only has a hammer, and who applies this hammer to all his work, whether it involves driving nails or inserting screws. We should be flexible in considering new techniques, and rigorous in making sure that we use the right experience, the right data, and models that are valid for the problem at hand.

Model Choice

Model choice is the next step in this process. There are many models to choose from. The analyst has many decisions to make in selecting a model, such as whether to use an explicit equation form, or, for example, a decision tree. An example of a decision tree might be something like “if the total dry weight of a spacecraft is more than 1,000 lbs., my estimate is $100 million, otherwise it is $25 million.” Or the analyst might select a model that estimates the costs based on user inputs, such as the NASA/Air Force Cost Model (NAFCOM). Using a mathematical equation to estimate the cost using one or more predictive cost drivers, such as weight, is a traditional approach. Also, if one selects to use an equation to estimate cost, the analyst must also decide what type of equation to use, whether linear or nonlinear.

Through data and experience, it is largely agreed in the NASA and DoD cost community that costs do not typically follow a linear pattern. Rather they tend to vary nonlinearly in relation to the cost drivers that are typically selected. In particular, the power law equation, or some variation of it has been widely adopted. The power law equation has the form

$$Y = aX^b.$$

In this case ***Y*** typically represents cost in $, but can also represent effort (hours, full-time equivalents). ***X*** typically represents weight or some other performance parameter. The

equation can also be modified to accommodate multiple cost drivers. The value of the ***b*** parameter in the power equation is usually less than ***1***, indicating economies of scale in design and production. This model has been found to do a good job explaining the relationship between cost and cost drivers, including weight and other performance characteristics, for a wide variety of spacecraft and other programs. For example, if the equation has the form,

$$\textit{Estimated Cost} = 1.5 \cdot \textit{Weight}^{0.5}$$

then as weight doubles, cost is increased by a factor equal to the square root of weight, rather than a simple linear relationship. Such equations are so commonly used they are referred to as "cost estimating relationships," or CERs.

If an equation form is selected, the form of the residuals between predicted and actual costs must also be chosen. For example given an equation of the form

$$Y = a + bX$$

and a set of data

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

the residuals are defined as:

$$\varepsilon_i = Y_i - (a + bX_i) = \textit{Actual} - \textit{Estimated}.$$

This is also referred to as the "error" term since it is the difference between the actual cost and the estimated cost. Residuals or "errors" are an important consideration in modeling since they often drive the methods used for parameter calculation. For example, linear regression finds the "best fit" by finding the parameters ***a*** and ***b*** that minimize the sum of the squares of the residuals

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} \left(Y_i - (a + bX_i)\right)^2 = \sum_{i=1}^{n} \left(\textit{Actual}_i - \textit{Estimated}_i\right)^2 \quad .$$

This method was first developed by the mathematicians Legendre and Gauss in the early 19th century, who used it to predict the orbits of heavenly bodies using observed data. Francis Galton later applied this technique to find linear predictive relationships between various phenomena, such as the relationship between the heights of fathers and sons. Galton found a positive correlation between these heights but found a tendency to return or "regress" toward the average height, hence the term "regression analysis"

The residuals of the power equation can either be additive or multiplicative.

Additive residuals have the form

$$Y = aX^b + \varepsilon$$

while multiplicative residuals have the form

$$Y = aX^b \varepsilon.$$

Multiplicative residuals are more appropriate for the spacecraft and defense industry in most applications because of wide variations in size, scope, and scale of the systems that are estimated. For example, if historical data ranges from $50 million to $1 billion, it is better to analyze percentage differences, since this provides a more meaningful comparison of accuracy than absolute dollar values. As a result we are primarily interested in the percentage difference between actual and estimated costs, not the absolute difference. See Figure 3 for a graphical comparison of multiplicative and additive errors. The commonly-used regression techniques considered in this paper are all based on the multiplicative error assumption.

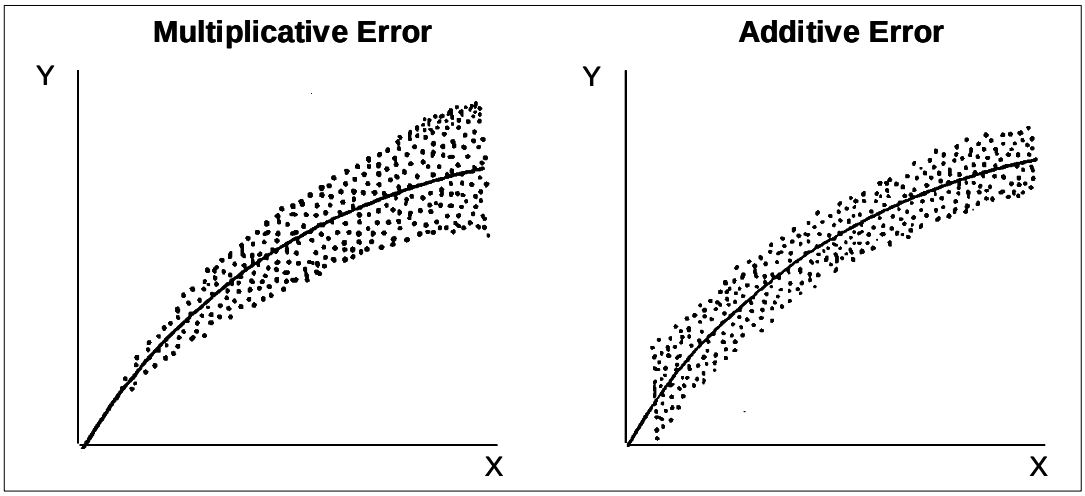

Figure 2. Multiplicative Vs. Additive Errors (Ref. 10).

For the power equation with multiplicative residuals, i.e.,

$$Y = aX^b\varepsilon$$

the estimates vary based on the variation of the residual

$$\varepsilon = \frac{Y}{aX^b}.$$

It's also common to adjust this to treat ε as a percentage, i.e., set

$$Y = aX^b(1+\varepsilon)$$

$$\varepsilon = \frac{aX^b - Y}{aX^b} = \frac{Estimate - Actual}{Estimate}.$$

That is, actual cost is equal to the estimate plus or minus a percentage of the estimate. Note that if the estimate is greater than the actual cost the residual is greater than zero. If the estimate is less than the actual the residual is less than zero. Note the lack of symmetry. For estimates above the actual, the maximum value of the residual is ***1***, and for estimates below the actual, the minimum value has no bound!

See Figure 3 for an illustration of multiplicative residuals for a subsystem CER in the NASA/Air Force Cost Model.

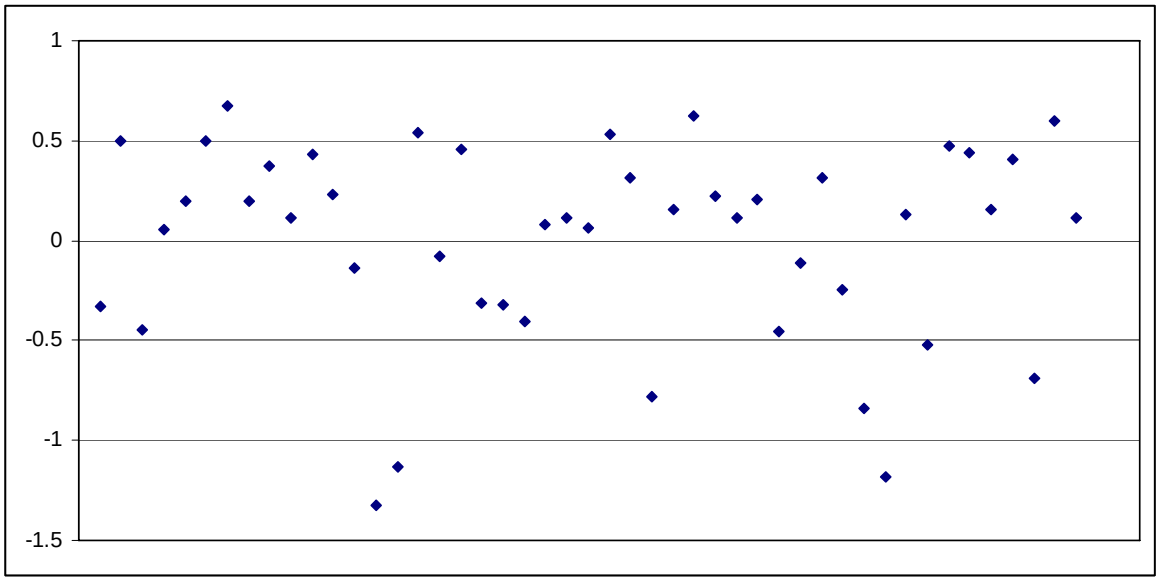

Figure 3. Multiplicative Residuals for a Subsystem CER in the NASA/Air Force Cost Model (NAFCOM)

For a "good" model, the cost drivers explain all (or most) of the variation in the historical data that can be explained. Therefore it is typically assumed that any remaining variation is random, either due to non-repeatable random phenomena (e.g., test failures) that are truly random phenomena and will not help predict future cost. The multiplicative residuals that represent this unexplained variation are thus treated as random variables. Thus after choosing the equation, and choosing the type of residual, we have to make yet a third choice, namely, the type of distribution that this unexplained variation follows. For CER development, residuals are typically assumed to follow normal, lognormal, gamma, or they are treated without making such an assumption (non-parametric).

The normal distribution is the most common probability distribution. Many random phenomena follow this distribution. It is also known as the "bell curve," due to its symmetry and small tails. If cost is a sum of many random independent phenomena, the central limit theorem indicates this may be the appropriate distribution. See Figure 4 for a depiction of a normal distribution.

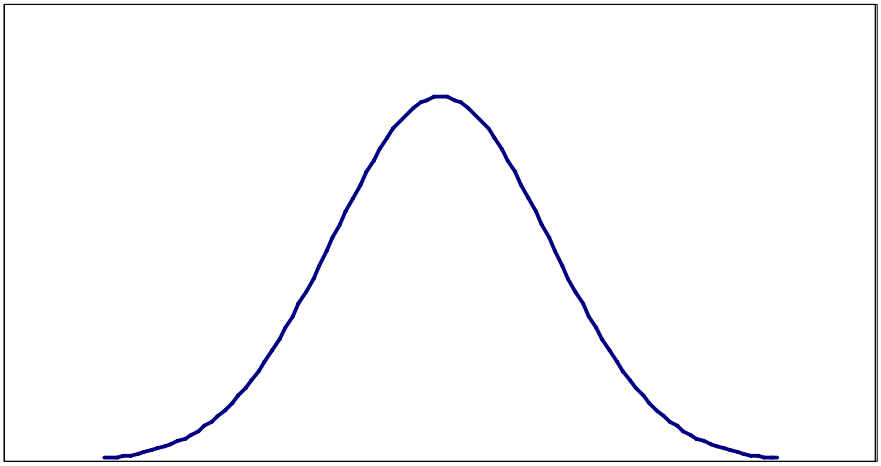

Figure 4. Example of a Normal Distribution.

The lognormal distribution is a skewed distribution. If ***X*** is lognormally distributed, ***y = ln(x)*** is normally distributed. The lognormal has a heavier right tail than the normal distribution, is bounded below by zero, and unbounded above. If cost is a function of multiplicative factors, for example, test failures cause a percentage increase in cost rather than a fixed amount increase, project costs are likely to be lognormally distributed. This is a multiplicative analog to the central limit theorem. These aspects make the lognormal appealing for cost modeling. See Figure 5 for a graphical depiction of a lognormal distribution.

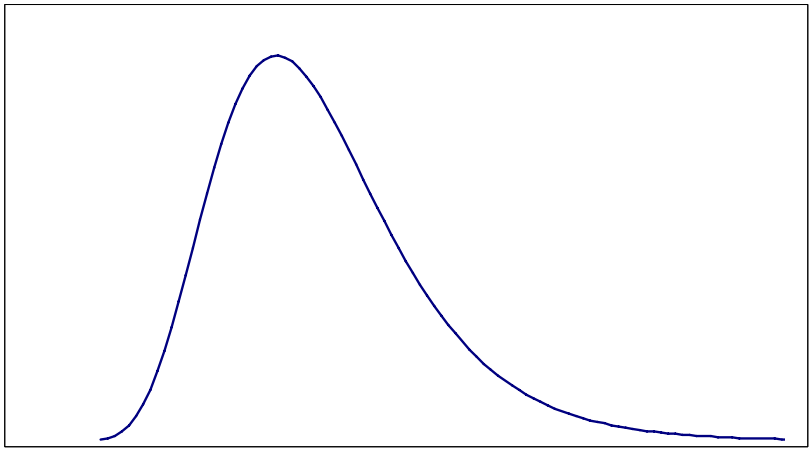

Figure 5. Example of a Lognormal Distribution.

The gamma distribution is a flexible distribution. It can to some extent resemble a lognormal, and can also resemble an exponential distribution. Indeed the gamma distribution is the sum of independent exponential distributions, so the exponential is a special case of a gamma distribution. See Figure 6 for representations of gamma distributions.

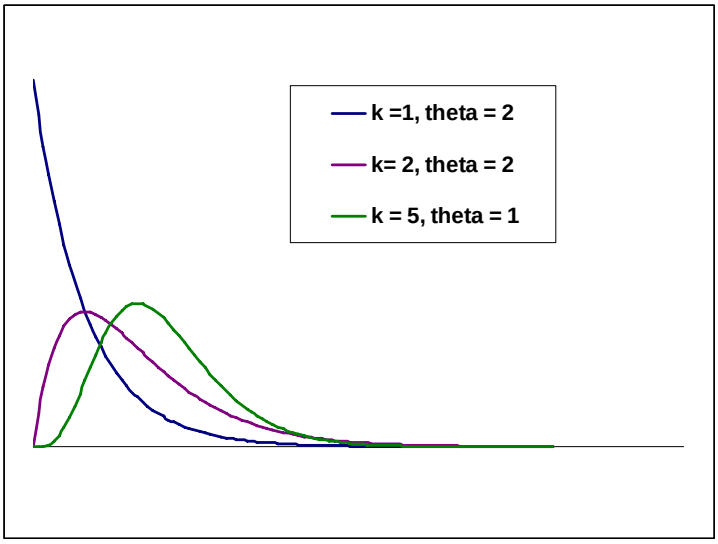

Figure 6. Examples of Gamma Distributions.

The decision to choose or not choose an underlying distribution for the residuals represents a fourth choice. When the data follow an observable pattern, based either on preliminary data analysis or through experience, parametric analysis is preferred. For example, assume that residuals follow a lognormal, gamma, or normal distribution. For example, NASA cost data are skewed, which makes intuitive sense, because cost cannot be less than zero, but there is no upper limit. This often leads to the assumption of lognormal or gamma residuals. However when the data do not follow an observable pattern, or there is no reason to assume an underlying pattern in the data, non-parametric analysis may be suitable. This may be the case if data sets are small, or if there is no reason to assume similarity with other available data.

However, if non-parametric techniques are used, the analyst must be careful to ensure models are valid, since techniques may assume an inherent pattern in the data or be similar enough to a parametric technique that the non-parametric version inherits some features of the parametric version.

Another issue with non-parametric techniques is the lack of rich techniques for developing confidence intervals, prediction intervals, covariance matrices, and other useful metrics and methods available for parametric models. Indeed, some statistical techniques do not exist for nonparametric problems. As shown by Bahadur and Savage in their 1956 paper “The Nonexistence of Certain Statistical Procedures in Nonparametric Problems” (Ref. 11), in such cases there is no effective hypothesis test for the population mean, no effective confidence interval for the population mean, and no effective point estimate for the population mean. They also showed that no confidence interval will fit the data well. This makes model validation problematic for non-parametric methods. However, note that parametric techniques do not necessarily involve assuming the residuals follow a particular probability distribution. This assumption can be much weaker, such as assuming finite variance.

Parameter Calculation

There are numerous ways to calculate the parameters of a cost-estimating relationship. One powerful statistical technique commonly used for parameter calculation is the method of maximum likelihood, which is sometimes referred to as maximum likelihood estimation (MLE). MLE is a widely used technique that serves as a unifying framework for the three CER methods we shall discuss in this paper.

Let $a_1,\ldots, a_n$ represent the observed data and $x_1,\ldots,x_n$ represent random variables where a_i results from observing the random variable x_i. The likelihood function, which represents the likelihood of obtaining the sample results, is defined as

$$L(\theta) = \prod_{i=1}^{n} Pr(X_i = A_i \mid \theta)$$

The maximum likelihood estimate of θ is the vector that maximizes the likelihood function. This technique is appealing because maximizing the likelihood of finding the true underlying parameters of this distribution is exactly what we hope to accomplish in developing our CER. One major advantage of this technique is that the likelihood function is almost always available. Maximum likelihood methods have good statistical properties, like consistency and efficiency, which will be revisited later. Also a rich body of statistical theory has been developed for maximum likelihood estimation.

The three CER methods considered in this paper, namely, log-transformed OLS, MUPE, and MPE-ZPB, all have a connection to maximum likelihood estimation, in the sense that parameter calculation for each of the methods considered can be viewed in the context of maximum likelihood. Maximum likelihood is used together with an assumption about the underlying residuals to calculate the parameters of the error distribution. Each of the CER techniques we consider has a strong connection to maximum likelihood estimation paired with either the lognormal, normal, or gamma distribution.

For $y_i=f(x_i,b)\cdot u_i$, let b = vector of coefficients of the CER, y_i = actual cost of the i^{th} data point, x_i = vector of cost drivers for the i^{th} data point, and u_i = residual of the i^{th} data point. The likelihood for a lognormal distribution for the i^{th} data point is

$$L(\mu,\theta) = \frac{1}{u_i\sqrt{2\pi\theta}} e^{-\frac{(\ln u_i-\mu)^2}{2\theta}}.$$

If we set $\mu=0$, we are estimating the median of the distribution. Why would we want to estimate the median? The three most commonly encountered measures of centrality are the mean, median, and mode. The mean is the "expected value," so for a sample of n data points this is

$$\sum_{i=1}^{n} \frac{x_i}{n}.$$

The median is the 50th percentile, the point at which half the population is less than this value, and half is greater. The mode is the "most likely" point of the density function, that is, the peak of the distribution. For a normal distribution, the mean, median, and mode are

all equal. For a lognormal, the mode is always less than the median, and the median is always less than the mean. Thus for a lognormal, the mean is always greater than the 50th percentile, and can be any percentile greater than the 50th, such as the 90th or 95th percentile. For this reason, a better metric for the center of a lognormal is the median. That is why it is common to report the median rather than the mean as the "average" of skewed data. Whenever average income or average house price data are reported in the media for example, the average reported is always the median, and for exactly this reason. See Figure 7 for a graphical comparison of the mode, median, and mean of a lognormal distribution.

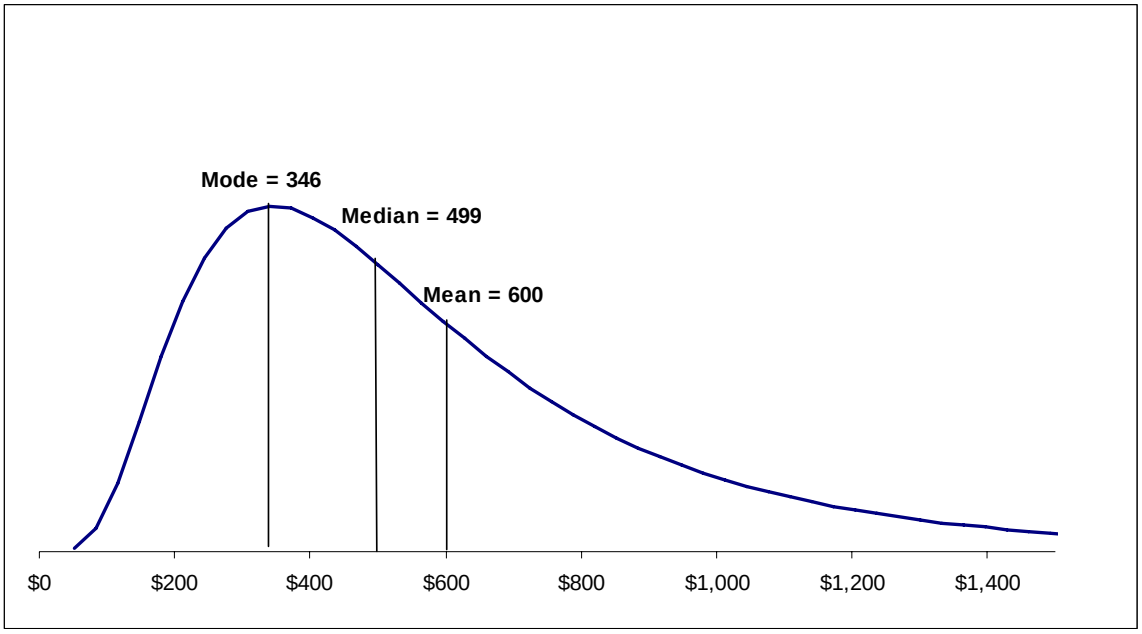

Figure 7. Comparing the Mode, Median, and Mean for a Lognormal Distribution.

Thus in order to estimate the median of a lognormal distribution, we start with

$$Pr(U \le u) = \int_{-\infty}^{u} \frac{1}{U\sqrt{2\pi\theta}} e^{-\frac{(\ln u - 0)^2}{2\theta}} dU.$$

However, we want to analyze this in terms of $y=f(x,b)\cdot u$. **Since** $u=y/f(x,b)$**,**

$$Pr(Y \le y) = Pr(u \cdot f \le y) = Pr\left(u \le \frac{y}{f}\right)$$

$$= \int_{-\infty}^{y} \frac{1}{y/f\sqrt{2\pi\theta}} \frac{1}{f} e^{-\frac{(\ln(y/f))^2}{2\theta}} dy$$

$$= \int_{-\infty}^{y} \frac{1}{y\sqrt{2\pi\theta}} e^{-\frac{(\ln(y/f))^2}{2\theta}} dy.$$

Therefore, the likelihood function is

$$L(\beta,\theta) = \prod_{i=1}^{n} \left(\frac{1}{y_i (2\pi\theta)^{\frac{n}{2}}} \right) \cdot e^{-\frac{1}{2\theta}\sum_{i=1}^{n} \ln\left(\frac{y_i}{f_i}\right)^2}.$$

Since the logarithm function is monotonically increasing, we can take the log of the likelihood function and maximize that instead, that is,

$$l(\beta,\theta) = -\frac{n}{2}\ln\theta - \sum_{i=1}^{n} \ln y_i - \frac{1}{2\theta}\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2.$$

Note that we ignore constants since they do not affect the maximization. Also note that this is the same as minimizing the negative of the log-likelihood function, that is,

$$l(\beta,\theta) = \frac{n}{2}\ln\theta + \sum_{i=1}^{n} \ln y_i + \frac{1}{2\theta}\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2.$$

In order to minimize the likelihood function, we first minimize with respect to θ.
To minimize, we take the partial derivative with respect to θ, set equal to zero, and solve for θ. Taking the derivative yields

$$\frac{\partial l}{\partial \theta} = \frac{n}{2\theta} - \frac{1}{2\theta^2}\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2.$$

Setting this equal to zero and solving gives

$$\hat{\theta} = \frac{\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2}{n}.$$

Plugging in the value for θ into the log likelihood function yields

$$l^*(\beta) = \frac{n}{2}\ln\frac{\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2}{n} + \sum_{i=1}^{n} \ln y_i + \frac{n}{2}.$$

Ignoring constants this simplifies to

$$l^*(\beta) = \ln\sum_{i=1}^{n} (\ln y_i - \ln f(x_i,\beta))^2,$$

which is equivalent to minimizing

$$L^*(\beta) = \sum_{i=1}^{n} (\ln y_i - \ln f(x_i, \beta))^2 .$$

This is the least squares of the log of the differences between the actual and the estimated costs. Notice the similarity to linear regression. What we have derived is a generalization of log-transformed ordinary least squares in the context of maximum likelihood. In log-transformed ordinary least squares, logarithmic transformation is applied to both the actual and the estimated costs. For the power equation $Y=aX^b$ this transforms the equation from a nonlinear equation to a linear one, i.e.

$$\ln Y = \ln(aX^b) = \ln a + b \ln X .$$

The parameters can be easily calculated in a spreadsheet. However, note that the maximum likelihood median estimator is more general. In the past this method has merely been viewed as a simplistic way of converting a nonlinear power equation to linear space and applying ordinary least squares to the resulting linear equation (Ref. 3). We have proven instead that any equation form may be used as there is nothing in the derivation that forces a particular equation type to be used. One simply minimizes the sum of the log-squared differences between the actual and the estimated costs. Thus equation forms such as $y=a+bx^c$ can be calculated with this generalized method, which we term Generalized Lognormal Maximum Likelihood Estimation, or GLMLE ("Glimly"). However, note that unless the log transformed equation is linear, one may need a computer to solve.

Although for skewed data, the median is a better representative of a distribution's centrality, the mean is the focus of most statistical estimators. The other two methods we will present estimate the mean, rather than the median of the error distribution, and one criticism often levied on log-transformed OLS is that it is biased low, since the median of a lognormal distribution is always less than its mean. However, this can be corrected, since there is a mathematical relationship between the median and the mean. The mean of a lognormal distribution is $\exp(\mu + \sigma^2/2)$ and the median is simply $\exp(\mu)$, so the mean is the quantity $\exp(\sigma^2/2)$ multiplied the estimate. The only complicating factor is that the population variance is not known with certainty and so it must be estimated using statistical samples. Several methods for estimating this factor have been proposed. A simple one, termed the "Ping" factor (Ref. 8) is

$$\exp\left(\left(1 - \frac{p}{n}\right)\frac{s^2}{2}\right)$$

where p is the number of parameters, n is the number of data points in the sample, and s^2 is the sample variance.

For the equation $y_i=f(x_i,b)\cdot u_i$, when the residuals are normally distributed, with mean 1 and variance θ, the likelihood function, as demonstrated by Lee (Ref. 12), is

$$L(\beta,\theta)=\frac{exp\left(\frac{-1}{2\theta}\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2\right)}{(2\pi\theta)^{\frac{n}{2}}\prod_{i=1}^{n}f(x_i,\beta)}.$$

The log-likelihood function is thus

$$l(\beta,\theta)=\frac{-1}{2\theta}\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2-\frac{n}{2}ln(2\pi)-\frac{n}{2}ln\,\theta-\sum_{i=1}^{n}ln\,f(x_i,\beta).$$

Maximizing this expression for θ and then substituting back into $l(\beta,\theta)$ yields the concentrated log-likelihood function

$$l^*(\beta)=-\frac{n}{2}ln\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2-\sum_{i=1}^{n}ln\,f(x_i,\beta).$$

This is the same as minimizing

$$l^*(\beta)=\frac{n}{2}ln\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2+\sum_{i=1}^{n}ln\,f(x_i,\beta).$$

As noted in Goldberg and Tuow (Ref. 13), this method is very similar to the minimum percent error method developed by Book and Young (Refs. 1 and 2), who ignore the final term and instead minimize the sum of squared percentage errors.

The minimum percent error method minimizes

$$\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)^2.$$

Thus the minimum percent error method is a pseudo-likelihood estimator in the case of normally distributed residuals since it is equivalent to minimizing the first term in the concentrated likelihood function. Note that the minimum percent error (MPE) method is biased. Instead of being biased below the mean like with log-transformed OLS, the MPE method is biased high, since one way to make the error term small is to make the estimates large. To correct for this Book and Lao (Ref. 3) introduced a bias constraint. The objective function is the same, but now sample bias is constrained to be zero, that is

$$\sum_{i=1}^{n}\left(\frac{y_i-f(x_i,\beta)}{f(x_i,\beta)}\right)=0.$$

This method is referred to as MPE-ZPB or ZMPE ("Zimpy"). It's not explicitly a parametric method, but it is similar to the normal MLE. In fact, it's so similar that MPE-ZPB provides a good approximation of the normal MLE, and in many cases will provide the same solution. To see that this is the case, note that the normal MLE objective function has two terms. The dominant term in this optimization is the first term. The second term provides a penalty for overestimating, but the gradient is much more sensitive to departures from the actual as a percentage of the estimate, unless the actuals are close to zero. The first term is unbounded below. Near the actual costs, this term approaches negative infinity. This only occurs for the second term when the estimates approach zero. For the case of one single actual cost equal to 10 see Figure 8 for an example that illustrates the large differences in the log-likelihood function terms in terms of the gradient. The gradient for the log-likelihood function is dominated by the first term, which is the minimum percent error term. Note that the first term and the total objective function are so similar that it is difficult to distinguish them on the graph.

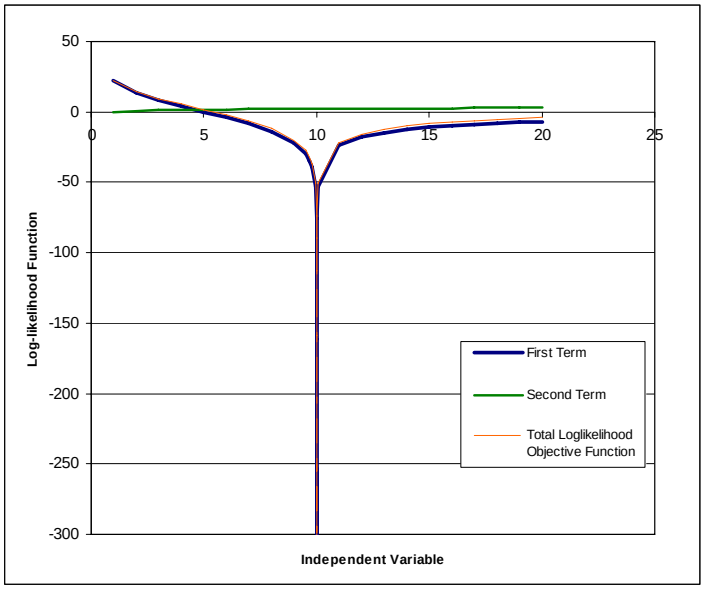

Figure 8. Graph Illustrating the Loglikelihoods for Normal MLE.

The function of the second term, together with weighting the first term by ***n/2*** ensures that the estimate is asymptotically unbiased. This has a similar function to the bias constraint in the MPE-ZPB method. The major difference is that MPE-ZPB is unbiased regardless of sample size, while the normal MLE will only be unbiased for large samples, since maximum likelihood estimates are asymptotically unbiased, but not necessarily unbiased for all samples, particularly small samples.

As a simple example comparing MPE-ZPB with the normal MLE, consider the data in Table 1.

Weight	Cost
2	4
4	6
5	8
10	12
15	15
20	37
30	25
40	22
50	35
55	40

Table 1. Example Cost, Weight Data.

For this example the normal MLE fit is

$$Estimated\ \ Cost = 2.41Wt.^{0.716},$$

while the MPE-ZPB fit is

$$Estimated\ Cost = 2.39Wt.^{0.719}.$$

See Figure 9 for a graphical comparison. A loglinear fit has been added to provide a contrast, since the normal MLE and MPE-ZPB graphs are coincident.

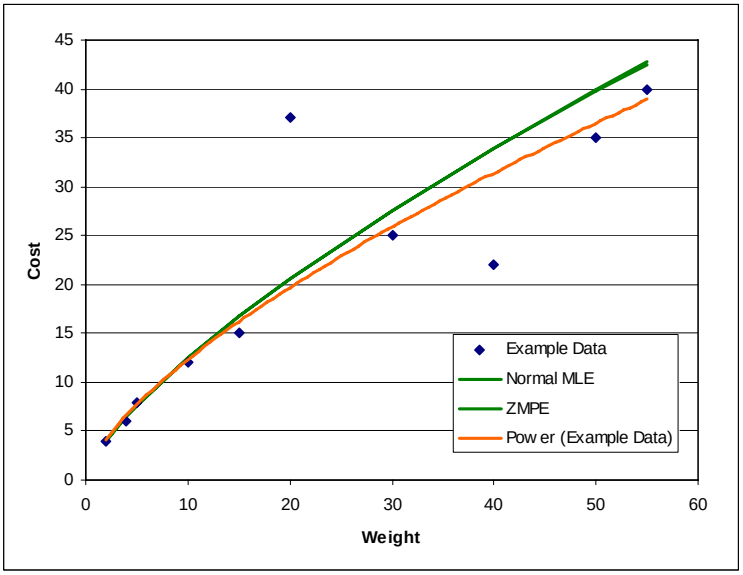

Figure 9. Normal MLE, ZMPE, and Log-Transformed OLS Comparison.

The MPE-ZPB method is tied to the assumptions of the normal MLE. Therefore we need normally distributed (multiplicative) residuals to ensure consistent solutions. Also, the normal MLE is very sensitive to departures from normality assumption. This will be treated in more detail when we come to model validation.

When the residuals follow a gamma distribution, the negative log-likelihood function is

$$l(\beta) = \sum_{i=1}^{n}\left(\frac{y}{f(x_i,\beta)} + ln\ f(x_i,\beta)\right).$$

This can be minimized by iteratively minimizing the sum of percent squared errors until the estimates converge, that is

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta_k)}{f(x_i,\beta_{k-1})}\right)^2,$$

where k denotes the iteration number. This method was first developed by Nelder (Ref. 14) and Wedderburn (Ref. 15), who called the method iteratively re-weighted least squares (IRLS). It was re-discovered by Hu in the 1990s, who called it minimum unbiased percentage error (MUPE) (Refs. 4, 8).

In the case of gamma residuals, IRLS/MUPE is a maximum likelihood estimate. However IRLS/MUPE does not depend upon the assumption of gamma residuals. The likelihood method was generalized by Wedderburn to consider quasi-likelihood, which has good statistical properties, but only requires a finite variance.

Thus log-transformed OLS, MPE-ZPB, and IRLS/MUPE all share a common connection in maximum likelihood estimation. Log-transformed OLS/GLMLE is a maximum likelihood estimator of the median of lognormally distributed multiplicative residuals. Thus it is a parametric method. MPE-ZPB is a pseudo-likelihood estimator of the mean of normally distributed multiplicative residuals with a bias constraint added. It is not directly parametric but it has parametric properties because it is typically a good approximation of the normal MLE solution. IRLS/MUPE is a maximum likelihood estimator of the mean of gamma distributed residuals. But it is also more general, since it is a quasi-likelihood parametric method.

Model Validation

Parameter calculation is the end of the process for many cost analysts. Once coefficients have been calculated, many analysts begin applying the new equations. But we are not done yet! We still need to check model validity. We need to determine whether or not the estimates do a good job of replicating actual historical cost, and whether or not the underlying assumptions, if any, hold true.

There are numerous goodness-of-fit metrics that can be used to determine whether or not the estimates do a good job of fitting the actual historical cost. Three commonly used measures are Pearson's R^2, standard percent error, and percent bias. Keep in mind that these three CER methods are nonlinear ones. As such, the linear R^2 commonly discussed in textbooks is problematic, since in the nonlinear case, this measure can be negative, as discussed by Book and Young (Ref. 5). As a measure of goodness-of-fit that does not suffer from this issue, Book and Young proposed a measure, Pearson's R^2, that circumvents this issue by calculating the square of the Pearson correlation coefficient. This is similar to the coefficient of determination and has a similar interpretation. Higher values are desired, and the metric ranges from 0 to 1. "Good" CERs often have Pearson's R^2 values above 90%. A second goodness of fit measure that measures deviations away from the fit is standard percent error. The typical linear standard error doesn't work well for the nonlinear multiplicative error case, since we are interested in percentage

deviations from the actual cost. The standard percent error is a nonlinear analog to the regression standard error, and is defined as

$$\%SEE = \sqrt{\frac{1}{n-k}\sum_{i=1}^{n}\left[\frac{y_i - f(x_i)}{f(x_i)}\right]^2} \times \quad 100\%,$$

where n is the sample size, and k is the number of fitted coefficients. In this case, lower values are desired. These values can be quite large, even for CERs with Pearson R^2s above 90%. A standard percent error for spacecraft CERs below 30% is considered excellent (and rare). Note the similarity of this metric to the MPE-ZPB objective function. Thus by design the MPE-ZPB method will have the lowest standard percent errors of any unbiased estimate. This method is simple, straightforward, and analogous to the standard error for linear CERs, but it has some drawbacks. The measure is not symmetric, and over estimates are not penalized the same as under estimates. Under estimates are heavily penalized compared to over estimates. For example, if the actual value is equal to 9 an estimate equal to 27 has squared percentage error equal to

$$\left(\frac{9}{27} - 1\right)^2 \cong 0.44 .$$

On the other hand, if the estimate is only one-third the amount of the actual value, or 3, the squared percentage error is

$$\left(\frac{9}{3} - 1\right)^2 = 4 .$$

Thus the minimum percent error method penalizes underestimation much more heavily than overestimation. Note the upper bound of the squared error for overestimation is 1, while the squared error for underestimation is not bounded above at all! Foussier has also noted that this metric distorts the true underlying error because of the squaring of the percent residual, and has proposed absolute value as a better measure than the squared error (Ref. 16).

Percentage bias is another important metric. It is defined as

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)$$

where n denotes the sample size. MPE without the bias constraint produces estimates that are biased upwards. As we have discussed log-transformed ordinary least squares produces estimates that are biased low. The MUPE method is asymptotically unbiased but can be biased for small samples. It is desirable to have estimates that have zero bias if you are interested in estimating the mean.

These three are the primary goodness-of-fit measures used for CERs developed in the cost estimating industry but there are numerous others, such as the trimmed percent error, which measures percent error of the estimate based on the average of the absolute value of the percent errors for the middle 80% of the data points. This trimming allows the

analyst to see how well the CER fits the bulk of the data by avoiding excessive influence of extreme data points.

Another important measure or quality of an estimator that is often overlooked is consistency. An estimator is consistent if for all $\delta > 0$ and any θ,

$$\lim_{n\to\infty} Pr\left(\left|\hat{\theta}_n - \theta\right| > \delta\right) = 0.$$

It's important that the technique converges to the true parameter as the sample size increases. Without this we have no guarantee our estimated parameter resembles the true underlying population parameter. Thus consistency is necessary to have a reliable model. Maximum likelihood methods are consistent when the underlying assumptions are valid. Thus log-transformed OLS is consistent when the (multiplicative) residuals are lognormally distributed. Due to its similarity with the normal MLE, MPE-ZPB will typically be consistent only if the residuals are normally distributed. MUPE on the other hand, will be consistent as long as the variance is finite. Thus this method is more robust in the face of departures from assumptions about the underlying shape of the residuals than the other two.

Another often overlooked metric is the mean square error. This is the error about the coefficient, something of primary importance in developing a CER, since we want to be able to accurately measure the true population parameter.

The Mean-Squared Error (MSE) of an estimator is

$$E\left[\left(\hat{\theta} - \theta\right)^2 \mid \theta\right].$$

An estimator $\hat{\theta}$ is a uniformly minimum variance unbiased estimator (UMVUE) if it is unbiased and for any true value of θ there is no other unbiased estimator that has a smaller variance. An estimator that is UMVUE is efficient, in that it achieves its theoretical lower bound. In practice this means that the estimated coefficient will likely be closer to the true coefficient than that calculated with another estimator. A (finite) data set is often considered as a random sample from an underlying population. The variance of the coefficients is a decreasing function of the sample size, so for small samples, the variance can be quite large relative to the coefficient. In these cases, the variance of the coefficient, if large, can mean than the estimated coefficient is far away from the true coefficient. Smaller variances mean a quicker convergence to the true underlying population coefficient as the sample size increases, as long as the estimator is consistent. Consider for example two consistent estimators for a coefficient, and suppose that one has variance equal to 50% of the size of the estimate, and the other has variance equal to 100% of the size of the estimate. Suppose for the sake of simplicity that the estimator's mean is equal to the true population coefficient. The coefficient for a single data set can be viewed as one random sample from a Monte Carlo simulation of the coefficient distribution. A single draw drawn for each of these is likelier to be closer to the true coefficient for the distribution with smaller variance. For example, a single Monte Carlo draw for the distribution with smaller variance is 0.89 while the same random draw for

the distribution with higher variance results in 1.58. The first estimator is thus 11% below the true coefficient, while the second is 58% greater. Thus small variance is a highly desirable property. Since maximum likelihood estimates are UMVUE, it is desirable to use them whenever their use can be justified. In general parametric models can be seen as more accurate predictors whenever the hypotheses required for their use can be supported.

In addition to measures of goodness-of-fit and consistency, and mean square error, we also need to validate model hypotheses. If a maximum likelihood method has been used, we need to check to see if residuals fit the assumed shape. Fit is used here in the negative sense. We can never truly prove anything statistically. We can use data to disprove conjectures but the best we can hope for when we make hypotheses and test them, these tests will fail to disprove or reject our hypotheses. As the great philosopher of science Karl Popper once remarked “Our knowledge can only be finite, while our ignorance must necessarily be infinite.” Three commonly used tests for the goodness of fit of a distribution are chi-square, Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D). Chi-square and Kolmogorov-Smirnov are both simple and easy to compute. Anderson-Darling is more powerful and considered a good test for departure from normality. Anderson-Darling gives more weight to the tails of the distribution, while chi-square gives more weight to low probability intervals.

In terms of what works in practice, given that our error is bounded below since neither our estimate nor our actual can be less than zero, but there is an infinite amount of room on the upside, we should expect positive skew in our error distributions. The lognormal distribution is a natural choice for modeling positive skew. Don Mackenzie (Ref. 17) has given empirical evidence in support of the lognormal for modeling the residuals. This is in accordance with what the authors have found with respect to the NASA/Air Force Cost Model, for which the lognormal distribution has been shown fits the CER residuals for all subsystems for which there are more a sufficient number of data points to provide a meaningful test. An example of a CER developed for NAFCOM will be discussed in the next section. The gamma distribution also has positive skew, which makes it an appealing choice. The normal distribution, which is symmetric, with thin tails, is not such a good choice. As an example of this consider the example data in Table 2. This example was specifically designed to provide a small example that accurately captures the positive skew effects associated found in NASA cost data.

See Figure 10 for a graphical comparison of the three methods. Note that while MPE-ZPB method produces lowest standard percent error, the overall trend does not match the general trend of the data. Note that MUPE and log-transformed OLS have similar fit.

Weight	Cost
10	3.2
20	4.5
30	30
40	7.1
70	8.4
100	10
200	14.1
300	17.3
500	22.4
1000	31.6

Table 2. Example Data.

This is similar to results reported by Mackenzie (Ref. 18). Also note the similar fits for MPE-ZPB and the normal MLE. The reason for the odd-looking fit for these cases is a sensitivity to the apparent outlier in the data set, namely the pair ***(30, 30)***. This has undue influence on the equation because such an outlier would be nearly impossible if the errors were normally distributed. The normal distribution has thin tails, so the odds of a large outlier such as appears in the data should be infinitesimal. In this case, the standard percent error for the normal MLE fit is 70.8%. The "true" data trend, not taking into account the outlier is the square root function, that is, $\boldsymbol{Y = \sqrt{X}}$. The input value ***30*** should therefore correspond to an output equal to $\boldsymbol{\sqrt{30} \approx \$5.48}$. This is far below the actual cost. Suppose instead that the normal MLE actually fit this trend exactly so that the estimated cost for the normal MLE is ***5.48*** when weight is equal to ***30***. Note that since the residuals are assumed to be normal, the conditional variation about the estimate is

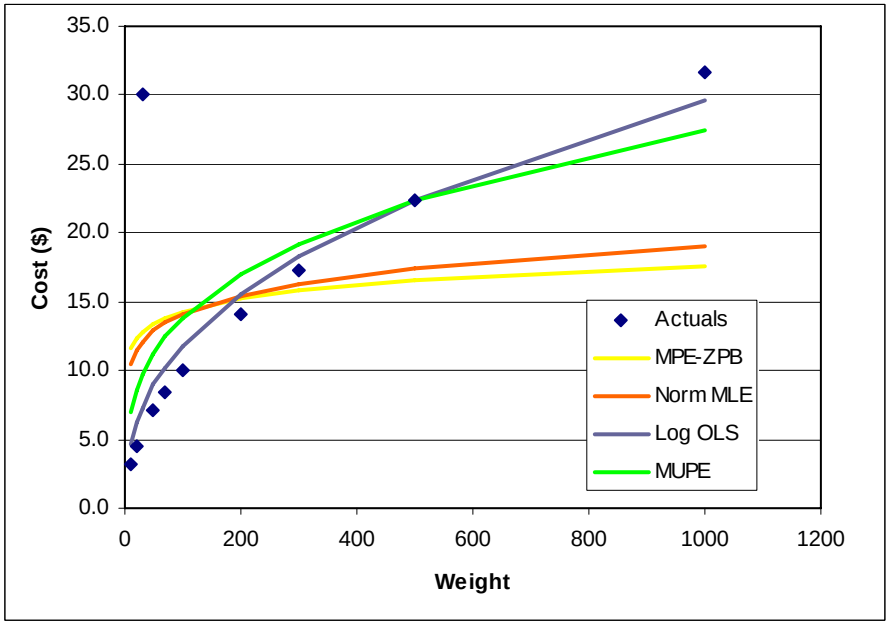

Figure 10. Comparison of Three CER Methods.

normal, with mean equal to ***5.48*1 = 5.48*** and standard deviation equal to ***5.48*** times the standard percent error, which is ***5.48*0.708***, or approximately ***3.88***. Translating this to a standard normal distribution,

$$\frac{30 - 5.48}{3.88} \approx 6.32 .$$

Thus, given that the mean of a standard normal fits the general data trend, the actual cost data is over six standard deviations away from the mean. The odds of such an event are roughly 1 in 7.6 *billion*. See Figure 11 for a graph of the normal distribution with standard deviations and areas for the curve. Because of this, there is a severe penalty assessed by the objective function for underestimating this data point by such a large

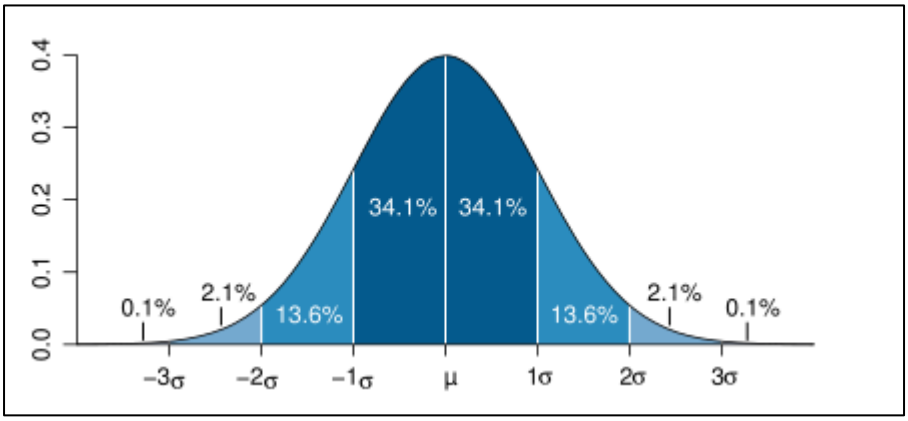

Figure 11. Standard Deviations from the Mean for a Normal Distribution.

amount. The normal MLE estimate when the weight is equal to ***30*** is ***$12.8***.
For this fit the actual cost is approximately ***1.89*** standard deviations above the mean. Based on the standard normal distribution there is a 2.9% chance of observing such an outlier by chance. A large penalty is still assessed for missing this data point, but the normal distribution assumption forces the fit higher until the estimate is within a few standard deviations of the actual estimate. Unless the project being estimated is a close analogy with a particular data point, single data points should not have such large outsized influence on the equation. This outlier could be valid in that it could be due to legitimate events not generally repeatable in practice. Thus undue weight is given to this outlier, since the assumption of normality tries to assume that this outlier is not in the tails of the distribution but closer to the center. The MPE-ZPB estimate is similarly affected because its similarity to the Normal MLE. Thus both the normal MLE and the MPE-ZPB method are not robust in the presence of departures from normality.

Summary of the Methods

Log-transformed ordinary least squares is the oldest of the three methods. This is because of ease of computation. The coefficients can be calculated by hand, while the other methods require the use of a computer. Its strengths are that it works well on skewed data, and it has optimal properties, because it is a maximum likelihood estimator of the median when the residuals are lognormally distributed. All empirical evidence published to date supports the assumption of lognormally distributed residuals for space applications, so this method is optimal in many cases. Since it is a parametric method, we have access to covariance matrices and confidence intervals. Also, based on the generalization derived in this paper, we can use any equation form in the generalized maximum likelihood estimator form. The drawback of this method is that it underestimates the mean. However this can be corrected for by applying an adjustment factor.

In contrast to log-transformed OLS, the MPE-ZPB method is a recent innovation. It requires the use of a computer. It has its merits, chief among them being that it minimizes the standard (percent) error of the estimate, so this method has lower standard percent errors than other methods. By design, the estimator is unbiased. In terms of drawbacks, the method is not explicitly a parametric method, so we don't have access to confidence intervals for the population mean or many other statistical properties. Also as we demonstrated, the method is not robust, because it is sensitive to departures from normality. This aspect makes the method particularly dubious for estimating skewed data. Indeed in some respects the model has the worst of both worlds. While similar to a parametric method, it retains some of its constraints, but on the other hand, it's not a true parametric method, so we don't have access to many of the statistical measures available to us when we use parametric methods.

The MUPE/IRLS method is older than the MPE-ZPB method, since the IRLS method dates back to the 1960s. Unlike log-transformed OLS and MPE-ZPB, it requires weak assumptions to be consistent, but still has confidence intervals and covariance matrices available to fully parametric methods. This method is maximum likelihood if residuals are gamma distributed. In such a case it is asymptotically unbiased, consistent, and efficient. On the other hand, if not an MLE, the method has weak optimality properties ("quasi-likelihood"). Thus the method may not be efficient. Also, the method can be biased for small samples.

Comparison with Other Industries

The analogy with cost estimating in insurance is "loss modeling." In insurance parlance, a "loss" is the amount of a loss experienced by a policyholder. Parametric models are used to estimate both loss size and claim frequency. Both log-transformed OLS and MUPE/IRLS are frequently used to model claim frequency. Fu and Moncher's 2004 presentation (Ref. 19) reports that the gamma and lognormal are the most widely used distributions in loss modeling. They mention 31 recent papers that reported the use of lognormal distributions and 37 that reported the use of gamma distributions for residual modeling. The lognormal is also used in ratemaking and reserve setting, a process not unlike cost risk analysis. Fu and Moncher also study the normal distribution, but find the lognormal and gamma much better for modeling skewed, positive data, like "loss" and "cost." They recommend against use of normal distribution for modeling skewed data because the normal distribution is symmetric. Ismail and Jemain (Ref. 20) also report the widespread use of the lognormal and gamma distributions in loss modeling as of 2009.

Costs are modeled parametrically in health care economics as well (Refs. 21-24). Both log-transformed OLS and IRLS/MUPE are widely used. No mention was found of a normal MLE or MPE-ZPB type method.

Thus log-transformed OLS and MUPE/IRLS are the two leading parametric methods for cost modeling in other industries. Despite wearing the mantle of the oldest method around, log-transformed OLS is a modern method that is still relevant for cost modeling.

The Framework in Action: Application to NAFCOM CERs

The authors have successfully applied this framework to the development of CERs included in the latest version of the NASA/Air Force Cost Model (NAFCOM), which was released in Spring 2009. For previous versions, estimates in NAFCOM have been developed by estimating cost at the subsystem and component level, and then aggregating these estimates to the total system level. The individual subsystem and component level estimates are developed using statistical methods that use historical NASA data as inputs. We again chose this type of model to develop our CERs and the aggregation method to build up our total systems-level estimates.

For this version, we revised the CERs for the first time since 2004. NAFCOM has built-in multivariate CERs that estimate cost using a variety of drivers, including weight, management parameters, and technical factors. For this update we added newly collected data to the data sets, and re-investigated cost drivers for the model. We interviewed subject matter experts for specific hardware elements in order to uncover new, better cost drivers. For example, we previously had included thrust as a cost driver in our reaction control CER. However, we when interviewed a reaction control hardware engineer, he provided us with the insight that thrust level was also important, which led us to stratify the thrust data using attribute variables, enabling us to improve the fit for this particular CER.

With our experience that log-transformed OLS has been a good method for CER development and that users of NAFCOM had reported produced good, credible cost estimates with these CERs, and that the residuals of previous equations had fit lognormal distributions, we formed the hypothesis that the residuals of our equations are lognormally distributed, and developed CERs via log-transformed OLS. As before we were able to derive CERs with good fits, and were able to verify our lognormal hypothesis for every single subsystem CER. We chose the Anderson-Darling test, since it is particularly good at detecting deviation from normality (and hence, lognormality as well). In every case, we were not able to reject the hypothesis at the 5% significance level for those CERs for which we had sufficient number of data points to perform the test.

We did not solely consider log-transform OLS, but also investigated the application of the MPE-ZPB method. We found the MPE-ZPB method not be a good method for our data, mainly for the reasons cited in previous section – the data provided unusual fits and proved to be overly sensitive to individual outlying data points, much more so than any other method. We found this to be the case for every single subsystem CER in NAFCOM. As an example consider the attitude determination and control subsystem design and development CER. For this subsystem, we thoroughly re-investigated the cost drivers, and made some changes to the cost drivers for the new version of NAFCOM. Heritage is one of the most important cost drivers. Heritage is often conflated with technology readiness, since the two are naturally related. Technology development will necessarily require a significant amount of new design effort. However, the converse is not true since it is possible to undertake a completely new design even when using existing state-of-the-practice technology. For this update, we leveraged recent research on separating the impact of these two distinct cost drivers on spacecraft electronics systems

(Ref. 25). As a result we added a new cost driver, a factor similar to technology readiness level that we have termed the technology maturity index. We also revised the heritage driver to only measure the impact of heritage. The complete list of cost drivers for this subsystem is displayed in Table 3.

Weight
Mission Class
Management Rating
Heritage
Technology Maturity Index
Year of Technology
Stabilization Method
Sensors Rating

Table 3. Cost Drivers for Attitude Determination and Control NAFCOM CER.

For this CER we applied log-transformed OLS, the MPE-ZPB method, and the normal MLE method. See Table 4 for a comparison of the goodness-of-fit statistics.

	Log-Transformed OLS	Adjusted Log-Transformed OLS	MPE-ZPB	Normal MLE
Pearson R^2	99.2%	99.2%	99.3%	99.3%
Std. % Error	50.2%	46.4%	41.4%	41.5%
Absolute % Error	36.5%	35.0%	36.3%	36.2%
Bias	-7.2%	-0.4%	0.0%	0.0%

Table 4. Comparison of Goodness-of-Fit Metrics for Comparison of the CER Methods.

In Table 4, note that "Adjusted Log-Transformed OLS" is the traditional log-transformed OLS method with the Ping factor applied to remove the bias. While the standard percent error is lowest for the MPE-ZPB method, the average absolute percent error, which as the name indicates measures the average of the absolute percent errors and as mentioned earlier has been recommended by Foussier as a better alternative to the standard percent error, is lowest for the adjusted log-transformed OLS CER. All four CERs have acceptable goodness-of-fit statistics. However more work is needed to confirm validity. In the case of log-transformed OLS, the residuals must be checked to see if they fit a lognormal distribution. Also the normal MLE residuals must be checked for normality. And in this instance, the normal MLE and the MPE-ZPB CER have approximately the same coefficients and provide almost exactly the same estimates, so if the normal MLE is rejected then so is the MPE-ZPB CER. See Figure 12 for a graphical comparison of the fits, with percentage differences between the MPE-ZPB and the adjusted log-transformed OLS added for contrast. Note that the maximum difference in the normal MLE and MPE-ZPB fits is approximately 2%. The Anderson-Darling test statistic for the log-transformed OLS residuals is 0.3351, much less than the critical value equal to 0.752 at 5% significance, so we cannot reject the hypothesis that the residuals are lognormally distributed. On the other hand the Anderson-Darling statistic for the normal MLE

residuals is 1.0251, above the critical value equal to 0.752 at 5% significance, so we reject the hypothesis that the residuals are normally distributed. Thus even though the

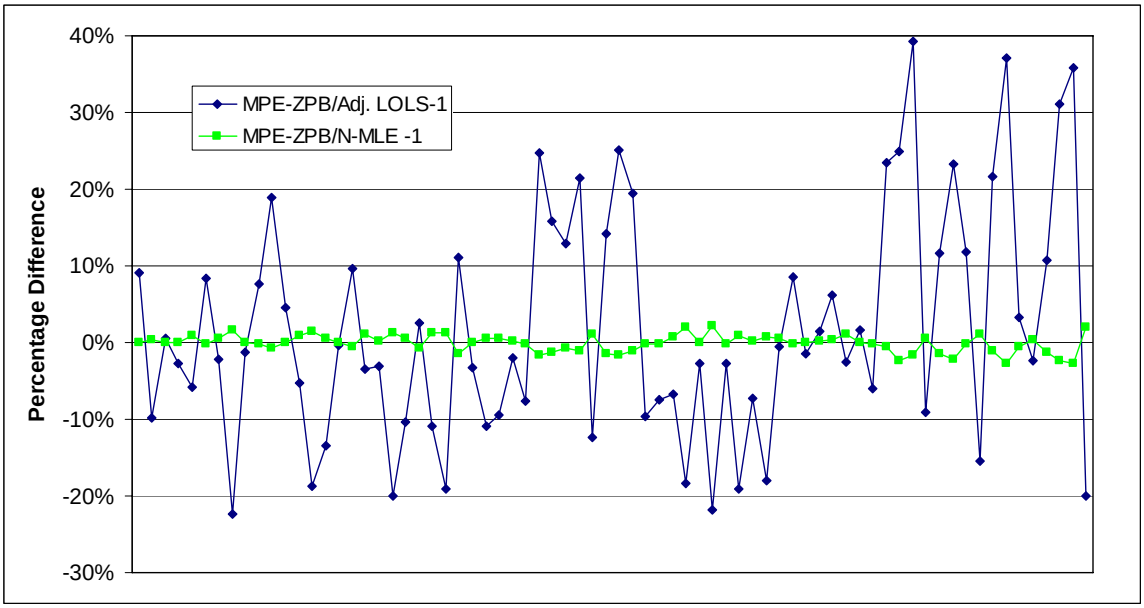

Figure 12. Comparing the MPE-ZPB Estimates with the Log-Transformed OLS and the Normal MLE Estimates.

goodness-of-fit statistics are good for all the CERs considered, the log-transformed OLS method is considered valid for this particular subsystem, while we reject the normal MLE and MPE-ZPB as not valid for this particular application. As we have demonstrated with other examples, the lack of fit means that the method cannot be considered robust or consistent, which means we have no confidence that the normal MLE and hence MPE-ZPB estimated coefficients approximate the underlying true coefficients. Thus sometimes a good fit is not good enough!

Summary

A framework has been introduced for thoroughly and rigorously developing credible cost models. All phases of the process are important, and although many such structured approaches to modeling can be developed, all should at minimum include the phases outlined in this paper, including model selection, parameter calculation, and model validation. All are important but model validation is an often-overlooked and neglected aspect. Proper model validation lends important insights into the cost modeling process.

While some aspects of the framework seem obvious, since a model cannot be developed at all without first selecting a model and then calculating its parameters, a structured approach can be valuable in assisting the analyst in thinking about methods and models in a general framework. In this paper, setting up such a structure allowed the authors to provide a unifying link between three popular CER methods that had not been considered before, namely maximum likelihood estimation. Examination of these methods through the powerful lens of maximum likelihood provided new insight into the log-transformed OLS and MPE-ZPB methods. Log-transformed OLS was found to not be strictly bound

to traditional “ordinary” least squares, but rather a more general method that can admit any equation form. Also, MPE-ZPB was found to have strong links to the normal maximum likelihood estimator. This forces the residuals to have some of the properties of the normal MLE, the most important of which is sensitivity to departure from normal residuals. This leads to a lack of robustness in the method when the residuals are skewed, as is typically the case with spacecraft cost.

References

1. Book, S.A., and P.H. Young, "General-Error Regression for USCM-7 CER Development," The Aerospace Corporation, El Segundo, CA, 1995
2. Book, S.A., and P.H. Young, "General-Error Regression for Deriving Cost-Estimating Relationships," Journal of Cost Analysis, Fall 1997, pp. 1-28.
3. Book, S.A., and N.Y. Lao, "Deriving Minimum-Percentage-Error CERs Under Zero-Bias Constraints," The Aerospace Corporation, El Segundo, CA, July 1996.
4. Book, S.A., "IRLS/MUPE CERs Are Not MPE-ZPB CERs," Presented at the International Society for Parametric Analysts Annual Conference, Seattle, WA, May 23-26, 2006.
5. Book, S.A., and P.H. Young, "The Trouble with R^2," The Journal of Parametrics, Vol. 26, No. 1, Summer 2006, pp. 87-112.
6. Foussier, P. M., From Product Description to Cost: A Practical Approach, Vols. 1 and 2, Springer-Verlag, London, 2006.
7. Goldberg, M.S., and A.E. Tuow, Statistical Methods Learning Curves and Cost Analysis, Institute for Operations Research and Management Sciences, Linthicum, MD, 2003.
8. Hu, S., "The Impact of Using Log-Error CERs Outside the Data Range and Ping Factor," Presented at the Annual Joint ISPA-SCEA Conference, Denver, CO, June, 2005.
9. Klugman, S.A., et al., *Loss Models*, 3rd Ed., John Wiley & Sons, Hoboken, 2008.
10. Eskew, H.L. and K.S. Lawler, "Correct and Incorrect Error Specifications in Statistical Cost Models," Journal of Cost Analysis, Spring 1994.
11. Bahadur, R.R., and L.J. Savage, "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," Annals of Mathematical Statistics 27, 1956, pp 1115-1122.
12. Lee, D.A., *The Cost Analyst's Companion*, Logistics Management Institute, McLean, VA, 1997.
13. Goldberg, M.S., and A.E. Tuow, *Statistical Methods Learning Curves and Cost Analysis*, Institute for Operations Research and Management Sciences, Linthicum, MD, 2003.
14. Nelder, J. A., "Weighted Regression, Quantal Response Data, and Inverse Polynomials," Biometrics, Vol. 24 (1968), pages 979-985.
15. Wedderburn, R.W.M., "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," Biometrika, Vol. 61, Number 3 (1974), pages 439-447.
16. Foussier, P. M. and P. Foussier, "Should We Use the Median Instead of the OLS?," Parametric World, Fall 2008, pp. 8-11.
17. Mackenzie, D., et al., "Top Level Spacecraft Cost Distribution Study," Joint Annual ISPA-SCEA Conference, Noordwijk, May, 2008.
18. Mackenzie, D., "Cost Estimating Relationship Variance Study," AIAA Space Conference, 2003, Long Beach, CA.
19. Fu, L., and R. Moncher, "Severity Distributions for GLMs: Gamma or Lognormal?", 2004 CAS Spring Meeting, Colorado Springs, CO, 2004.
20. Ismail, N., and A.A. Jemain, "Comparison of Minimum Bias and Maximum Likelihood Methods for Claim Severity," Casualty Actuarial Society E-Forum, Winter 2009.
21. Manning, W.G., and J. Mullahy, "Estimating Log Models: To Transform or Not to Transform?", Journal of Health Economics, 2001

22. Basu, A. et al., “Comparing Alternative Models: Log Vs. Cox Proportional Hazard,” Health Economics, 2004.
23. Manning, W.G., et al., “Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data,” Journal of Health Economics, 2005.
24. Gallin, J.H., “Net Migration and State Labor Market Dynamics,” Journal of Labor Economics, 2004.
25. Smart, C., “Assessing the Effect of Technology Development on Avionics,” Presented at the NASA Cost Symposium, Cleveland, OH, June, 2006.