Bayesian Parametrics: How to Develop a CER with Limited Data and Even Without Data

Christian Smart, Ph.D.
Director, Cost Estimating and Analysis
Missile Defense Agency
Christian.smart@mda.mil
256-450-4936

April 14, 2014

Abstract

During college, my mathematics and economics professors were adamant in telling me that I needed at least two data points to define a trend. You may have been taught this same dogma. As it turns out, this is wrong. You can define a trend with one data point, and even without any data at all. A cost estimating relationship (CER), a mathematical equation that relates cost to one or more technical inputs, is a specific application of trend analysis. The purpose of this paper is to introduce methods for applying CERs to small data sets, including the cases of one data point and no data.

The only catch is that you need some prior information for one or more of the CER's parameters. For example, consider a linear CER with one explanatory variable: $\tilde{y} = a + bX$. The slope of the equation, b, can be interpreted as an economies of scale factor. As such, it is typically between 0 and 1. When using weight as the explanatory variable, rules of thumb are 0.5 for development cost and 0.7 for production cost [4]. Bayes' Theorem can be applied to combine prior information with sample data to produce CERs using small data sets.

This paper discusses Bayes' Theorem and applies it to linear and nonlinear CERs, including ordinary least squares and log-transformed ordinary least squares.

Introduction

You may be familiar with the Law of Large Numbers (sample mean converges to the expected value as the size of the sample increases), but you may not be familiar with the Law of Small Numbers. The Law of Small Numbers, coined by the mathematician Richard Guy [3], states that there are never enough small numbers to meet all the demands placed upon them. This points out the difficulty in conducting statistical analysis with small data sets. However, there is a need for developing cost estimates for programs that have limited prior experience. For example, NASA has not developed many launch vehicles, yet there is a need to understand how much a new launch vehicle will cost. For insurance companies, there is a need to write policies for people who have never been insured. In political polling, there is a need to forecast voter turnout and election results for rural counties with low populations that report little data. One way to approach these problems is to use Bayesian analysis. This is named after the Rev. Thomas Bayes, who first developed this approach [8]. Bayesian analysis combines prior experience, or opinion, with sample data to conduct statistical analysis. This prior experience can be subjective, or based on similar, more general data, applied to a specific subset, as in a hierarchical model.

Bayesian analysis has proven to be successful in a multitude of applications. These techniques were used in World War II to help crack the Enigma code used by the Germans, thus helping to shorten the war. John Nash's equilibrium for games with incomplete or imperfect information is a form of Bayesian analysis. Actuaries have used these technicques for over 100 years to set insurance premiums. Bayesian voice recognition researchers applied their skills as leaders of the portfolio and technical trading team for the Medallion Fund, a $5 billion hedge fund which has averaged annual returns of 35% after fees since 1989 [8].

Recently, one of the most well-known recent examples of applying Bayesian methods is Nate Silver's election predictions. In his 2012 book *The Signal and the Noise*, advocates the use of Bayesian techniques in forecasting. Silver provides election forecasts via his blog, http://fivethirtyeight.blogs.nytimes.com. His recent results have been extremely accurate. In 2008 he correctly predicted the winner of 49 of 50 states in the presidential election. The only state Silver missed was Indiana, which went to Barack Obama by one percentage point. He correctly predicted the winner of all 35 U.S. Senate races that year. In the 2012 United States presidential election between Barack Obama and Mitt Romney, Silver correctly predicted the winner of all 50 states and the District of Columbia. That same year, his predictions of U.S. Senate races were correct in 31 of 33 states; he predicted Republican victory in North Dakota and Montana, where Democrats won.

The Bayesian method can also be applied to cost analysis. Cost estimating relationships (CERs) are important tools for cost estimators. The limitation is that they require a significant amount of data. However, it is often the case that we have small amounts of data in cost estimating. We show how to apply Bayes' Theorem to regression-based CERs, and in the process, leveraging prior experience and information to be able to apply CERs to small data sets.

Indeed, small data sets are the ideal setting for the application of Bayesian techniques for cost analysis. Given large data sets that are directly applicable to the problem at hand, a straightforward regression analysis will likely be probably the preferred method. However, when applicable data is limited, I suggest leveraging prior experience to develop accurate estimates. The idea of applying significant prior experience with limited data has been termed "thin-slicing" by Malcolm Gladwell in his best-selling book *Blink* [2]. In his book, Gladwell presents several examples of how experts can make accurate predictions with limited data. One example is the case of a marriage expert who can analyze a conversation between a husband and wife for an hour and can predict with 95% accuracy whether the couple will be married 15 years later. If he analyzes a couple for 15 minutes he can predict the same result with 90% accuracy.

Bayes' Theorem

The conditional probability of event A given event B is denoted by $Pr(A \mid B)$. In its discrete form, Bayes' Theorem states that

$$Pr(A \mid B) = \frac{Pr(A)Pr(B \mid A)}{Pr(B)}. \tag{1}$$

Let's consider the example of testing for the use of illegal drugs. Many have had to take such a test as a condition of employment with the federal government. What is the probability that someone who fails a drug test does not actually use illegal drugs? Bayes' Theorem can be used to answer such questions.

Suppose that 95% of the population does not use illegal drugs. Also suppose that the drug test is highly accurate. If someone is a drug user, it returns a positive result 99% of the time. If someone is not a drug user, the test returns a false positive only 2% of the time.

In this case: A is the event that someone does not use illegal drugs, and B is the event that someone tests positive for illegal drugs. The complement of A, denoted A', is the event that an individual is a user of illegal drugs.

From the law of total probability,

$$Pr(B) = Pr(B \mid A)Pr(A) + Pr(B \mid A')Pr(A'). \tag{2}$$

Thus, substituting eq. 2 into eq. 1, Bayes' Theorem is equivalent to:

$$Pr(A \mid B) = \frac{Pr(B \mid A)Pr(A)}{Pr(B \mid A)Pr(A) + Pr(B \mid A')Pr(A')}. \tag{3}$$

We see that the probability of someone who fails a drug test but is not an illegal drug user can be calculated by plugging in the appropriate values into eq. 3,

$$Pr(A \mid B) = \frac{0.02(0.95)}{0.02(0.95) + 0.99(0.05)} \approx 27.7\%. \tag{4}$$

Therefore, even with accurate drug tests, it is easy to obtain false positives. This is a case of inverse probability, a kind of statistical detective work where we try to determine whether someone is innocent or guilty based on revealed evidence.

More typical of the kind of problem that we want to solve is the following: We have some prior evidence or opinion about a subject, and we also have some direct empirical evidence. How do we take our prior evidence and combine it with the current evidence to form an accurate estimate of a future event?

It's simply a matter of interpreting Bayes' Theorem. Let $Pr(A)$ be the probability that we assign to an event before seeing the data, the *prior* probability. Then let

$Pr(A \mid B)$ be the probability after we see the data, the *posterior* probability. Thus $\frac{Pr(B|A)}{Pr(B)}$ is the probability of observing these data given the hypothesis, the *likelihood*.

Bayes' Rule can be re-stated as

$$Posterior \propto Prior \times Likelihood. \tag{5}$$

An example of this application of Bayes' Theorem can be found in the Monty Hall Problem. This is based on the television show *Let's Make a Deal*, whose original host was Monty Hall. In this version of the problem, there are three doors. Behind one door is a car. Behind each of the other two doors is a goat. Suppose you pick door #1. Monty, who knows what is behind each door, then opens door #3, showing you a goat behind it. He then asks if you want to pick door #2 instead, see Figure 1. Is it to your advantage to switch doors?

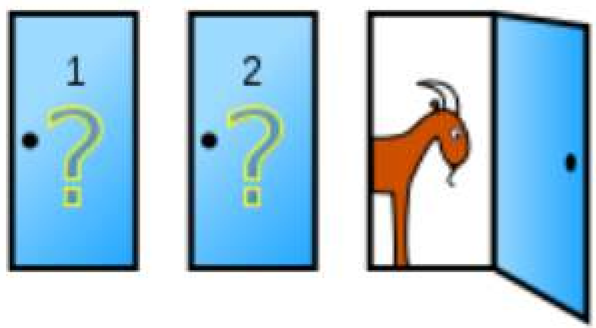

Figure 1: Monty Hall Problem.

To solve this problem, let A_1 denote the event that the car is behind door #1, A_2 the event that the car is behind door #2, and A_3 the event that the car is behind door #3. Your original hypothesis is that there was an equally likely chance that the car was behind any one of the three doors. Thus the prior probability, before the third door is opened, that the car was behind door #1, which we denote $Pr(A_1) = \frac{1}{3}$. Also, $Pr(A_2) = Pr(A_3) = \frac{1}{3}$.

Once you picked door #1, you were given additional information. You were shown that a goat is behind door #3. Let B denote the event that you are shown that a goat is behind door #3. The probability that there is a goat behind door #3 is best calculated by considering three conditional probabilities.

The probability that you are shown the goat is behind door #3 is an impossible event if the car is behind door #3. Thus $Pr(B \mid A_3) = 0$. Since you picked door #1, Monty will open either door #2 or door #3, but not door #1. Thus, if the car is actually behind door #2, it is a certainty that Monty will open door #3 and show you a goat. Thus $Pr(B \mid A_2) = 1$. If you have picked correctly and have chosen the right door, then there are goats behind both door #2 and door #3. In this case, there is a 50% chance that Monty will open door #2 and a 50% chance that he will open door #3. Thus $Pr(B \mid A_1) = \frac{1}{2}$.

By Bayes' Theorem,

$$Pr(A_1 \mid B) = \frac{Pr(A_1)Pr(B \mid A_1)}{Pr(A_1)Pr(B \mid A_1) + Pr(A_2)Pr(B \mid A_2) + Pr(A_3)Pr(B \mid A_3)}. \quad (6)$$

Plugging in the probabilities that we have derived, we find that

$$Pr(A_1 \mid B) \;=\; \frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 1+\frac{1}{3}\cdot 0} = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}} = \frac{1}{3}, \quad (7)$$

and

$$Pr(A_2 \mid B) = \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 1+\frac{1}{3}\cdot 0} = \frac{\frac{1}{3}}{\frac{1}{6}+\frac{1}{3}} = \frac{2}{3}. \quad (8)$$

And since you already know that the car is not behind door #3, $Pr(B) = 0$.

Thus you have a $\frac{1}{3}$ chance of picking the car if you stick with you initial choice of door #1, but a $\frac{2}{3}$ chance of picking the car if you switch doors. It is in your interest to switch doors.

Did you think there was no advantage to switching doors? You are not alone. Marilyn Vos Savant, famous for having the world's highest IQ at 228, wrote a column for Parade magazine for many years about this same question. In 1990, a reader posed the Monty Hall problem to her and she provided the correct answer. But many people, including people with Ph.D.s, some mathematicians, derided Marilyn for being wrong (see http://marilynvossavant.com/game-show-problem/ for more information). Even the famous mathematician Paul Erdos found the problem to be counterintuitive [5]. But the correct answer is that once door #3 is opened and revealed to have a goat behind it, there is a two-thirds chance that the car is behind door #2. If you are still not convinced, conduct a Monte Carlo simulation to see that this is the correct answer.

For our application of Bayes' Theorem to cost estimating we will need the continuous form of eBayes' Theorem. If the *prior* distribution is continuous, Bayes' Theorem is written as

$$\pi(\theta \mid x_1, \ldots, x_n) = \frac{\pi(\theta) f(x_1, \ldots, x_n \mid \theta)}{f(x_1, \ldots, x_n)} = \frac{\pi(\theta) f(x_1, \ldots, x_n \mid \theta)}{\int \pi(\theta) f(x_1, \ldots, x_n \mid \theta) d\theta} \quad (9)$$

where:

$\pi(\theta)$ is the *prior density*, the initial density function for the parameters that varies in the model. It is possible to define an *improper prior density*, one which is nonnegative but whose integral is infinite.

$f(x \mid \theta)$ is the conditional probability density function of the model. It defines the model's probability given the parameter θ;

$f(x_1, \ldots, x_n \mid \theta)$ is the conditional joint probability density function of the data given θ. Typically the observations are assumed to be independent given θ, and in this case,

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) \quad (10)$$

where $f(x_1, \ldots, x_n)$ is the unconditional joint density function of the data $x_1, \ldots, x_n$. It is calculated from the conditional joint density function by integrating over the prior density function of θ:

$$f(x_1, \ldots, x_n) = \int \pi(\theta) f(x_1, \cdots, x_n \mid \theta) d\theta. \tag{11}$$

$\pi(\theta \mid x_1, \ldots, x_n)$ is the *posterior density function*, the revised density function for the parameter θ based on the observations $x_1, \ldots, x_n$.

$f(x_{n+1} \mid x_1, \ldots, x_n)$ is the *predictive density function*, the revised unconditional density based on the sample data. It is calculated by integrating the conditional probability density function over the posterior density of θ:

$$f(x_{n+1} \mid x_1, \ldots, x_n) = \int f(x_{n+1} \mid \theta) \pi(\theta \mid x_1, \ldots, x_n) d\theta. \tag{12}$$

Application of Bayes' Theorem to Ordinary Least squares CERs

In this section, we consider ordinary least squares (OLS) CERs of the form

$$Y = a + bX + \epsilon \tag{13}$$

The application of Bayes' Theorem involves prior distributions about a and b, as well as ϵ.

For the application of Bayes' Theorem, we will write this in mean deviation form:

$$Y = \alpha_{\bar{X}} + \beta(X - \bar{X}) + \epsilon. \tag{14}$$

This form makes it easier to establish prior inputs, since it is easier to think of an average value for prior cost than it is for the intercept of the least-squares equation.

Given a sample of data points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, the likelihood function can be written as

$$L(\alpha_{\bar{X}}, \beta) \propto \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(Y_i - (\alpha_{\bar{X}} + \beta(X_i - \bar{X})))^2} \tag{15}$$

$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - (\alpha_{\bar{X}} + \beta(X_i - \bar{X})))^2}. \tag{16}$$

The expression $\sum_{i=1}^{n}(Y_i - (\alpha_{\bar{X}} + \beta(X_i - \bar{X})))^2$ can be simplified as

$$\sum_{i=1}^{n} \left(Y_i - \bar{Y} + \bar{Y} - \left(\alpha_{\bar{X}} + \beta\left(X_i - \bar{X}\right)\right)\right)^2 \tag{17}$$

$$= \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 + 2\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)\left(\bar{Y} - \left(\alpha_{\bar{X}} + \beta\left(X_i - \bar{X}\right)\right)\right) + \sum_{i=1}^{n} \left(\bar{Y} - \left(\alpha_{\bar{X}} + \beta\left(X_i - \bar{X}\right)\right)\right)^2$$

which reduces to

$$SS_y - 2\beta SS_{xy} + n\left(\bar{Y} - \alpha_{\bar{X}}\right)^2 + \beta^2 SS_x \tag{18}$$

since $\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right) = 0$ and $\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = 0$, where

$$\begin{aligned} SS_y &= \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 && (19)\\ SS_x &= \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 && (20)\\ SS_{xy} &= \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right). && (21) \end{aligned}$$

Thus, the joint likelihood of $\alpha_{\bar{X}}$ and β is proportional to

$$\begin{aligned} e^{-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n\left(\alpha_{\bar{X}} - \bar{Y}\right)^2\right)} &= e^{-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x\right)} e^{-\frac{1}{2\sigma^2}\left(n\left(\alpha_{\bar{X}} - \bar{Y}\right)^2\right)} && (22)\\ &= e^{-\frac{1}{2\frac{\sigma^2}{SS_x}}\left(\frac{SS_y}{SS_x} - \frac{2\beta SS_{xy}}{SS_x} + \beta^2\right)} e^{-\frac{1}{2\frac{\sigma^2}{n}}\left(\alpha_{\bar{X}} - \bar{Y}\right)^2} \end{aligned}$$

Completing the square on the innermost expression in the first term yields

$$\begin{aligned} \beta^2 - 2\beta\frac{SS_{xy}}{SS_x} + \frac{SS_y}{SS_x} &= \beta^2 - 2\beta\frac{SS_{xy}}{SS_x} + \frac{SS_{xy}^2}{SS_x^2} - \frac{SS_{xy}^2}{SS_x^2} + \frac{SS_y}{SS_x} && (23)\\ &= \left(\beta - \frac{SS_{xy}}{SS_x}\right)^2 + \text{constant}. \end{aligned}$$

This means that likelihood is proportional to

$$e^{-\frac{1}{2\frac{\sigma^2}{SS_x}}\left(\beta - \frac{SS_{xy}}{SS_x}\right)^2} e^{-\frac{1}{2\frac{\sigma^2}{n}}\left(\alpha_{\bar{X}} - \bar{Y}\right)^2} = L(\beta)\, L(\alpha_{\bar{X}}) \tag{24}$$

so the likelihoods are independent.

Note $\frac{SS_{xy}}{SS_x} = B$, the least squares slope, and $\bar{Y} = A_{\bar{X}}$, the least squares estimate of the average. The likelihood of the slope β follows a normal distribution with mean B and variance $\frac{\sigma^2}{SS_x}$. The likelihood of the average $\alpha_{\bar{X}}$ follows a normal distribution with mean, $A_{\bar{X}}$, and variance, $\frac{\sigma^2}{n}$.

The joint prior for β and $\alpha_{\bar{X}}$, $g(\alpha_{\bar{X}}, \beta)$, has the property that

$$g(\alpha_{\bar{X}}, \beta) = g(\alpha_{\bar{X}})g(\beta) \tag{25}$$

By Bayes' Theorem, the joint posterior density function is proportional to the joint prior times the joint likelihood:

$$g(\alpha_{\bar{X}}, \beta \mid \{(x_1, y_1), \ldots, (x_n, y_n)\}) = g(\alpha_{\bar{X}}, \beta) \text{ sample likelihood } (\alpha_{\bar{X}}, \beta). \tag{26}$$

If the prior density for β is normal with mean m_β and variance s_β^2, we obtain a normal posterior with mean m'_β and variance s'^2_β, where

$$m'_\beta = \frac{\frac{1}{s_\beta^2}}{\frac{1}{s'^2_\beta}} m_\beta + \frac{\frac{SS_x}{\sigma^2}}{\frac{1}{s'^2_\beta}} B \tag{27}$$

and

$$\frac{1}{s'^2_\beta} = \frac{1}{s^2_\beta} + \frac{SS_x}{\sigma^2}. \tag{28}$$

If the prior density for $\alpha_{\bar{X}}$ is normal with mean $m_{\alpha_{\bar{X}}}$ and variance $s^2_{\alpha_{\bar{X}}}$, then we obtain a normal posterior with mean $m'_{\alpha_{\bar{X}}}$ and variance $s'^2_{\alpha_{\bar{X}}}$ where

$$m'_{\alpha_{\bar{X}}} = \frac{\frac{1}{s^2_{\alpha_{\bar{X}}}}}{\frac{1}{s'^2_{\alpha_{\bar{X}}}}} m_{\alpha_{\bar{X}}} + \frac{\frac{n}{\sigma^2}}{\frac{1}{s'^2_{\alpha_{\bar{X}}}}} A_{\bar{X}} \tag{29}$$

and

$$\frac{1}{s'^2_{\alpha_{\bar{X}}}} = \frac{1}{s^2_{\alpha_{\bar{X}}}} + \frac{n}{\sigma^2}. \tag{30}$$

What we really need is the predictive equation, not just the posterior values of the parameters. But in this case the predictive mean is equal to the posterior mean, i.e.,

$$\mu_{n+1} = m'_{\alpha_{\bar{X}}} + m'_\beta(X_{n+1} - \bar{X}). \tag{31}$$

This result follows from the fact that the prior distribution is normal, and the likelihood is normal, resulting in a predictive distribution that is normal as well. This is the case of a conjugate prior.

Consider the case of a non-informative improper prior such as $\pi(\alpha_{\bar{X}}) = 1$ for all $\alpha_{\bar{X}}$. By independence, β is calculated as before and $\alpha_{\bar{X}}$ is calculated as

$$L(\alpha_{\bar{X}}) = e^{\frac{-1}{2\frac{\sigma^2}{n}}(\alpha_{\bar{X}} - \bar{Y})^2} \tag{32}$$

which follows a normal distribution with mean $\bar{Y}$ and variance $\frac{\sigma^2}{n}$. This is equivalent to the sample mean of $\alpha_{\bar{X}}$ and the variance of the sample mean $s^2_{\alpha_{\bar{X}}}$. Thus in the case where we only have information about the slope, the sample mean of actual data is used for $\alpha_{\bar{X}}$.

For each parameter, the updated estimate incorporating both prior information and sample data is weighted by the inverse of the variance of each estimate. The inverse of the variance is called the precision. This result can be generalized to the linear combination of any two estimates that are independent and unbiased.

Theorem: If two estimators are unbiased and independent, then the minimum variance estimate is the weighted average of the two estimators with weights that are inversely proportional to the variance of the two estimators.

Proof: Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be two independent, unbiased estimators of a random variable θ, where $E(\tilde{\theta}_1) = E(\tilde{\theta}_2) = \theta$. Since both are unbiased, the weighted average of the two is also unbiased. In order to confirm this, let w and $1 - w$ be the weights used in the weighted average of $\tilde{\theta}_1$ and $\tilde{\theta}_2$, respectively. Then the weighted average is also an unbiased estimator of θ since

$$E(w\tilde{\theta}_1 + (1-w)\tilde{\theta}_2) = wE(\tilde{\theta}_1) + (1-w)E(\tilde{\theta}_2) = w\theta + (1-w)\theta = \theta, \tag{33}$$

$\tilde{\theta}_1$ and $\tilde{\theta}_2$ are independent. Thus, the variance of the weighted average is

$$Var(w\tilde{\theta}_1 + (1-w)\tilde{\theta}_2) = w^2 Var(\tilde{\theta}_1) + (1-w)^2 Var(\tilde{\theta}_2). \tag{34}$$

To determine the weights that minimize the variance, define ϕ to be a function of w, that is,

$$\phi(w) = w^2 Var(\tilde{\theta}_1) + (1-w)^2 Var(\tilde{\theta}_2) \tag{35}$$

Now take the first derivative of ϕ with respect to w and set it equal to zero:

$$\phi'(w) = 2wVar(\tilde{\theta}_1) - 2(1-w)Var(\tilde{\theta}_2) = 2wVar(\tilde{\theta}_1) + 2wVar(\tilde{\theta}_2) - 2Var(\tilde{\theta}_2) = 0. \tag{36}$$

Note the second derivative is $\phi''(w) = 2Var(\tilde{\theta}_1) + 2Var(\tilde{\theta}_2)$, confirming the value of w which satisfies the equation $\phi'(w) = 0$ will be a minimum.

Solving for w, we find that

$$w = \frac{Var(\tilde{\theta}_2)}{Var(\tilde{\theta}_1) + Var(\tilde{\theta}_2)}. \tag{37}$$

Multiplying both the numerator and the denominator by $\frac{1}{Var(\tilde{\theta}_1)Var(\tilde{\theta}_2)}$ yields

$$w = \frac{\frac{1}{Var(\tilde{\theta}_1)}}{\frac{1}{Var(\tilde{\theta}_1)} + \frac{1}{Var(\tilde{\theta}_2)}}. \tag{38}$$

It immediately follows that

$$1 - w = \frac{\frac{1}{Var(\tilde{\theta}_2)}}{\frac{1}{Var(\tilde{\theta}_1)} + \frac{1}{Var(\tilde{\theta}_2)}}. \tag{39}$$

Note this rule extends to more than two estimates.

This leads to the Precision-Weighting Rule for combining prior experience and sample data:

Precision-Weighting Rule for Combining Two Parametric Estimates

Given two independent, unbiased parametric estimates $\tilde{\theta}_1$ and $\tilde{\theta}_2$ with precisions $\rho_1 = \frac{1}{Var(\tilde{\theta}_1)}$ and $\rho_2 = \frac{1}{Var(\tilde{\theta}_2)}$, respectively, the minimum variance estimate of weighted average is given by

$$\frac{\rho_1}{\rho_1+\rho_2}\tilde{\theta}_1 + \frac{\rho_2}{\rho_1+\rho_2}\tilde{\theta}_2.$$

The precision-weight approach has desirable properties. It is an uniformly minimum variance unbiased estimator (UMVUE). This approach minimizes the *mean squared error*, which is defined as

$$MSE_{\tilde{\theta}}(\theta) = E\left[(\tilde{\theta} - \theta)^2 \mid \theta\right]. \tag{40}$$

In general, the lower the mean squared error, the better the estimator. The mean squared error is widely accepted as a measure of accuracy [7]. Thus, the precision-weighted approach which minimizes the mean squared error, has optimal properties. You may be familiar with this as the ”least squares criterion” from linear regression. We provide examples of applying this method to updating the parameters in the next section.

Application to Log-Transformed Ordinary Least Squares CERs

For an example based on real data, consider earth-orbiting satellite cost and weight trends. Goddard Space Flight Center's Rapid Spacecraft Development Office (RSDO) is designed to procure satellites cheaply and quickly. Their goal is to quickly acquire a spacecraft for launching already designed payloads using fixed-price contracts. They claim that this approach mitigates cost risk. If this is the case their cost should be less than the average earth-orbiting spacecraft. For more on RSDO see http://rsdo.gsfc.nasa.gov/.

Data on earth-orbiting spacecrafts is plentiful, while data for RSDO has a much smaller sample size. When I did some analysis in 2008 to compare the cost of non-RSDO earth-orbiting satellites with RSDO missions, I had a database with 72 non-RSDO missions from NAFCOM and 5 RSDO missions. Figure 2 is a scatter plot showing cost vs. weight for these two data sets, along with best-fit LOLS trend lines.

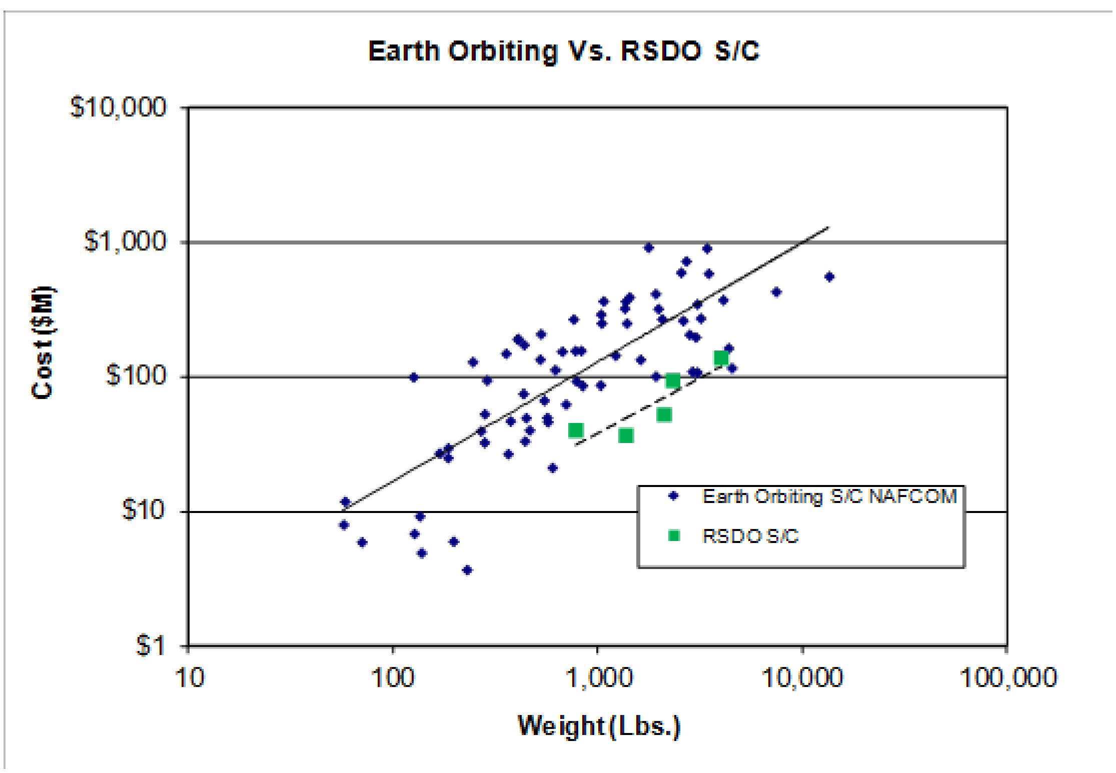

Figure 2: Earth-orbiting and RSDO Spacecraft Cost and Weight Trends.

Note Figure 2 is a log-scale graph. Power equations of the form $\tilde{Y} = aW^b$ were fit to both data sets using log-transformed ordinary least squares. The b-value previously mentioned is a measure of the economy of scale. For NAFCOM the value is 0.89 and for RSDO it is 0.81. This would seem to indicate greater economies of scale for the RSDO spacecraft. Even more significant is the difference in the magnitude of costs between the two data sets. The log-scale graph understates the difference, so seeing a significant difference between two lines plotted on a log-scale graph is very telling. For example, for a weight equal to 1,000 lbs. the estimate based on RSDO data is 70% less than the data based on earth-orbiting spacecraft data from NAFCOM.

Since we have five data points, why not just use the regression analysis from the RSDO data? Because the RSDO data set consists of only five data points! However, the Bayesian approach allows us to combine the Earth-Orbiting Spacecraft data with the smaller data set. We use a hierarchical approach, treating the earth-orbiting spacecraft data from NAFCOM as the prior, and the RSDO data as the sample. This is exactly the type of approach that Nate Silver uses to develop highly accurate elections forecasts in small population areas and areas with little reported data. This is also the approach that actuaries use when setting premiums for insurances with little data.

Because we have used log-transformed OLS to develop the regression equations, we are assuming that the residuals are lognormally distributed, and thus normally distributed in log-space. We will use the approach for updating normally distributed priors with normally distributed data to estimate the precisions. These precisions will then determine the weights we assign the parameters.

To apply LOLS, we transform the equation $\tilde{Y} = aW^b$ to log-space by applying the natural log function to each side, i.e.

$$\ln(\tilde{Y}) = \ln(aW^b) = \ln(a) + b\dot{\ln}(W). \tag{41}$$

Once the data are transformed, ordinary least squares regression is applied to both the NAFCOM data and to the RSDO data.

As data are available for both data sets, opinion is not used. The precisions used in evaluating the combined equations are calculated from the regression statistics. In keeping with the model,

$$\ln(Cost) = \ln(X) - \sum_{i=1}^{n} \frac{\ln X_i}{n}. \tag{42}$$

From the regressions we need the values of the parameters as well as the variances of the parameters. Statistical software packages provide both the parameters and their variances as outputs. Using the Data Analysis add-in in Excel, the Summary Output table provides these values. Table 1 is the output of the regression for the NAFCOM data. Note the "Coefficients" and "Standard Errors" for Intercept and X Variable 1 are the mean and variance of the parameters.

In Table 1, "Intercept" is $\alpha_{\tilde{X}}$ and "X Variable 1" is the slope parameter β. The "Standard Error" is the standard deviation of the respective parameter. We square this value to obtain the variance. Table 2 contains the values for the means, variances, precisions, and combined values for the parameters.

SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.79439689				
R Square	0.63106642				
Adjusted R^2	0.62579595				
Standard Error	0.81114468				
Observations	72				
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	78.78101763	78.78101763	119.7361	8.27045E-17
Residual	70	46.05689882	0.657955697		
Total	71	124.8379164			
	Coefficients	Standard Error	t-Stat	P-value	Lower 95%
Intercept	4.60873098	0.095594318	48.1134863	1.95E-55	4.418074125
X Variable 1	0.88578231	0.080949568	10.942397	8.27E-17	0.724333491

Table 1: Regression Statistics Summary for the NAFCOM Data.

Parameter	NAFCOM Mean	NAFCOM Variance	NAFCOM Precision	RSDO Mean	RSDO Variance	RSDO Precision	Combined Mean
$\alpha_{\bar{X}}$	4.6087	0.0091	109.4297	4.1359	0.0201	49.8599	4.4607
β	0.8858	0.0065	152.6058	0.8144	0.0670	14.9298	0.8794

Table 2: Combining the Parameters.

In Table 2, the mean of each parameter is the value calculated by the regression. The variance is the square of the standard error. The precision is the inverse of the variance. The combined mean is calculated by weighting each parameter by its relative precision. Using eq.38, for the intercept the relative precision weights are

$$\frac{\frac{1}{0.0091}}{\frac{1}{0.0091}+\frac{1}{0.0201}} = \frac{109.4297}{109.4297+49.8599} \approx 0.6870 \tag{43}$$

for the NAFCOM data, and $1-0.6870 = 0.3130$ for the RSDO data. Likewise if we use eq. 38 to calculate the slope the relative precision weights, we obtain

$$\frac{\frac{1}{0.0065}}{\frac{1}{0.0065}+\frac{1}{0.0670}} = \frac{152.6058}{152.6058+14.9298} \approx 0.9109 \tag{44}$$

for the NAFCOM data, and $1-0.9109 = 0.0891$ for the RSDO data. The combined intercept is $0.6870 \cdot 4.6087 + 0.3130 \cdot 4.1359 \approx 4.4607$ and the combined slope is $0.9109 \cdot 0.8858 + 0.0891 \cdot 8144 \approx 0.8794$.

The log-space equation is $\tilde{Y} = \alpha_{\bar{X}} + \beta(X-\bar{X})$, thus the combined equation in log-space is $\tilde{Y} = 4.4607 + 0.8794(X-\bar{X})$. The only remaining question is what value to use for $\bar{X}$. We have two data sets. But since we consider the first data set as the prior information, the mean is calculated from the second data set, that is, from the RSDO data. The log-space mean of the RSDO weights is 7.5161. Thus the log-space equation is

$$\begin{aligned}\tilde(Y) &= 4.4607 + 0.8794(X - \bar{X}) \quad (45)\\ &= 4.4607 + 0.8794(X - 7.5161) \quad (46)\\ &= -2.1491 + 0.8794X. \quad (47)\end{aligned}$$

This equation is in log-space, that is,

$$\ln(\widetilde{\text{Cost}}) = -2.1491 + 0.8794 \ln(Wt). \quad (48)$$

In linear space, this is equivalent to

$$\widetilde{\text{Cost}} = 0.1166Wt^{0.8794}. \quad (49)$$

See Figure 3 for a comparison of all three trendlines: the solid line based on NAFCOM earth-orbiting Spacecraft (EO) data, the dashed line based only on RSDO data, and the dotted line is the Bayesian combination of the two estimates.

One RSDO data point not in the data set that launched in 2011 was the Landsat Data Continuity Mission (now Landsat 8). The Landsat Program provides repetitive acquisition of high resolution multispectral data of the Earth's surface on a global basis. The Landsat satellite bus weighs $4,566$ lbs. Using the Bayesian equation, eq.49, the predicted cost is

$$\widetilde{\text{Cost}} = 0.1166 \cdot 4566^{0.8794} \approx \$192\text{Million} \quad (50)$$

which is equal to the actual cost for the spacecraft bus! The RSDO data alone predicts a cost equal to $131 Million 31% below the actual cost), while the earth-orbiting data alone predicts a cost equal to $492 million (156% higher than the actual cost). While this is only one data point, this seems promising. It is significantly more accurate than using regression analysis on the earth-orbiting NAFCOM data or regression analysis on the RSDO data alone.

Note that the range of the RSDO data is narrow compared to the larger NAFCOM data set. The weights of the missions in the NAFCOM data set range from 57 lbs. to 13,448 lbs. The range of the missions in the RSDO data set range from 780 lbs. to 4,000 lbs. One issue with using the RSDO data alone is that it is likely you will need to estimate outside the range of the data, which is problematic for a small data set. Combining the RSDO data with a larger date set with a wider range provides confidence in estimating outside the limited range of a small data set.

To summarize the hierarchical approach we begin by regressing the prior data. Record the parameters of the prior regression. Calculate the precisions of the parameters of the prior. Next regress the sample data. Record the parameters of the sample regression. Calculate the precisions of the parameters. Once these two regressions are complete, combine the two regression equations by precision weighting the means.

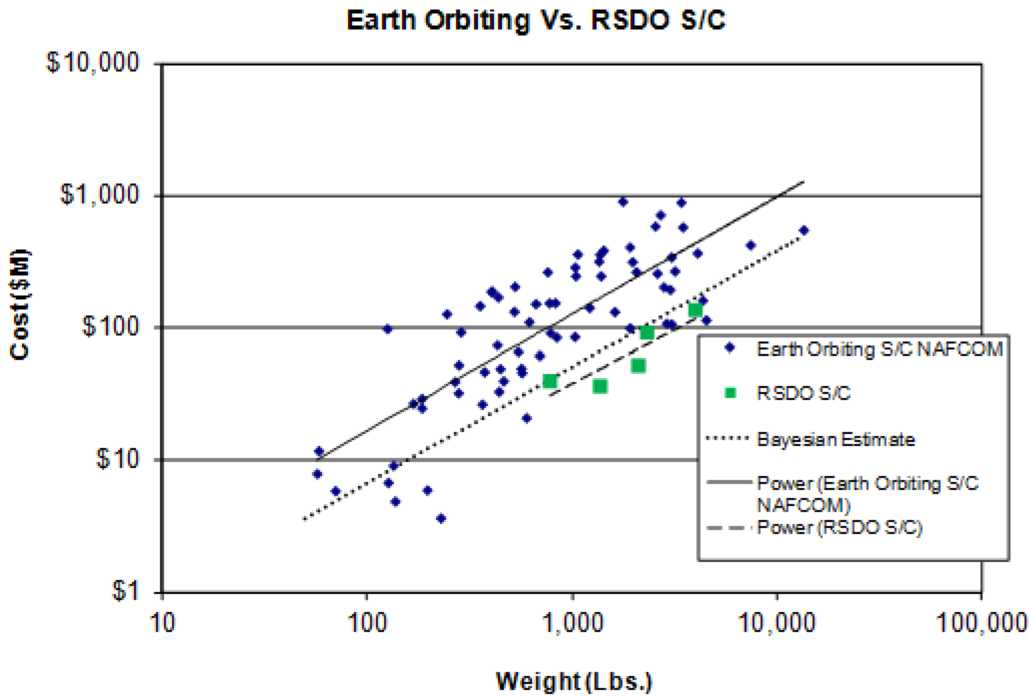

Figure 3: Comparison of Three Trendlines.

NAFCOM's First Pound Methodology

The NASA/Air Force Cost Model includes a method called "First Pound" CERs. These equations have the power form

$$\tilde{Y} = aW^b, \tag{51}$$

where $\tilde{Y}$ is the estimate of cost and W is dry spacecraft mass in pounds. The "First Pound" method is used for developing CERs with limited data. A slope b that varies by subsystem is based on prior experience. As documented in NAFCOM v2012, "NAFCOM subsystem hardware and instrument b-values were derived from analyses of some 100 weight-driven CERs taken from parametric models produced for MSFC, GSFC, JPL, and NASA HQ. In depth analyses also revealed that error bands for analogous estimating are very tight when NAFCOM b-values are used" [9].

As explained by Hamaker, the slope is assumed, then the a parameter is calculated by calibrating the data to one data point or to a collection of data points. Hamaker credits Frank Freiman with the inspiration for this approach, who implemented something similar in an early version of the PRICE cost model [4].

The slope values are provided in Table 3 by group and subsystem. In the table DDT&E is an acronym for Design, Development, Test and Evaluation, which is the nonrecurring cost. Flight Unit is the recurring cost.

The slope parameter in a power equation represents the economies of scale experienced. If there are no economies of scale, as the size of a project grows, the cost increases at a linear rate and $b = 1$. If there are diseconomies of scale then $b > 1$. In most cases, however, there are economies of scale and $b < 1$. The lower the value of b, the greater the economies of scale.

AsJoe Hamaker explained,

"The engineering judgment aspect of NAFCOM assumed slopes is based on the structural/mechanical content of the system versus the electronics/software content of the system. Systems that are more structural/mechanical are expected to demonstrate more economies of scale (i.e. have a lower slope) than systems with more electronics and software content. Software for example, is well known in the cost community to show diseconomies of scale (i.e. a CER slope of $b > 1.0$); that the larger the software project (in for example, lines of code) the more the cost per line of code. Larger weights in electronics systems implies more complexity generally, more software per unit of weight and more cross strapping and integration costs-all of which dampens out the economies of scale as the systems get larger. The assumed slopes are driven by considerations of how much structural/mechanical content each system has as compared to the system's electronics/software content" [4].

Among hardware item structural elements exhibit the greatest economies of scale while electronic elements exhibit the least. Propulsion elements, which are a mix of structures and electronics have economies of scale between these two extremes.

The data in Table 3 includes group and subsystem information. The spacecraft is the system. Major sub-elements are called subsystems, and include elements such as structures, reaction control, etc. A group is a collection of subsystems. For example the Avionics group is a collection of Command and Data Handling, Attitude Control, Range Safety, Electrical Power, and the Electrical Power Distribution, Regulation, and Control subsystems.

Some of these values come in handy when estimating cost for launch vehicles. For example, in 2007-2009, when I was developing cost estimates for the planned Ares launch vehicles, I had to develop estimates for range safety, crew accommodations, and environmental control and life support (ECLS) subsystems, among others. NASA has only developed a few launch vehicles, and very few human-rated systems that have required subsystems like crew accommodations and ECLS subsystems. The last of these was the International Space Station, and prior to that, the Space Shuttle.

As a notional example, suppose that you have one ECLS data point, with a dry weight equal to 7,000 pounds, and development cost equal to \$500 million. In Table 3, the b-value is equal to 0.65, plugging into the equation below, we have

$$500 = a(7000^{0.65}). \tag{52}$$

Solving this equation for a we find that

$$a = \frac{500}{7000^{0.65}} \approx 1.58 \tag{53}$$

Note that the slopes in Table 3 were verified based on log-transformed ordinary least squares regressions. The resulting CER is $\widetilde{\text{Cost}} = 1.58 \cdot \text{Weight}^{0.65}$. Thus as mentioned at the beginning of this paper, you only need one point to determine a trend line. Can we do even better? Can we develop a CER without any data at all? Yes we can! Consider the following which I called the "No Pound Methodology."

Subsystem/Group	DDT&E	Flight Unit
Antenna Subsystem	0.85	0.80
Aerospace Support Equipment	0.55	0.70
Attitude Control/Guidance and Navigation Subsystem	0.75	0.85
Avionics Group	0.90	0.80
Communications and Command and Data Handling Group	0.85	0.80
Communications Subsystem	0.85	0.80
Crew Accommodations Subsystem	0.55	0.70
Data Management Subsystem	0.85	0.80
Environmental Control and Life Support Subsystem	0.50	0.80
Electrical Power and Distribution Group	0.65	0.75
Electrical Power Subsystem	0.65	0.75
Instrumentation Display and Control Subsystem	0.85	0.80
Launch and Landing Safety	0.55	0.70
Liquid Rocket Engines Subsystem	0.30	0.50
Mechanisms Subsystem	0.55	0.70
Miscellaneous	0.50	0.70
Power Distribution and Control Subsystem	0.65	0.75
Propulsion Subsystem	0.55	0.60
Range Safety Subsystem	0.65	0.75
Reaction Control Subsystem	0.55	0.60
Separation Subsystem	0.50	0.85
Solid/Kick Motor Subsystem	0.50	0.30
Structures Subsystem	0.55	0.70
Structures/Mechanical Group	0.55	0.70
Thermal Control Subsystem	0.50	0.80
Thrust Vector Control Subsystem	0.55	0.60

Table 3: First Pound Slopes from NAFCOM.

To see this, we start in log-space, where we use the log-transformed OLS equation in mean difference form, i.e.,

$$\begin{aligned}
\widetilde{\ln(Y)} &= \alpha_{\tilde{X}} + \beta\left(\ln(X) - \frac{\sum_{i=1}^{n}\ln(X_i)}{n}\right) \qquad (54)\\
&= \frac{\sum_{i=1}^{n}\ln(Y_i)}{n} + \beta\left(\ln(X) - \frac{\sum_{i=1}^{n}\ln(X_i)}{n}\right)\\
&= \frac{1}{n}\ln\left(\prod_{i=1}^{n} Y_i\right) + \beta\left(\ln(X) - \frac{1}{n}\ln\left(\prod_{i=1}^{n} X_i\right)\right)\\
&= \ln\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}} + \beta\left(\ln(X) - \ln\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}\right)\\
&= \ln\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}} - \beta\ln\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}} + \beta\ln(X)\\
&= \ln\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}} - \ln\left(\prod_{i=1}^{n} X_i\right)^{\frac{\beta}{n}} + \ln(X)^{\beta}\\
&= \ln\left(\frac{\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}}{\left(\prod_{i=1}^{n} X_i\right)^{\frac{\beta}{n}}}\right) + \ln(X)^{\beta}\\
&= \ln\left(\frac{\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}}{\left(\prod_{i=1}^{n} X_i\right)^{\frac{\beta}{n}}}X^{\beta}\right).
\end{aligned}$$

Exponentiating both sides, we find

$$\tilde{Y} = \frac{\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}}{\left(\prod_{i=1}^{n} X_i\right)^{\frac{\beta}{n}}}X^{\beta}. \qquad (55)$$

Recall that $(\prod_{i=1}^{n} a_i)^{1/n}$ is the geometric mean. Therefore the term in the numerator is the geometric mean of the cost, and the term in the denominator is the geometric mean of the independent variable (such as weight) raised to the β. The geometric mean is distinct from the arithmetic mean, and is always less than or equal to the arithmetic mean. To apply this ”No-Pound” methodology you would need to apply insight or opinion to find the geometric mean of the cost, the geometric mean of the cost driver, and the economy of scale parameter, the slope.

We mention the ”No Pound” methodology as an aside, and now return to a description of the first-pound methodology and applying Bayes' Theorem to it.

The first-pound methodology bases the b-value entirely on the prior experience, and the a-value entirely on the sample data. No prior assumption for the a-value is applied. Denote the prior parameters by a_{Prior}, b_{Prior}, the sample parameters by a_{Sample}, b_{Sample}, and the posterior values by $a_{Posterior}$, $b_{Posterior}$. The first-pound methodology calculates the posterior values as

$$\begin{aligned}
a_{Posterior} &= a_{Sample} \qquad (56)\\
b_{Posterior} &= b_{Prior}. \qquad (57)
\end{aligned}$$

This is equivalent to a weighted average of the prior and sample information with a weight equal to 1.0 applied to the sample data for the a-value and a weight equal to 1.0 applied to the prior information for the b-value. The first-pound method in NAFCOM is not exactly the same as the approach we have derived but it is a Bayesian framework. Prior values for the slope are derived from experience and data, and this information is combined with sample data to provide an estimate based on experience and data. The first electronic version of NAFCOM in 1994 included the first-pound CER methodology. Thus, NAFCOM has included Bayesian statistical estimating methods for almost 20 years.

NAFCOM's calibration module is similar to the first pound method, but is an extension for multi-variable equations. Instead of assuming a b-value, the parameters for the built-in NAFCOM multi-variable CERs are used, but the intercept parameter (a-value) is calculated from the data, as with the first-pound method. The multi-variable CERs in NAFCOM have the form

$$\widetilde{\text{Cost}} = a \cdot \text{Weight}^{b_1} \cdot \text{New Design}^{b_2} \cdot \text{Technical}^{b_3} \cdot \text{Management}^{b_4} \cdot \text{Class}^{b_5} \tag{58}$$

In this equation, *New Design* is the percentage of new design for the subsystem ($0-100\%$). *Technical* cost drivers are determined for each subsystem and are weighted based upon their impact on the development or unit cost. *Management* cost drivers are based on new ways of doing business survey sponsored by the Space Systems Cost Analysis Group (SSCAG). The *Class* variable is a set of attribute (*dummy*) variables that are used to delineate data across mission classes: earth-orbiting, Planetary, Launch Vehicles, and Manned Launch Vehicles. See Smart's presentation on NAFCOM multivariate CERs for more information [10].

To apply the precision-weighted method to the first-pound CERs, we need an estimate of the variances of the b-values. Based on data from NAFCOM, these can be obtained by calculating average a-values for each mission class earth-orbiting, planetary, launch vehicle, or crewed system and then calculating the standard error and the sum of squares of the natural log of the weights. The results of these calculations are displayed in Table 4. Note that there is not enough data to calculate b-values for Range Safety or Separations subsystems.

In the absence of data, one way to calculate the standard deviation of the slopes is to estimate your confidence and express it in those terms. For example, if you are highly confident in your estimate of the slope parameter you may decide that means you are 90% confident that the actual slope will be within 5% of your estimate. For a normal distribution with mean μ and standard deviation σ the upper limit of a symmetric two-tailed 90% confidence interval is 20% higher than the mean, that is,

$$\mu + 1.645\sigma = 1.20\mu \tag{59}$$

from which it follows that

$$\sigma = \frac{0.20}{1.645\mu} \approx 0.12\mu. \tag{60}$$

Thus the coefficient of variation, which is the ratio of the standard deviation to the mean, is 12% in this case. See Table 5 for a list of suggested values for the coefficients

Subsystem/Group	DDT&E	Flight Unit
Antenna Subsystem	0.0338	0.0161
Aerospace Support Equipment	0.0698	0.0982
Attitude Control/Guidance and Navigation Subsystem	0.0120	0.0059
Avionics Group	0.0055	0.0044
Communications and Command and Data Handling Group	0.0053	0.0003
Communications Subsystem	0.0208	0.0141
Crew Accommodations Subsystem	0.0826	0.0565
Data Management Subsystem	0.0048	0.0025
Environmental Control and Life Support Subsystem	0.0439	0.2662
Electrical Power and Distribution Group	0.0064	0.0043
Electrical Power Subsystem	0.0878	0.0161
Instrumentation Display and Control Subsystem	0.1009	0.0665
Launch and Landing Safety	0.0960	0.0371
Liquid Rocket Engines Subsystem	0.1234	0.0483
Mechanisms Subsystem	0.0050	0.0167
Miscellaneous	0.0686	0.0784
Power Distribution and Control Subsystem	0.0106	0.0053
Propulsion Subsystem	0.2656	0.1730
Range Safety Subsystem	-	-
Reaction Control Subsystem	0.0144	0.0092
Separation Subsystem	-	-
Solid/Kick Motor Subsystem	0.0302	0.0105
Structures Subsystem	0.0064	0.0038
Structures/Mechanical Group	0.0029	0.0023
Thermal Control Subsystem	0.0055	0.0045
Thrust Vector Control Subsystem	0.7981	0.0234

Table 4: Calculate b-Value Variances for NAFCOM Data.

of variation based on the true mean being within 20% of the estimate with the stated confidence level.

Confidence Level	Coefficient of Variation
90%	12%
80%	16%
70%	19%
50%	30%
30%	52%
10%	159%

Table 5: Coefficient of variations for the confidence that the true mean is within 20% of the estimated mean.

For example, the Structures Subsystem in NAFCOM has a mean value equal to 0.55 for the b-value parameter of DDT&E, see Table 3. The calculated variance for 37 data points is 0.0064 found in Table 4. The standard deviation is the square root of this value, which is approximately 0.08. The calculated coefficient of variation is thus equal to

$$\frac{0.08}{0.55} \approx 14.5\%. \tag{61}$$

If I were 80% confident that the true value of the structures b-value is within 20% of 0.55 (i.e., between 0.44 and 0.66), then the coefficient of variation will equal 16% (see Table 5). This is similar to the calculated value.

As an example of applying the first pound priors to actual data, suppose we revisit the environmental control and life support (ECLS) subsystem. See Figure 4 for a scatter plot of the theoretical first unit cost and weight for six data points. The log-transformed ordinary least squares best fit is provided by the equation

$$\widetilde{\text{Cost}} = 0.4070Wt^{0.6300}. \tag{62}$$

The prior b-value for ECLS flight unit cost provided in Table 3 is 0.80. The first-pound methodology provides no prior for the a-value. Given no prior, the Bayesian method uses the calculated value as the a-value, and combines the b-values. The variance of the b-value from the regression is 0.1694 and thus the precision is $\frac{1}{0.1694} \approx 5.9032$. For the prior, the ECLS 0.8 b-value is based largely on electrical systems. The environmental control system is highly electrical, so I subjectively place high confidence in this value. I have 80% confidence that the true slope parameter is within 10% of the true value. From Table 5 this implies a coefficient of variation equal to 16%. Thus, the standard deviation of the b-value prior is $0.80 \cdot 0.16 = 0.128$ and the variance is the square of this value, i.e., $0.128^2 \approx 0.01638$. The precision is then $\frac{1}{0.01638} \approx 61.0352$. Therefore the precision-weighted b-value is

$$0.80 \cdot \frac{61.0352}{61.0352 + 5.9032} + 0.63 \cdot \frac{5.9032}{61.0352 + 5.9032} = 0.7850. \tag{63}$$

Thus, the adjusted equation combining prior experience and data is

$$\widetilde{\text{Cost}} = 0.4070Wt^{0.7850}. \tag{64}$$

The predictive equation, eq.(64) produced by the Bayesian analysis is very similar to the NAFCOM first-pound method. Recall that the first-pound methodology produces an a-value that is equal to the average a-value (in log-space). This is the same as the a-value produced by the regression since

$$\ln(\tilde{Y}) = \ln(a) + b\ln(X). \tag{65}$$

For each of the n data points, the a-value is calculated in log-space as

$$\ln(a) = \ln(\tilde{Y}) - b\ln(X). \tag{66}$$

The overall log-space a-value is the average of these a-values, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n}\ln(a_i) = \frac{1}{n}\sum_{i=1}^{n}\ln(Y_i) - \frac{b}{n}\sum_{i=1}^{n}\ln(X_i). \tag{67}$$

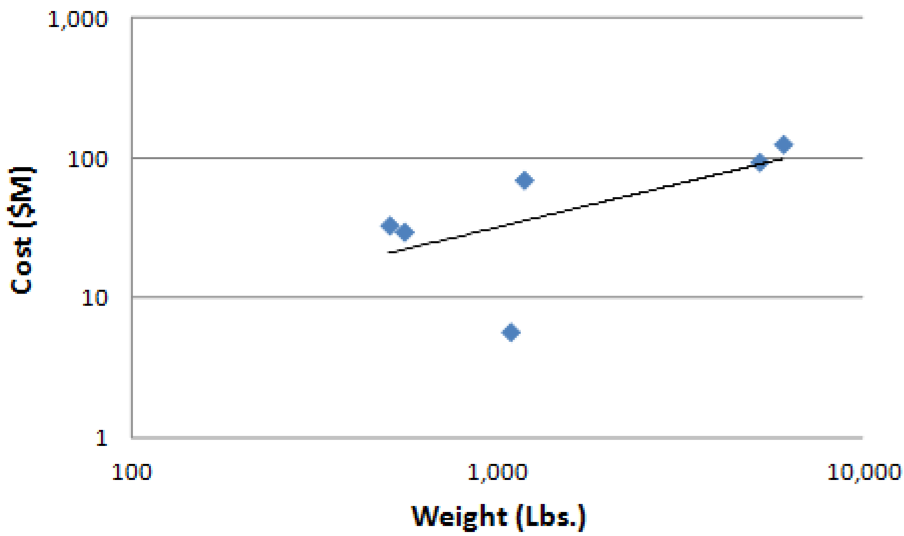

Figure 4: Environmental Control and Life Support Subsystem Data.

In the case $b = \beta$, this is the same as the calculation of the a-value from the normal equations in the regression. For small data sets we expect the overall b-value to be similar to the prior b-value. Thus, NAFCOM's first-pound methodology is very similar to the Bayesian approach.

However, the NAFCOM first-pound methodology and calibration modules can be enhanced by incorporating more aspects of the Bayesian approach. The first-pound methodology can be extended to incorporate prior information about the a-value as well. Neal Hulkower describes how Malcolm Gladwell's "thin-slicing" can be applied to cost estimating [2]. Hulkower suggests that experienced cost estimates can use prior experience to develop accurate cost estimates with limited information. I attended his presentation in 2008 where Dr. Hulkower asked the audience to provide estimates for a 300 kg small satellite, 10,000 lines of equivalent source lines of code for military mobile software in C++, a recent model F16, and an aircraft carrier. All audience responses were with 25% of the actual costs for the systems. Even with limited information it is possible to provide accurate estimates with limited information. Dr. Hulkower's experiment was an expert solicitation for the prior average cost of such a system. Providing prior information on average cost to develop a prior a-value can also be useful. And if a large data set is available, the average cost of a mission can be calculated from data to be used for the prior value [6].

Conclusion

The Bayesian framework involves taking prior experience, combining it with sample data, sometimes a small sample of data, and uses it to make accurate predictions of future events. Examples of successful applications in real world scenarios involved predicting election results, setting insurance premiums, and decoding encrypted messages, to name just a few.

This paper introduced Bayes' Theorem, and demonstrated how to apply it to regression analysis. An example of applying this method to prior experience with data, termed the hierarchical approach, was demonstrated. The idea of developing CER parameters based on logic and experience was discussed. A method for applying the Bayesian approach to this situation was presented, and an example of this approach to actual data was examined.

There are advantages to using this approach. This method enhances estimating costs when there is limited data. We showed one method for developing a CER with one data point, and introduced another for developing a CER without any data whatsoever! One particular advantage is that a small data set can have a narrow range, like we saw with the RSDO data. Estimating outside the range of the data is problematic for a small data set. For a small data set you have little confidence that the trend will hold outside the range of the data, or even that the trend itself will be valid. Combining a small data set with prior experience provides confidence in estimating outside the limited range of that set.

However, this method has some challenges. You must have some prior experience or information that can be applied to the problem. Without this you are left to frequency-based approaches. Nonetheless, there are ways to derive this information from logic, as discussed by Hamaker [4].

The examples in this paper have focused on log-transformed ordinary least squares. We did not discuss how to apply this method to other CER methods, such as the general error regression model (GERM), the interactively reweighted least squares (IRLS), or minimum unbiased percentage error (MUPE) approaches. The precision-weight approach does not depend on any particular underlying assumption. The only requirement is that the variance of the parameters are needed. The variances required for the GERM approach can be derived from a bootstrap approximation, as proposed by Book [1], the details are left for a future paper.

We did not discuss how to incorporate risk analysis in the Bayesian approach to parameters. The formula for the posterior variance was calculated for the ordinary least squares case, and this can also be applied to the log-transformed ordinary least squares method as well. We leave the details of this approach to a future paper as well.

References

[1] S. Book. Prediction Intervals for CER-Based Estimates. *Department of Defense Cost Analysis Symposium*, 2004.

[2] M. Gladwell. *Blink: The Power of Thinking Without Thinking*. Little, Brown, and Company, New York, New York, 2005.

[3] R. Guy. The Strong Law of Small Numbers. *American Math Monthly*, 95:697–712, 1988.

[4] J. Hamaker. A Monograph on CER Slopes. 2008. Unpublished white paper.

[5] P. Hoffman. *The Man Who Loved Only Numbers: The Story of Paul Erdos and the Serach for Mathematical Truth*. Hyperion, New York, New York, 1998.

[6] N. Hulkower. 'Thin-Slicing' for Costers: Estimating in the Blink of an Eye. *Huntsville Chapter of the Society of Cost Estimating and Analysis*, 2008.

[7] S. Klungman, H. Panjer, and G. Wilmott. *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., Hoboken, New Jersey, third edition, 2008.

[8] S. McGrayne. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant From Two Centuries of Controversy*. Yale University Press, New Haven, Connecticut, 2011.

[9] NASA. NASA/Air Force Cost Model (NAFCOM). version 2012.

[10] C. Smart. Multivariate CERs in NAFCOM. Cleveland, Ohio, 2006.