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

Christian Smart, Ph.D., CCEA
Director, Cost Estimating and Analysis
Missile Defense Agency

Introduction

- **When I was in college, my mathematics and economics professors were adamant in telling me that I needed at least two data points to define a trend**
  - **It turns out this is wrong**
  - **You can define a trend with only one data point, and even without any data**
- **A cost estimating relationship (CER), which is a mathematical equation that relates cost to one or more technical inputs, is a specific application of trend analysis which in cost estimating is called parametric analysis**
- **The purpose of this presentation is to discuss methods for applying parametric analysis to small data sets, including the case of one data point, and no data**

The Problem of Limited Data

- **A familiar theorem from statistics is the Law of Large Numbers**
  - **Sample mean converges to the expected value as the size of the sample increases**
- **Less familiar is the Law of Small Numbers**
  - **There are never enough small numbers to meet all the demands placed upon them**
- **Conducting statistical analysis with small data sets is difficult**
  - **However, such estimates have to be developed**
  - **For example NASA has not developed many launch vehicles, yet there is a need to understand how much a new launch vehicle will cost**
  - **There are few kill vehicles, but there is still a need to estimate the cost of developing a new kill vehicle**

One Answer: Bayesian Analysis

- **One way to approach these problems is to use Bayesian statistics**
  - **Bayesian statistics combines prior experience with sample data**
- **Bayesian statistics has been successfully applied to numerous disciplines (McGrayne 2011, Silver 2012)**
  - **In World War II to help crack the Enigma code used by the Germans, shortening the war**
  - **John Nash's (of *A Beautiful Mind* fame) equilibrium for games with partial or incomplete information**
  - **Insurance premium setting for property and casualty for the past 100 years**
  - **Hedge fund management on Wall Street**
  - **Nate Silver's election forecasts**

Application to Cost Analysis

- Cost estimating relationships (CERs) are important tool for cost estimators
- One limitation is that they require a significant amount of data
  - It is often the case that we have small amounts of data in cost estimating
- In this presentation we show how to apply Bayes' Theorem to regression-based CERs

Small Data Sets

- **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 is preferred**
- **However when applicable data are limited, leveraging prior experience can aid in the development of accurate estimates**

"Thin-Slicing"

- **The idea of applying significant prior experience with limited data has been termed "thin-slicing" by Malcolm Gladwell in his best-selling book *Blink* (Gladwell 2005)**
- **In his book Gladwell presents several examples of how experts can make accurate predictions with limited data**
- **For example, Gladwell presents 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 the same expert analyzes a couple for 15 minutes he can predict the same result with 90% accuracy**

Bayes' Theorem

- **The distribution of the model given values for the parameters is called the *model distribution***
- ***Prior* probabilities are assigned to the model parameters**
- **After observing data, a new distribution, called the *posterior* distribution, is developed for the parameters, using 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**

$$Pr(A|B) = \frac{Pr(A)Pr(B|A)}{Pr(B)}$$

Example Application (1 of 2)

- **Testing for illegal drug use**
  - **Many of you have had to take such a test as a condition of employment with the federal government or with a government contractor**
- **What is the probability that someone who fails a drug test is not a user of illegal drugs?**
- **Suppose that**
  - **95% of the population does not use illegal drugs**
  - **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**

Example Application (2 of 2)

- **In this case**
  - ***A*** **is the event that someone is not a user of illegal drugs**
  - ***B*** **is the event that someone test positive for illegal drugs**
  - **The complement of *A*, denoted *A'*, is the event that someone is a user of illegal drugs**
- **From the law of total probability**

$$\mathbf{Pr(B) = Pr}(\boldsymbol{B}|\boldsymbol{A})\,\mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A'})\,\mathbf{Pr}(\boldsymbol{A'})$$

- **Thus Bayes' Theorem in this case is equivalent to**

$$\mathbf{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A})\,\mathbf{Pr}(\boldsymbol{A})}{\mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A})\,\mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A'})\,\mathbf{Pr}(\boldsymbol{A'})}$$

- **Plugging in the appropriate values**

$$\mathbf{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\mathbf{0.02(0.95)}}{\mathbf{0.02(0.95) + 0.99(0.05)}} \approx \mathbf{27.7\%}$$

Forward Estimation (1 of 2)

- **The previous example 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?**

Forward Estimation (2 of 2)

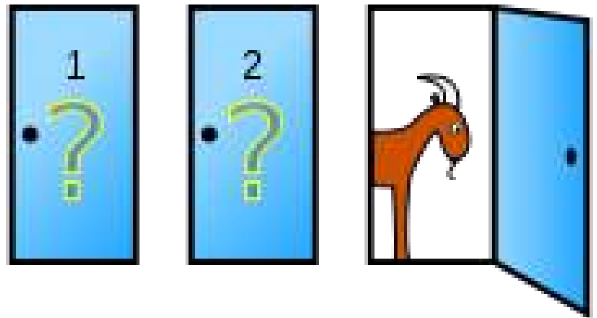
- **It's simply a matter of interpreting Baye's Theorem**
- ***Pr(A)* is the probability that we assign to an event before seeing the data**
  - **This is called the prior probability**
- ***Pr(A|B)* is the probability after we see the data**
  - **This is called the posterior probability**
- **Pr(B|A)/Pr(B) is the probability of the seeing these data given the hypothesis**
  - **This is the likelihood**
- **Bayes' Theorem can be re-stated as**

$$\textit{Posterior} \propto \textit{Prior*Likelihood}$$

Example 2: Monty Hall Problem (1 of 5)

- 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
- You pick a door and Monty, who knows what is behind the doors, then opens one of the other doors that has a goat behind it
- Suppose you pick door #1
  - Monty then opens door #3, showing you the goat behind it, and ask you if you want to pick door #2 instead
  - Is it to your advantage to switch your choice?

Monty Hall Problem (2 of 5)



- **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**
  - **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**
  - **Prior probability, before the third door is opened, that the car was behind door #1, which we denote $Pr(A_1)$, is 1/3. Also, $Pr(A_2)$ and $Pr(A_3)$ are also equal to 1/3.**

Monty Hall Problem (3 of 5)

- **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 you are shown the goat is behind door #3 is an impossible event is the car is behind door #3**
  - **$Pr(B|A_3) = 0$**
- **Since you picked door #1, Monty will open either door #2 or door #3, but not door #1**
- **If the car is actually behind door #2, it is a certainty that Monty will open door #3 and show you a goat.**
  - **$Pr(B|A_2) = 1$**
- **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**
  - **$Pr(B|A_2) = 1/2$**

Monty Hall Problem (4 of 5)

- **By Baye's Theorem**

$$Pr(A_1|B) = \frac{Pr(A_1)\,Pr(B|A_1)}{Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2) + Pr(A_3)\,Pr(B|A_3)}$$

- **Plugging in the probabilities from the previous chart**

$$Pr(A_1|B) = \frac{(1/3)(1/2)}{(1/3)(1/2) + (1/3)(1) + (1/3)(0)} = \frac{1/6}{1/6 + 1/3} = 1/3$$

$$Pr(A_2|B) = \frac{(1/3)(1)}{(1/3)(1/2) + (1/3)(1) + (1/3)(0)} = \frac{1/3}{1/6 + 1/3} = 2/3$$

$$Pr(A_3|B) = 0$$

Monty Hall Problem (5 of 5)

- **Thus you have a 1/3 of picking the car if you stick with you initial choice of door #1, but a 2/3 chance of picking the car if you switch doors**
  - **You should switch doors!**
- **Did you think there was no advantage to switching doors? If so you're not alone**
- **The Monty Hall problem created a flurry of controversy in the "Ask Marilyn" column in *Parade Magazine* in the early 1990s (Vos Savant 2012)**
- **Even the mathematician Paul Erdos was confused by the problem (Hofmann 1998)**

Continuous Version of Bayes' Theorem (1 of 2)

If the prior distribution is continuous, Bayes' Theorem is written as

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

where

$\pi(\theta)$ **is the prior density function**

$f(x|\theta)$ **is the conditional probability density function of the model**

$f(x_1,\ldots,x_n|\theta)$ **is the conditional joint density function of the data given** θ

Continuous Version of Bayes' Theorem (2 of 2)

$f(x_1,\ldots,x_n)$ **is the unconditional joint density function of the data**

$$f(x_1,\ldots,x_n) = \int \pi(\theta) f(x_1,\ldots,x_n|\theta)\,d\theta$$

$\pi(\theta | x_1,\ldots,x_n)$ **is the posterior density function, the revised density based on the data**

$f(x_{n+1} | x_1,\ldots,x_n)$ **is the *predictive density function*, the revised unconditional density based on the sample data:**

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

Application of Bayes' Theorem to OLS: Background

- **Consider ordinary least squares (OLS) CERs of the form**

$$Y = a + bX + \varepsilon$$

where *a* and *b* are parameters, and ε is the residual, or error, between the estimate and the actual

- **For the application of Baye's Theorem, re-write this in mean deviation form**

$$Y = \alpha_{\bar{x}} + \beta(X - \bar{X}) + \varepsilon$$

- **This form makes it easier to establish prior inputs for the intercept (it is now the average cost)**

Application of Bayes' Theorem to OLS: Likelihood Function (1 of 6)

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

$$L(\alpha_{\overline{x}}, \beta) \propto \prod_{i=1}^{n} exp\left(-\frac{1}{2\sigma^2}\left(Y_i - \left(\alpha_{\overline{x}} + \beta(X_i - \overline{X})\right)\right)^2\right)$$

$$= exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \left(\alpha_{\overline{x}} + \beta(X_i - \overline{X})\right)\right)^2\right)$$

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

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

Application of Bayes' Theorem to OLS: Likelihood Function (2 of 6)

which is equivalent to

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

which reduces to

$$SS_y - 2\beta SS_{xy} + n(\bar{Y} - \alpha_{\bar{x}})^2 + \beta^2 SS_x$$

since $\sum_{i=1}^{n}(Y_i - \bar{Y}) = 0$

and $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$

Application of Bayes' Theorem to OLS: Likelihood Function (3 of 6)

where

$$SS_y = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$SS_x = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$SS_{xy} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Application of Bayes' Theorem to OLS: Likelihood Function (4 of 6)

The joint likelihood of $\alpha_{\overline{X}}$ and β is proportional to

$$\boldsymbol{exp}\left(-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\overline{X}} - \overline{Y})^2\right)\right)$$

$$= \boldsymbol{exp}\left(-\frac{1}{2\sigma^2}\left(SS_y - 2\beta SS_{xy} + \beta^2 SS_x\right)\right)\boldsymbol{exp}\left(-\frac{1}{2\sigma^2}\left(n(\alpha_{\overline{X}} - \overline{Y})^2\right)\right)$$

$$= \boldsymbol{exp}\left(-\frac{1}{2\sigma^2/SS_x}\left(\beta^2 - 2\beta SS_{xy}/SS_x + SS_y/SS_x\right)\right)\boldsymbol{exp}\left(-\frac{1}{2\sigma^2/n}\left((\alpha_{\overline{X}} - \overline{Y})^2\right)\right)$$

Application of Bayes' Theorem to OLS: Likelihood Function (5 of 6)

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

$$\beta^2 - \frac{2\beta SS_{xy}}{SS_x} + \frac{SS_y}{SS_x} = \beta^2 - \frac{2\beta SS_{xy}}{SS_x} + \frac{SS_{xy}^2}{SS_x^2} - \frac{SS_{xy}^2}{SS_x^2} + \frac{SS_y}{SS_x}$$

$$= \beta^2 - \frac{2\beta SS_{xy}}{SS_x} + \frac{SS_{xy}^2}{SS_x^2} - \frac{SS_{xy}^2}{SS_x^2} + \frac{SS_y}{SS_x} = \left(\beta - \frac{SS_{xy}}{SS_x}\right)^2 + \text{constant}$$

which means that likelihood is proportional to

$$exp\left(-\frac{1}{2\sigma^2/SS_x}\left(\beta - \frac{SS_{xy}}{SS_x}\right)^2\right) exp\left(-\frac{1}{2\sigma^2/n}((\alpha_{\overline{X}} - \overline{Y})^2)\right)$$

$$= L(\beta)L(\alpha_{\overline{X}})$$

Application of Bayes' Theorem to OLS: Likelihood Function (6 of 6)

- **Thus the likelihoods for $\alpha_{\overline{X}}$ and β are independent**

- **We have derived that**

  $\frac{SS_{xy}}{SS_x} = B$ **, the least squares slope**

  $\overline{Y} = A_{\overline{X}}$**, the least squares estimate for the mean**

  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 $\alpha_{\overline{X}}$ follows a normal distribution with mean $\overline{Y}$ and variance $\frac{\sigma^2}{n}$

Application of Bayes' Theorem to OLS: The Posterior (1 of 2)

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

$$g\left(\alpha_{\overline{X}}, \beta \mid \left\{(x_1, y_1), \ldots, (x_n, y_n)\right\}\right) = g\left(\alpha_{\overline{X}}, \beta\right) \cdot sample\ likelihood\left(\alpha_{\overline{X}}, \beta\right)$$

- **If the prior density for β is normal with mean m_β and variance s_β^2 the posterior is normal with mean m_β' and variance $s_\beta'^2$ where**

$$m_\beta' = \frac{1/s_\beta^2}{1/s_\beta'^2} m_\beta + \frac{SS_x/\sigma^2}{1/s_\beta'^2} B$$

and

$$\frac{1}{s_\beta'^2} = \frac{1}{s_\beta^2} + \frac{SS_x}{\sigma^2}$$

Application of Bayes' Theorem to OLS: The Posterior (2 of 2)

- **If the prior density for $\alpha_{\overline{X}}$ is normal with mean $m_{\alpha_{\overline{X}}}$ and variance $s^2_{\alpha_{\overline{X}}}$ the posterior is normal with mean $m'_{\alpha_{\overline{X}}}$ and variance $s'^2_{\alpha_{\overline{X}}}$ where**

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

$$\frac{1}{{s'_{\alpha_{\overline{X}}}}^2} = \frac{1}{s^2_{\alpha_{\overline{X}}}} + \frac{n}{\sigma^2}$$

Application of Bayes' Theorem to OLS: The Predictive Equation

- **In the case of a normal likelihood with a normal prior, the mean of the predictive equation is equal to the mean of the posterior distribution, i.e.,**

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

Non-Informative Priors

- For a non-informative improper prior such as $\pi(\alpha_{\overline{X}}) = 1$ for all $\alpha_{\overline{X}}$
- By independence, β is calculated as in the normal distribution case, and $\alpha_{\overline{X}}$ is calculated as

$$exp\left(-\frac{1}{2\sigma^2/n}(\alpha_{\overline{x}} - \overline{Y})^2\right)$$

- which follows a normal distribution with mean equal to $\overline{Y}$ and variance equal to σ^2/n
  - This is equivalent to the sample mean of $\alpha_{\overline{X}}$ and the variance of the sample mean
- Thus in the case where we only information about the slope, the sample mean of actual data is used for $\alpha_{\overline{X}}$

Estimating with Precisions

- **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***
- **We next generalize this result to the linear combination of any two estimates that are independent and unbiased**

The Precision Theorem (1 of 4)

- **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**
- **Proof**
  - **Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be two independent, unbiased estimators of a random variable θ**
    - **By definition** $E(\tilde{\theta}_1) = E(\tilde{\theta}_2) = \theta$
  - **Let w and $1-w$ denote the weights**
  - **The weighted average is unbiased since**

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

The Precision Theorem (2 of 4)

- **Since the two estimators are independent the variance of the weighted average is**

$$Var\left(w\tilde{\theta}_1 + (1-w)\tilde{\theta}_2\right) = w^2 Var\left(\tilde{\theta}_1\right) + (1-w)^2 Var\left(\tilde{\theta}_2\right)$$

- **To determine the weights that minimize the variance, define**

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

- **Take the first derivative of this function and set equal to zero**

$$\phi'(w) = 2wVar\left(\tilde{\theta}_1\right) - 2(1-w)Var\left(\tilde{\theta}_2\right)$$

$$= 2wVar\left(\tilde{\theta}_1\right) + 2wVar\left(\tilde{\theta}_2\right) - 2Var\left(\tilde{\theta}_2\right) = 0$$

The Precision Theorem (3 of 4)

- **Note that the second derivative is**

$$\phi''(w) = 2Var(\tilde{\theta}_1) + 2Var(\tilde{\theta}_2)$$

ensuring that the solution will be a minimum

- **The solution to this equation is**

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

The Precision Theorem (4 of 4)

- **Multiplying both the numerator and the denominator by**

$$1/\left(Var(\tilde{\theta}_1) \cdot Var(\tilde{\theta}_2)\right)$$

yields

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

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

which completes the proof

Precision-Weighting Rule

- **The Precision-Weighting Rule for combining two parametric estimates**
  - **Given two independent and unbiased estimates $\tilde{\theta}_1$ and $\tilde{\theta}_2$ with precisions $\rho_1 = 1/Var(\tilde{\theta}_1)$ and $\rho_2 = 1/Var(\tilde{\theta}_2)$ the minimum variance estimate is provided by**

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

Advantages of the Rule

- **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**

$$MSE_{\tilde{\theta}}(\theta) = E\left[\left(\tilde{\theta} - \theta\right)^2 | \theta\right]$$

  - **In general, the lower the mean squared error, the better the estimator**
    - **The mean square error is widely accepted as a measure of accuracy**
    - **You may be familiar with this as the "least squares criterion" from linear regression**
  - **Thus the precision-weighted approach which minimizes the mean square error, has optimal properties**

Examples

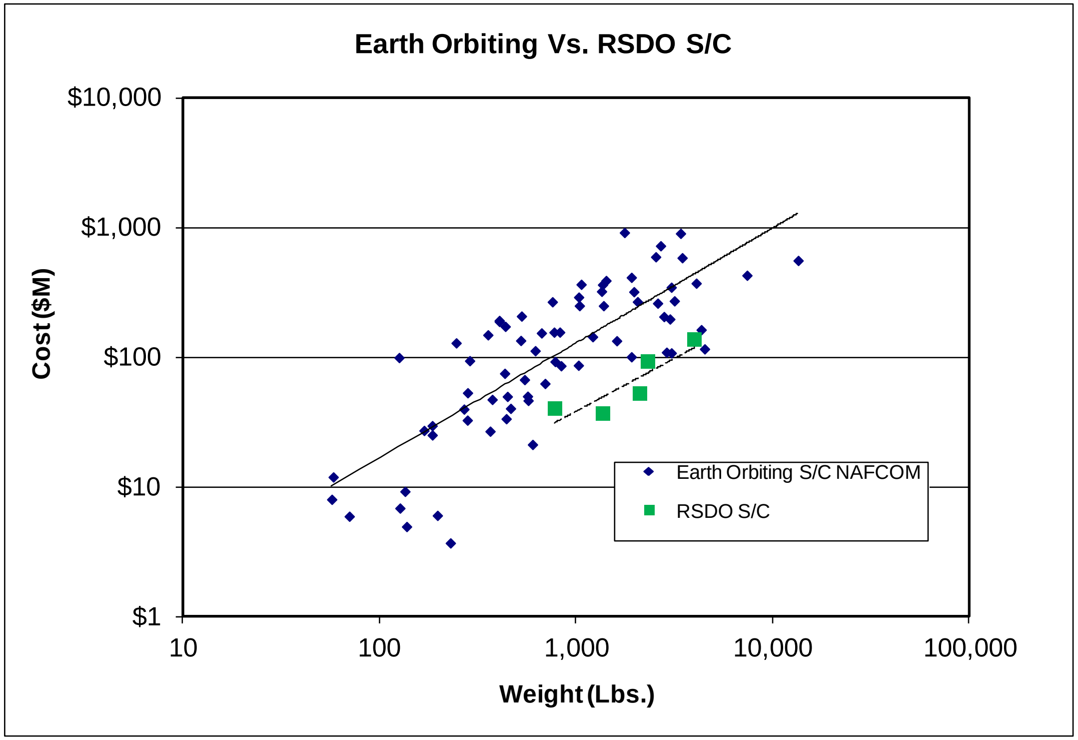
- **The remainder of this presentation focuses on two examples**
  - **One considers the hierarchical approach**
    - **Generic information is used as the prior, and specific information is used as the sample data**
  - **The second focuses on developing the prior based on experience and logic**

Example: Goddard's RSDO

- 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/

Comparison to Other Spacecraft (1 of 2)

- **Data on earth orbiting spacecraft is plentiful while data for RSDO is 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 the NASA/Air Force Cost Model (NAFCOM) and 5 RSDO missions**



Comparison to Other Spacecraft (2 of 2)

- **Power equations of the form $\widetilde{Y} = aW^b$ were fit to both data sets**
- **The b-value which we mentioned is a measure of the economy of scale, is .89 for the NAFCOM data, and 0.81 for the RSDO data 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**

Hierarchical Approach

- **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**
  - **Nate Silver used this method to develop accurate election forecasts in small population areas and areas with little data**
  - **This is also the approach that actuaries use when setting premiums for insurances with little data**

Transforming the Data (1 of 2)

- **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 thus 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 $\widetilde{Y} = aW^{b}$ to log space by applying the natural log function to each side, i.e.**

$$ln\widetilde{Y} = ln\left(aW^{b}\right) = ln(a) + b \cdot ln(W)$$

Transforming the Data (2 of 2)

- **In this case $\alpha_{\overline{X}} = a$ and $\beta = b$**
- **The average *Y*-value is the average of the natural log of the cost values**
- **Once the data are transformed, ordinary least squares regression is applied to both the NAFCOM data and to the RSDO data**
- **Data are available for both data sets - opinion is not used**
- **The precisions used in calculating the combined equation are calculated from the regression statistics**
- **We regress the natural log of the cost against the difference between the natural log of the weight and the mean of the natural log of the weight. That is, the dependent variable is *ln(Cost)* and the independent variable is**

$$ln(X) - \sum_{i=1}^{n} \frac{lnX_i}{n}$$

Obtaining the Variances

- From the regressions we need the values of the parameters as well as the variances of the parameters
- Statistical software package provide both the parameter and their variances as outputs
- Using the Data Analysis add-in in Excel, the Summary Output table provides these values

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.79439689
R Square	0.63106642
Adjusted R Square	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	48.21134863	1.95E-55	4.418074125
X Variable 1	0.88578231	0.080949568	10.942397	8.27E-17	0.724333491

Mean and variance of the parameters

Combining the Parameters (1 of 2)

Parameter	NAFCOM Mean	NAFCOM Variance	NAFCOM Precision	RSDO Mean	RSDO Variance	RSDO Precision	Combined Mean
$\alpha_{\overline{X}}$	*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*

- **The mean of each parameter is the value calculated by the regression and 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**
- **For the intercept the relative precision weights for the intercept are**

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

for the NAFCOM data, and $1-0.6870=0.3130$ **for the RSDO data**

Combining the Parameters (2 of 2)

- **For the slope the relative precision weights are**

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

for the NAFCOM data, and $1-0.9109=0.0891$ **for the RSDO data**

- **The combined intercept is**

$$0.6870\cdot 4.6087+0.3130\cdot 4.1359\approx 4.4607$$

- **The combined slope is**

$$0.9109\cdot 0.8858+0.0891\cdot 8144\approx 0.8794$$

The Predictive Equation

- **The predictive equation in log-space is**

$$\widetilde{\boldsymbol{Y}} = \mathbf{4.4607 + 0.8794}(\boldsymbol{X} - \bar{\boldsymbol{X}})$$

- **The only remaining question is what 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**

$$\widetilde{\boldsymbol{Y}} = \mathbf{4.4607 + 0.8794}(\boldsymbol{X} - \bar{\boldsymbol{X}}) = \mathbf{4.4607 + 0.8794}(\boldsymbol{X} - \mathbf{7.5161})$$
$$= \mathbf{-2.1491 + 0.8794}\boldsymbol{X}$$

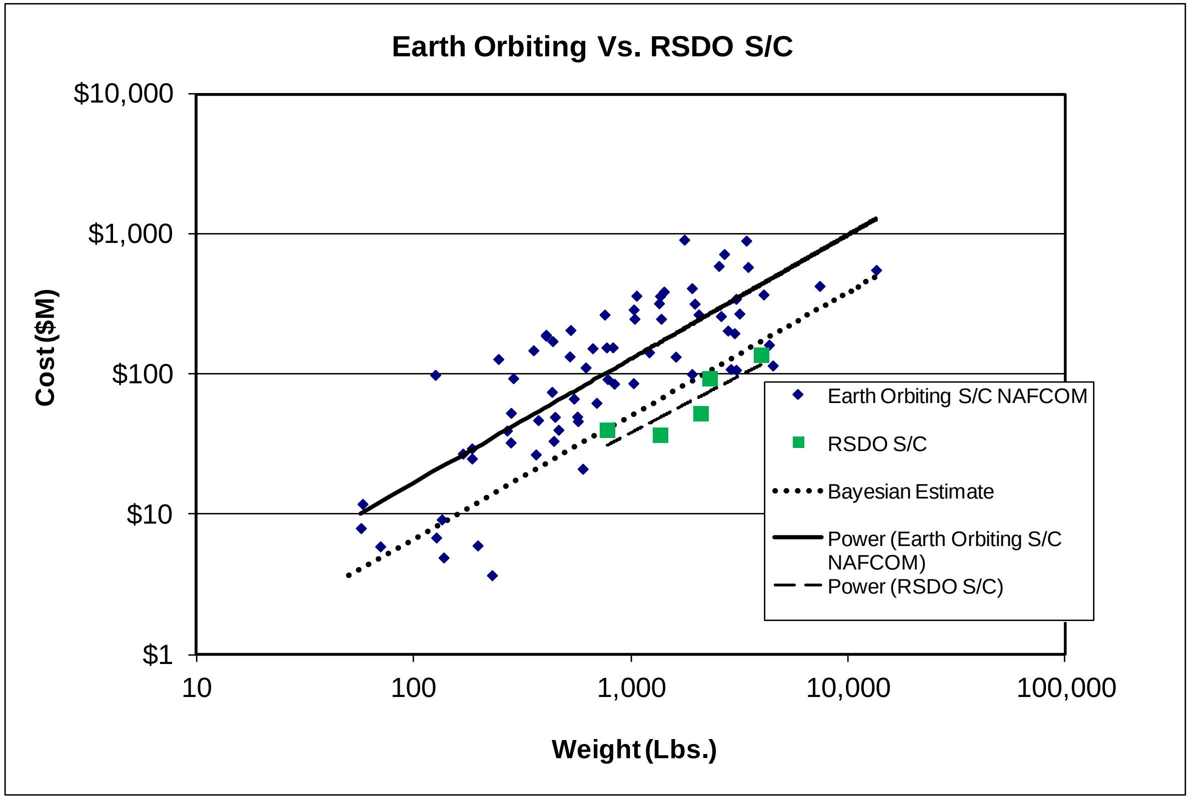
Transforming the Equation

- **This equation is in log-space, that is**

$$ln(\widetilde{Cost}) = -2.1491 + 0.8794 ln(Wt)$$

- **In linear space, this is equivalent to**

$$\widetilde{Cost} = 0.1166\, Wt^{0.8794}$$



Applying the Predictive Equation

- 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 dry weight is 3,280 lbs.
- Using the Bayesian equation the predicted cost is

$$\widetilde{Cost} = 0.1166 \cdot 3280^{0.8794} \approx \$144\ Million$$

which is 20% below the actual cost, which is approximately $180 million in normalized $

- The RSDO data alone predicts a cost equal to $100 Million
  - 44% below the actual cost
- The Earth-Orbiting data alone predicts a cost equal to $368 million
  - more than double the actual cost
- While this is only one data point, this seems promising

Range of the Data

- **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**

Summary of the Hierarchical Approach

- **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 steps are complete, combine the two regression equations by precision weighting the means of the parameters**

NAFCOM's First Pound Methodology (1 of 2)

- **The NASA/Air Force Cost Model includes a method called "First Pound" CERs**
- **These equations have the power form** $\tilde{Y} = aW^b$ **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 (NASA, 2012), "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. Further, actual regression historical models. In depth analyses also revealed that error bands for analogous estimating are very tight when NAFCOM b-values are used."**

NAFCOM's First Pound Methodology (2 of 2)

- **The slope is assumed, and then the *a* parameter is calculated by calibrating the data to one data point or to a collection of data points (Hamaker 2008)**
- **As explained by Joe Hamaker (Hamaker 2008), "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)—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.".**

NAFCOM's First Pound Slopes (1 of 2)

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

NAFCOM's First Pound Slopes (2 of 2)

- **In the table, DDT&E is an acronym for Design, Development, Test, and Evaluation**
  - **Same as RDT&E or Non-recurring**
- **The table 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**

First-Pound Methodology Example

- **As a notional example, suppose that you have one environmental control and life support (ECLS) data point, with dry weight equal to 7,000 pounds, and development cost equal to $500 million. In the table the *b*-value is equal to 0.65, which means that**

$$\mathbf{500 = a\left(7000^{0.65}\right)}$$

- **Solving this equation for *a* we find that**

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

- **The resulting CER is**

$$\widetilde{Cost} = 1.58 \cdot Weight^{0.65}$$

"No Pound" Methodology (1 of 3)

- **If we can develop a CER with only one data point, can we go one step further and develop a CER based on no data at all?**
  - **The answer if yes we can!**

- **To see what information we need to apply this method, start with the first pound methodology, and assume we have a prior value for β**
- **We start in log space**

$$\widetilde{ln(Y)} = \alpha_{\overline{X}} + \beta\left(ln(X) - \frac{\sum_{i=1}^{n} ln(X_i)}{n}\right)$$

$$= \frac{\sum_{i=1}^{n} ln(Y_i)}{n} + \beta\left(ln(X) - \frac{\sum_{i=1}^{n} ln(X_i)}{n}\right)$$

“No Pound” Methodology (2 of 3)

$$= \frac{1}{n} ln\left(\prod_{i=1}^{n} Y_i\right) + \beta\left(ln(X) - \frac{1}{n} ln\left(\prod_{i=1}^{n} X_i\right)\right)$$

$$= ln\left(\prod_{i=1}^{n} Y_i\right)^{1/n} + \beta\left(ln(X) - ln\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right)$$

$$= ln\left(\prod_{i=1}^{n} Y_i\right)^{1/n} - \beta ln\left(\prod_{i=1}^{n} X_i\right)^{1/n} + \beta \cdot ln(X)$$

$$= ln\left(\prod_{i=1}^{n} Y_i\right)^{1/n} - ln\left(\prod_{i=1}^{n} X_i\right)^{\beta/n} + ln\left(X^{\beta}\right)$$

$$= ln\left(\frac{\left(\prod_{i=1}^{n} Y_i\right)^{1/n}}{\left(\prod_{i=1}^{n} X_i\right)^{\beta/n}}\right) + ln\left(X^{\beta}\right)$$

$$= ln\left(\frac{\left(\prod_{i=1}^{n} Y_i\right)^{1/n}}{\left(\prod_{i=1}^{n} X_i\right)^{\beta/n}} X^{\beta}\right)$$

"No Pound" Methodology (3 of 3)

- Exponentiating both sides yields

$$\widetilde{Y} = \frac{\left(\prod_{i=1}^{n} Y_i\right)^{1/n}}{\left(\prod_{i=1}^{n} X_i\right)^{\beta/n}} X^{\beta}$$

- The term

$$\left(\prod_{i=1}^{n} Y_i\right)^{1/n}$$

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

First-Pound Methodology and Bayes (1 of 2)

- 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 parameters by $a_{posterior}$, $b_{posterior}$
- The first-pound methodology calculates the posterior values as

$$a_{posterior} = a_{sample}$$
$$b_{posterior} = b_{prior}$$

- This is equivalent to a weighted average of the prior and sample information with a weight equal to 1 applied to the sample data for the a-value, and a weight equal to 1 applied to the prior information for the b-value

First-Pound Methodology and Bayes (2 of 2)

- 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
  - NAFCOM has included Bayesian statistical estimating methods for almost 20 years

NAFCOM's Calibration Module

- NAFCOM's calibration module is similar to the first pound method, but is an extension for multi-variable equations
- Instead of assuming a value for the b-value, the parameters for the built-in NAFCOM multivariable 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

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

- "New Design" is the percentage of new design for the subsystem (0-100%)
- "Technical" cost drivers were determined for each subsystem and were weighted based upon their impact on the development or unit cost
- "Management" cost drivers based on a 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

Precision-Weighting First Pound CERs

- **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 calculated 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**
- **See the table on the next page for these data**

Variances of the b-Values

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

*** There is not enough data for Range Safety or Separation to calculate variance**

Subjective Method for b-Value Variance

- **One way to calculate the standard deviation of the slopes without data 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,**

$$\mu + 1.645\sigma = 1.20\mu$$

from which it follows that

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

  - **Thus the coefficient of variation, which is the ratio of the standard deviation to the mean, is 12%**

Coefficient of Variations Based on Opinion

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

- **The structures subsystem in NAFCOM has a mean value equal to 0.55 for the b-value parameter of DDT&E**
- **The calculated variance for 37 data points is 0.0064, so the standard deviation is approximately 0.08**
- **The calculated coefficient of variation is thus equal to**

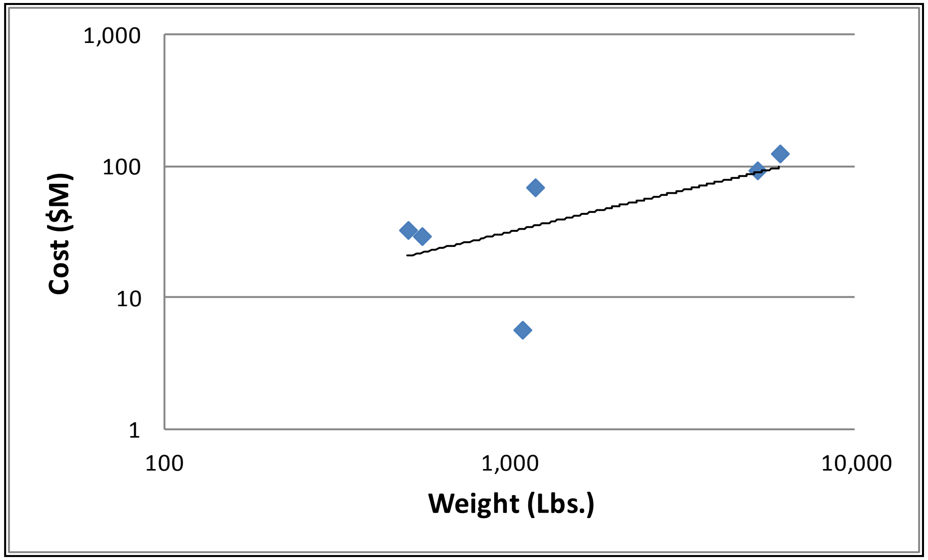
$$\frac{\mathbf{0.08}}{\mathbf{0.55}} \approx \mathbf{14.5\%}$$

- **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%**

Example

- As an example of applying the first pound priors to actual data, suppose we re-visit the environmental control and life support (ECLS) subsystem
- The log-transformed ordinary least squares best fit is provided by the equation

$$\widetilde{Cost} = 0.4070\, Wt^{0.6300}$$



Precision-Weighting the Means (1 of 2)

- **The prior b-value for ECLS flight unit cost provided 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**

Precision-Weighting the Means (2 of 2)

- **I have 80% confidence that the true slope parameter is within 10% of the true value which implies a coefficient of variation equal to 16%**
- **Thus the standard deviation of the b-value prior is equal to** $0.80 \cdot 0.16 \approx 0.128$ **and the variance is approximately 0.01638, which means the precision is**

$$\frac{1}{0.01638} \approx 61.0352$$

- **The precision-weighted b-value is thus**

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

- **Thus the adjusted equation combining prior experience and data is**

$$\widetilde{Cost} = 0.4070\, Wt^{0.7850}$$

Similarity Between Bayesian and First Pound Methods (1 of 2)

- **The predictive equation produced by the Bayesian analysis is very similar to the NAFCOM first-pound method**
- **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**

$$\ln(\tilde{Y}) = \ln(a) + b \cdot \ln(X)$$

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

$$\ln(a) = \ln(\tilde{Y}) - b \cdot \ln(X)$$

- **The overall log-space a-value is the average of these a-values**

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

Similarity Between Bayesian and First Pound Methods (2 of 2)

- **In the case 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**
- **Not only is the first-pound method a Bayesian framework but it can be considered as an approximation of the Bayesian method**

Enhancing the First-Pound Methodology

- **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 (Gladwell 2005, Hulkower 2008)**
  - **Hulkower suggests that experienced cost estimates can use prior experience to develop accurate cost estimates with limited information**

Summary (1 of 2)

- **The Bayesian framework involves taking prior experience, combining it with sample data, and uses it to make accurate predictions of future events**
  - **Examples include predicting election results, setting insurance premiums, and decoding encrypted messages**
- **This presentation 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 presented**
  - **The idea of developing CER parameters based on logic and experience was discussed**
  - **Method for applying the Bayesian approach to this situation was presented, and an example of this approach to actual data was discussed**

Summary (2 of 2)

- Advantages to using this approach
  - Enhances the ability to estimate costs for small data sets
  - Combining a small data set with prior experience provides confidence in estimating outside the limited range of a small data set
- Challenge
  - You must have some prior experience or information that can be applied to the problem
    - Without this you are left to frequency-based approaches
    - However, there are ways to derive this information from logic, as discussed by Hamaker (2008)

Future Work

- **We only discussed the application to ordinary least squares and log-transformed ordinary least squares**
  - **We did not discuss other methods, such as MUPE or the General Error Regression Model (GERM) framework**
  - **Can apply the precision-weighting rule to any CERs, just need to be able to calculate the variance**
  - **For GERM can calculate the variance of the parameters using the bootstrap method**
- **We did not explicitly address risk analysis, although we did derive the posteriors for the variances of the parameters, which can be used to derived prediction intervals**

References

1. Bolstad, W.M., *Introduction to Bayesian Statistics*, 2nd Edition, John Wiley & Sons, Inc., 2007, Hoboken, New Jersey.
2. Book, S.A., "Prediction Intervals for CER-Based Estimates", presented at the 37th Department of Defense Cost Analysis Symposium, Williamsburg VA, 2004.
3. Gladwell, M., *Blink: The Power of Thinking Without Thinking*, Little, Brown, and Company, 2005, New York, New York.
4. Guy, R. K. "The Strong Law of Small Numbers." Amer. Math. Monthly 95, 697-712, 1988.
5. Hamaker, J., "A Monograph on CER Slopes," unpublished white paper, 2008.
6. Hoffman, P., *The Man Who Loved Only Numbers: The Story of Paul Erdos and the Search for Mathematical Truth*, Hyperion, 1998, New York, New York.
7. Hulkower, N.D. "'Thin-Slicing' for Costers: Estimating in the Blink of an Eye," presented at the Huntsville Chapter of the Society of Cost Estimating and Analysis, 2008.
8. Klugman, S.A., H.J. Panjer, and G.E. Wilmott, Loss Models: From Data to Decisions, 3rd Edition, John Wiley & Sons, Inc., 2008, Hoboken, New Jersey.
9. McGrayne, S.B., *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, 2011, New Haven, Connecticut.
10. NASA, NASA/Air Force Cost Model (NAFCOM), version 2012.
11. Silver, N., *The Signal and the Noise – Why So Many Predictions Fail, but Some Don't*, The Penguin Press, 2012, New York, New York.
12. Smart, C.B., "Multivariate CERs in NAFCOM," presented at the NASA Cost Symposium, June 2006, Cleveland, Ohio.
13. Vos Savant, M., http://marilynvossavant.com/game-show-problem/