Bayesian Parametrics: Extending the Gaussian Model

GALORATH

Christian Smart, Ph.D., CCEA
David Jo, PCEA

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

Law of Large Numbers

A familiar theorem from statistics

-Sample mean converges to the expected value as the size of the sample increases

Law of Small Numbers

A less familiar theorem

-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

WHAT CAN BE DONE?

COLLECT MORE DATA

WORK WITH OTHER GOV'T AGENCIES (E.G., AIR FORCE, COMMERCIAL COMPANIES)

CAN BE DIFFICULT TO DO

IMPUTE

USE IMPUTATION TO FILL IN MISSING DATA

NEED TO BE CAREFUL TO CONSERVE CORRELATION STRUCTURES

GO LOWER

ESTIMATE AT THE COMPONENT LEVEL INSTEAD OF THE SUBSYSTEM LEVEL

MORE NOISE

BAYES

USE TECHNIQUES DESIGNED TO ESTIMATE WITH LIMITED DATA

SUBJECTIVE

BAYES’ THEOREM

BAYES' THEOREM

CONDITIONAL PROBABILITY

NUMEROUS APPLICATIONS

1 **ENIGMA**

Bayesian techniques were used to help crack the Enigma code in Word War II, shortening the war

2 **PROPERTY and CASUALTY INSURANCE**

Used for over a century to set premiums when there is limited data

3 **HEDGE FUND MANAGEMENT**

Also used in election forecasting and game theory

CONDTIONAL PROBABILITY:

$$Pr\left(A \mid B\right)$$

BAYES' THEOREM:

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

Bayesian techniques have been successfully used in a variety of applications - their use is no mere academic exercise in fancy statistics. They have *skin in the game*.

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

DATA – HOW SMALL IS SMALL?

FITTING TO NOISE

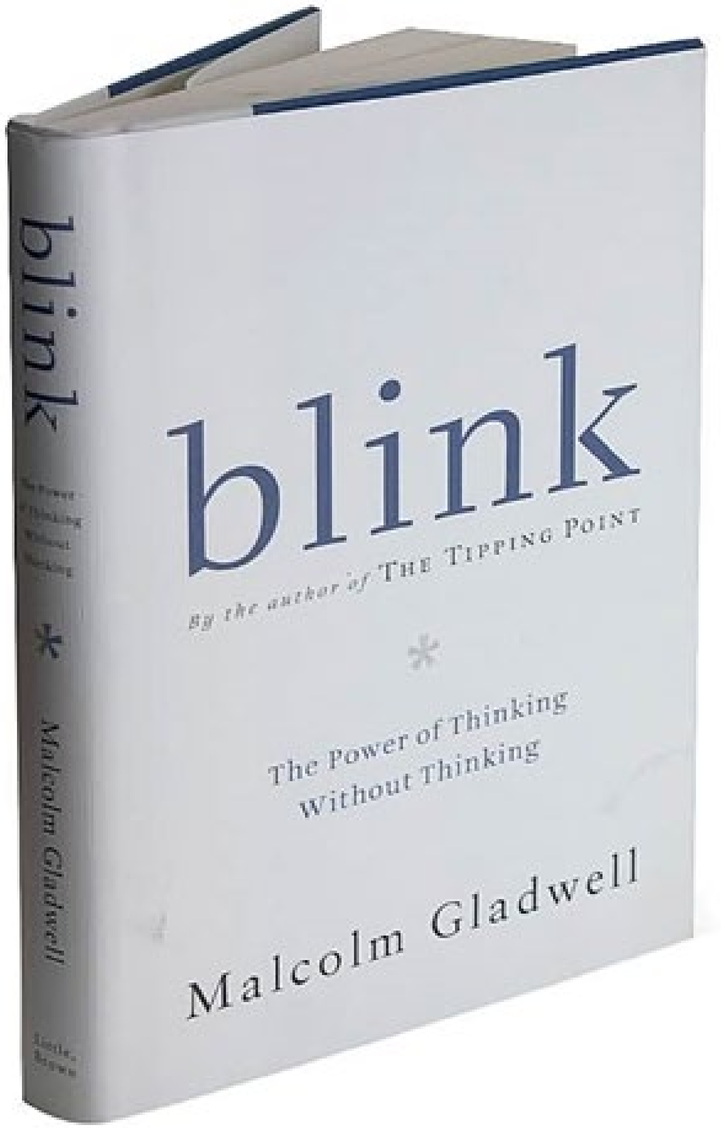
- SMALL DATA SETS ARE THE IDEAL SETTING FOR THE APPLICATION OF BAYESIAN TECHNIQUES FOR COST ANALYSIS
- GIVEN LARGE DATA SETS THAT ARE DIRECTLY APPLICABLE TO THE PROBLEM AT HAND A STRAIGHTFORWARD REGRESSION ANALYSIS IS PREFERRED

RULE OF THUMB

- 50 DATA POINTS PLUS 10 DATA POINTS PLUS 10 DATA POINTS FOR EVERY ADDITIONAL PARAMETER
- HOWEVER WHEN APPLICABLE DATA ARE LIMITED, LEVERAGING PRIOR EXPERIENCE CAN AID IN THE DEVELOPMENT OF ACCURATE ESTIMATES

BASIS

- THE RULE OF THUMB IS BASED ON SIMULATION STUDIES OF FITTING REGRESSIONS TO RANDOMLY GENERATED DATA



THIN SLICING

AN INTUITIVE FORM

INTRODUCED IN *BLINK*

The best-selling book *Blink* introduced the notion of thin-slicing, compiled numerous examples of experts who can make accurate predictions with limited data because of their deep experience

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

TAKEAWAY

Judgment is important in estimating – the Bayesian approach allows you to incorporate it in a mathematical framework

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

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Bayesian methods date back to the 18th century – computational advances and algorithms developed in the 1990s have made the techniques more useful

EXAMPLE 1: DRUG TESTING

LAW OF TOTAL PROBABILITY:

$$\Pr(\mathrm{B}) = \Pr(B|A)\Pr(A) + \Pr(B|A')\Pr(A')$$

BAYES' THEOREM:

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

ANSWER:

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

1

QUESTION

What is the probability that someone who fails a drug test is not a drug user?

2

ASSUMPTIONS

95% of the population are non-users

If someone is a drug user, it returns a positive result 99% of the time

If someone is not a user, it returns a positive result 2% of the time

3

DEFINING TERMS

A = Event that someone is NOT a drug user

B = Event that someone tests positive for drugs

4

COMPLEMENT

A' = Event that someone uses drugs

B' = Event that someone tests negative for drugs

FORWARD ESTIMATION WITH BAYES

A MATTER OF INTERPRETATION

PREVIOUS EXAMPLE

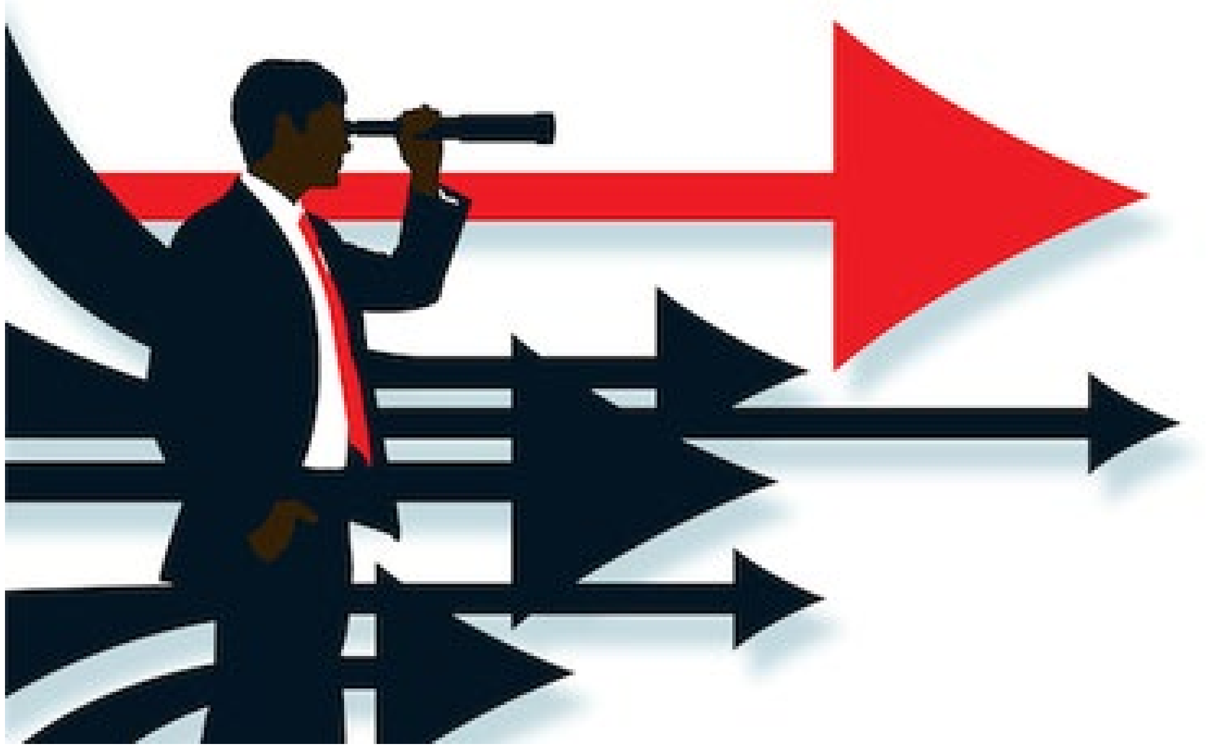
Inverse probability – statistical detective work, where you see the result of an action and infer whether someone is guilty or innocent based

MORE TYPICAL

Forward estimation is more common – you have some evidence from the past and want to use it to predict a future event – such as the final cost or schedule of a project.

RESTATEMENT OF BAYES' THEOREM

$$Posterior \propto Prior \cdot Likelihood$$



EXAMPLE 2: MONTY HALL PROBLEM

BAYES' THEOREM:

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

POSTERIOR PROBABILITIES:

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

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

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

ANSWER: YOU SHOULD SWITCH!

1

QUESTION

There are three doors. Behind one is a car. Behind the other two cars are goats. You pick a door. Monty opens a different door and shows you a goat, and offers you the chance to switch

2

DEFINING TERMS

Let A_i denote the event that the car is behind the i^{th} door, WLOG assume:

1. You pick door #1
2. You are shown a goat behind door #3, define this as event B

3

PRIOR PROBABILITY

Initial assumption is that

$P(A_1) = P(A_2) = P(A_3) = 1/3$

4

CONDITIONAL PROBABILITY

$P(B|A_3) = 0$

$P(B|A_2) = 1$

$P(B|A_1) = 1/2$

APPLICATION OF BAYES’ THEOREM TO REGRESSION

THE CONTINUOUS VERSION OF BAYES' THEOREM

- **The previous discussion has been for discrete distributions – for regression analysis we consider continuous distributions**
- **If the prior distribution is continuous, Bayes' Theorem is written as**

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

where

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

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

$f(x_1,\dots,x_n|\theta)$ **is the conditional joint density function of the model**

$f(x_1,\dots,x_n) = \int \pi(\theta)f(x_1,\dots,x_n|\theta)d\theta$ **is the unconditional joint density function**

THE CONTINUOUS VERSION OF BAYES' THEOREM (2)

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

$f(x_{n+1}|x_1,\ldots,x_n) = \int f(x_{n+1}|\theta)\pi(\theta|x_1,\ldots,x_n)d\theta$ **is the predictive density function, the revised unconditional density based on the data**

BAYESIAN OLS – SETUP

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

$$\boldsymbol{Y = a + bX + \varepsilon}$$

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**

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

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

BAYESIAN OLS – RESULTS

- **We're going to skip the gory details – the interested reader is welcome to consult Christian Smart's 2014 ICEAA paper for details**
- **Assumptions:**
  - **The likelihood of the slope β follows a Gaussian distribution with mean *B* and variance $\frac{\sigma^2}{SS_x}$**
  - **The likelihood of the average $\alpha_{\overline{X}}$ follows a Gaussian distribution with mean $\overline{Y}$ and variance $\frac{\sigma^2}{n}$**
- **If the prior density for β, is Gaussian with mean m_β and variance s_β^2 the posterior is Gaussian with mean m'_β and variance s'^2_β where**

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

BAYESIAN OLS – RESULTS (2)

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

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

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

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

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

HIERARCHICAL APPROACH EXAMPLE

Applying Bayes' Theorem to Parametrics

GODDARD's RSDO EXAMPLE

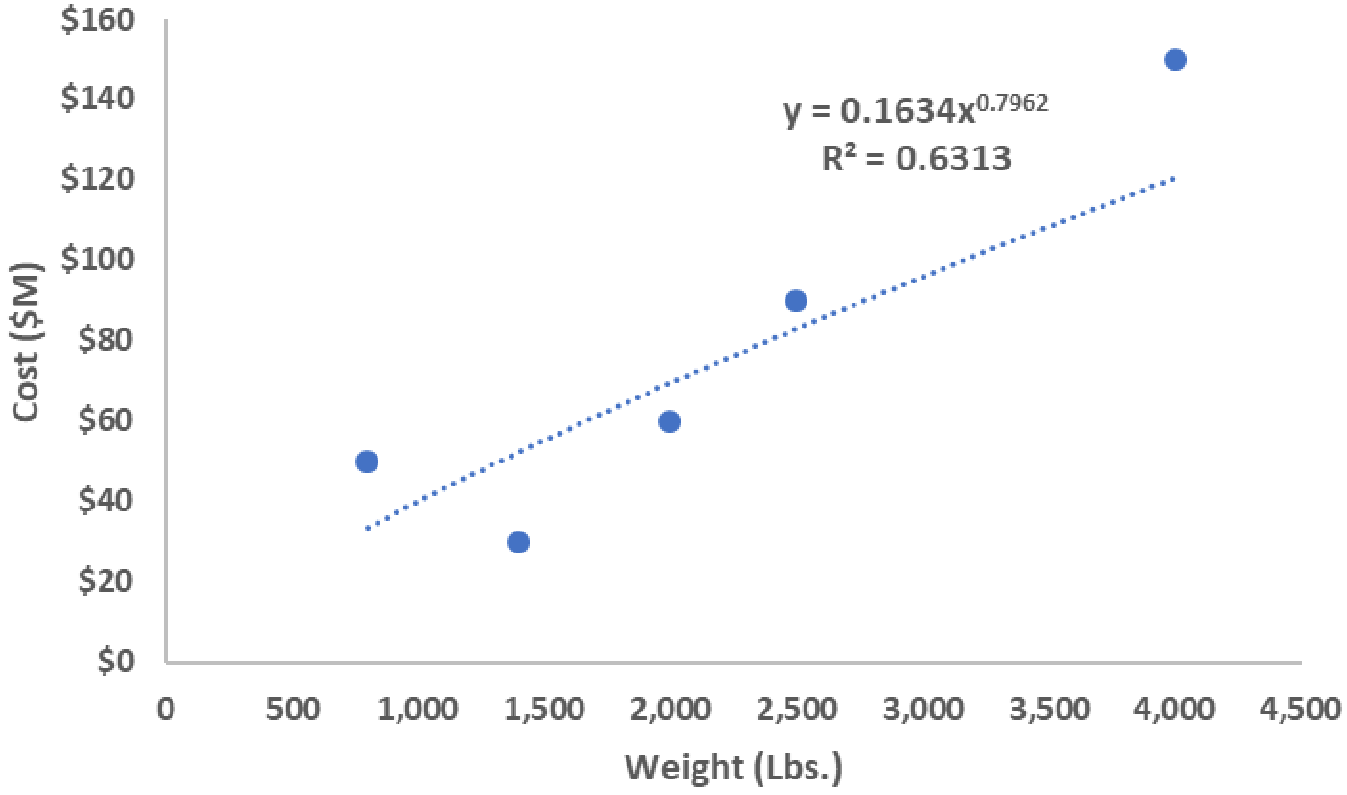
NASA's Rapid Spacecraft Development Office (RSDO) uses streamlined acquisition processes and fixed-price contracts to cut costs for robotic Earth-orbiting satellites

Give five historical data points of RSDO missions, what is the best way to estimate an RSDO mission

One option is to develop a CER with the five data points

However, given our discussion of the law of small numbers, a trend line developed from these five points may not be meaningful

Alternative – consider a larger set of robotic Earth-Orbiting satellites



HIERARCHICAL APPROACH EXAMPLE

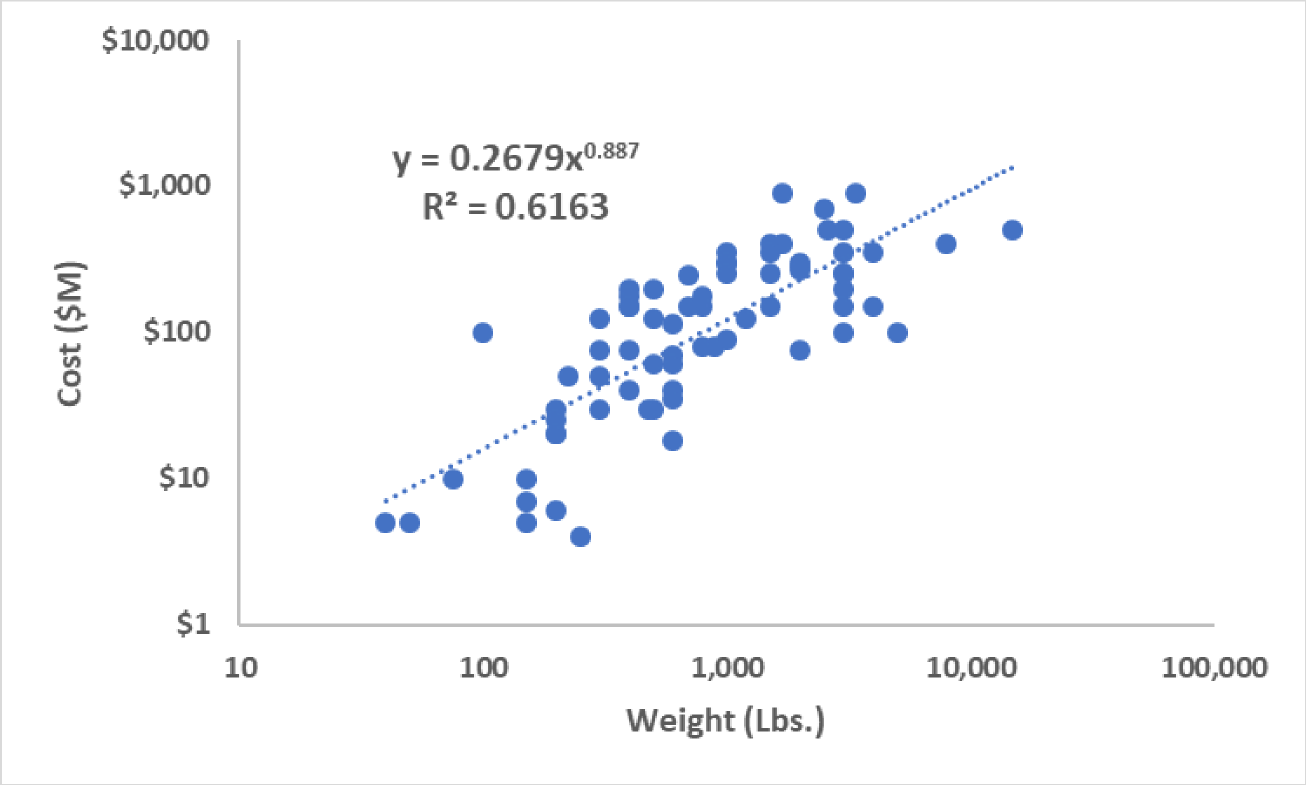
Applying Bayes' Theorem to Parametrics

EARTH-ORBITING DATA (FROM NAFCOM)

There are many more earth-orbiting data points if we do not restrict our attention to only RSDO missions

However, these missions are substantially more expensive than the RSDO analogues

Classical techniques apply here but they will likely overestimate a new RSDO mission by a significant amount



HIERARCHICAL APPROACH EXAMPLE

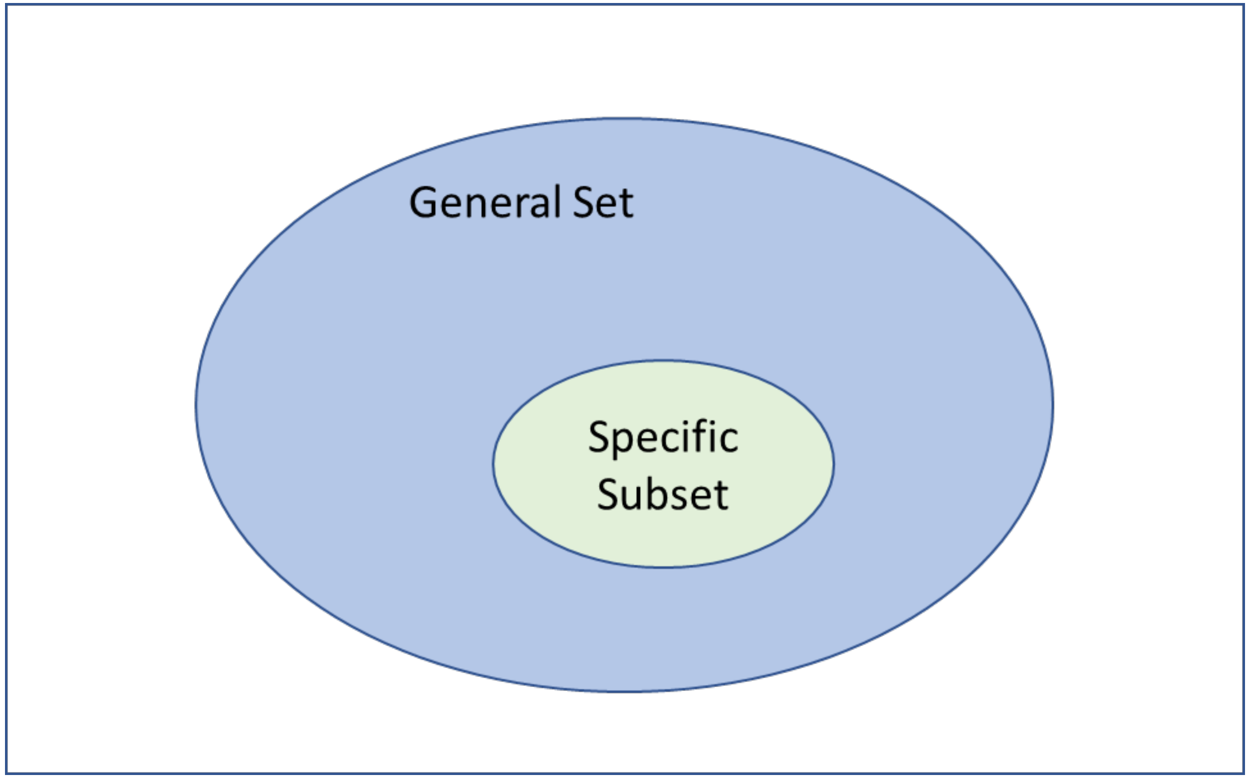
Applying Bayes' Theorem to Parametrics

COMBINING THE DATA

Bayes' Theorem provides a way to use both sources of data

The large, generic set of earth-orbiting missions can be used as a prior

The small specifically applicable data set can be used to update the prior probabilities



HIERARCHICAL APPROACH EXAMPLE

Transforming the Data

LOGARITHMIC

The equation has a power form

$$\tilde{Y} = aW^{b}$$

We use log-transformed OLS in this example

This allows for an analytical solution for the Bayesian update, using Gaussians in log-space

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

HIERARCHICAL APPROACH EXAMPLE

Obtaining the Variances

DIRECTLY AVAILABLE FROM THE REGRESSION ANALYSIS

Using the Data Analysis add-in for Excel, we conduct two regression

The output of the regression analysis for the prior is shown on the right

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

HIERARCHICAL APPROACH EXAMPLE

COMBINING THE PARAMETERS

DIRECTLY AVAILABLE FROM THE REGRESSION ANALYSIS

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 precision

For the intercept, the relative precision weights for the intercept of the NAFCOM data are

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

And for the RSDO data, it is $\mathbf{1-0.6870=0.3130}$

HIERARCHICAL APPROACH EXAMPLE

COMBINING THE PARAMETERS (2)

DIRECTLY AVAILABLE FROM THE REGRESSION ANALYSIS

For the slope the 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

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

HIERARCHICAL APPROACH EXAMPLE

THE PREDICTIVE EQUATION

DIRECTLY AVAILABLE FROM THE REGRESSION ANALYSIS

The predictive equation in log-space is

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

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

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

HIERARCHICAL APPROACH EXAMPLE

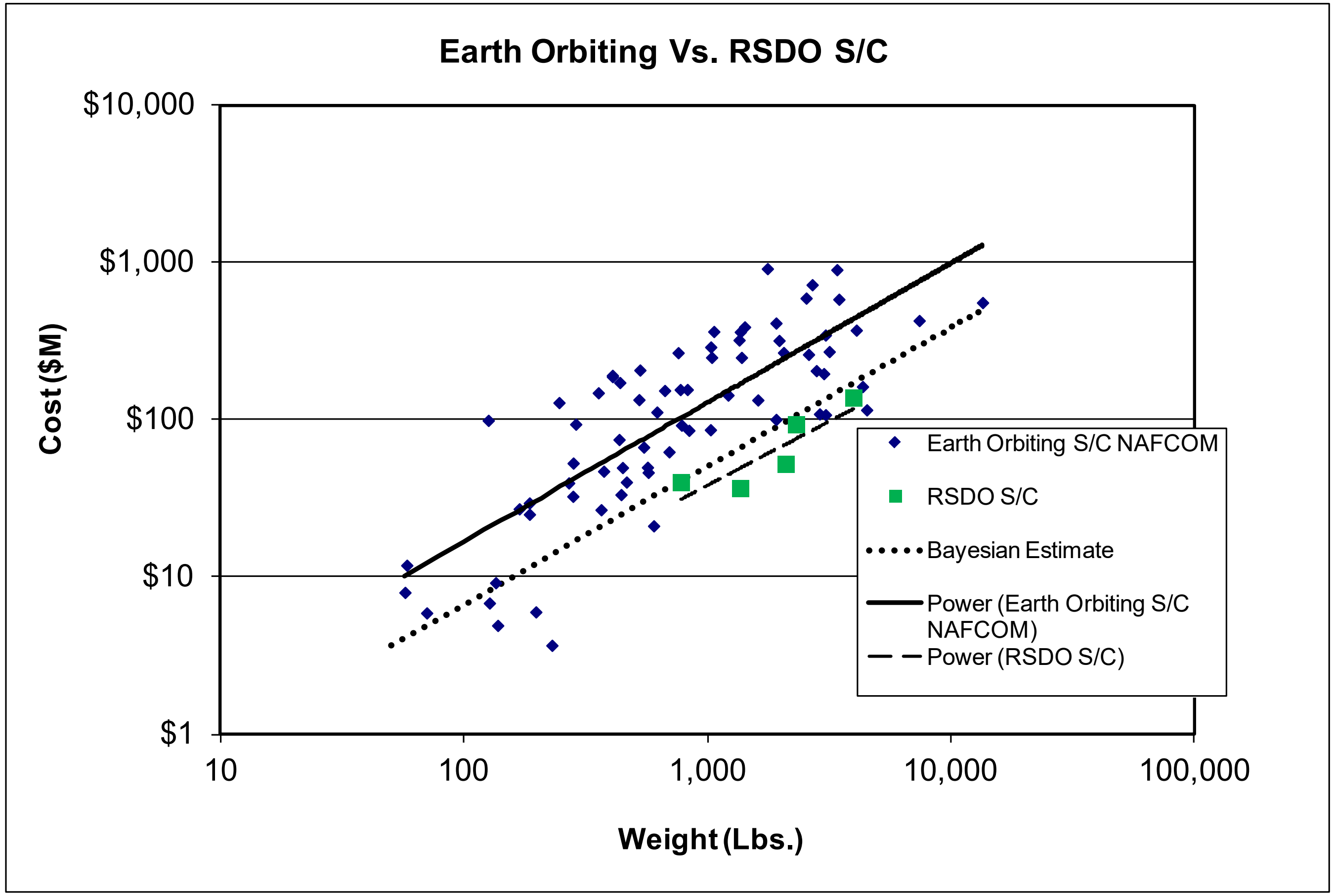
Applying Bayes' Theorem to Parametrics

RESULTS

The Bayesian CER coefficients are weighted averages of the coefficients of the CERs based on the two separate data sets

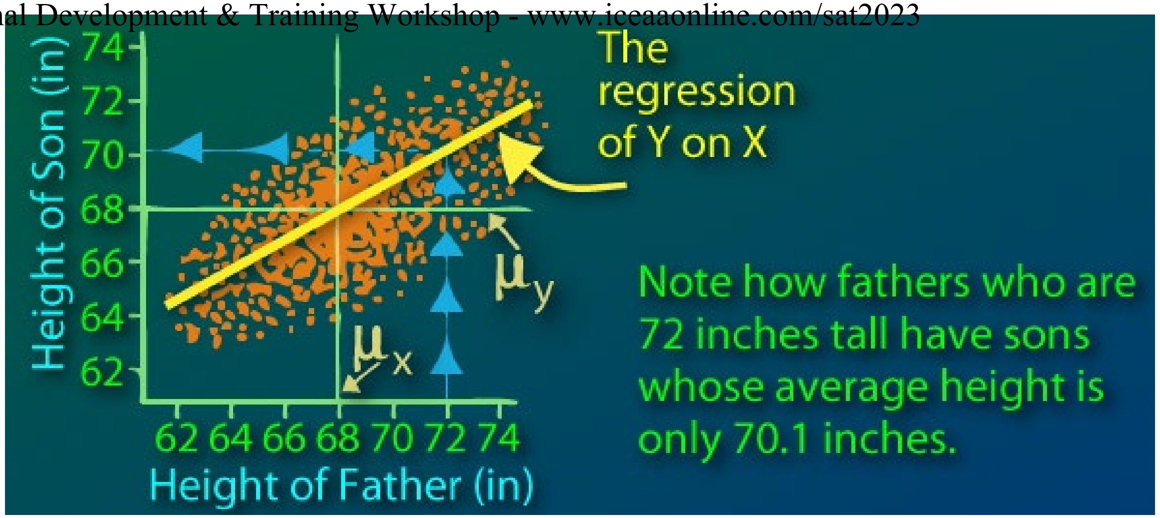
The Bayesian CER is applicable to a wider range of data than just the RSDO missions

In testing the CER on an RSDO mission not in the data set, the Bayesian CER was more accurate than either of the RSDO-only CER or the all Earth-orbiting CER



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

APPLYING THE EQUATION



IN PRACTICE

Applied the equation to predicting a new RSDO mission

VALUE

Estimates:

Bayesian regression $144 M

NAFCOM data: $368 million;

RSDO data: $100 million

ACTUAL

Actual cost for the mission was $180 m

Bayesian regression was the best estimate

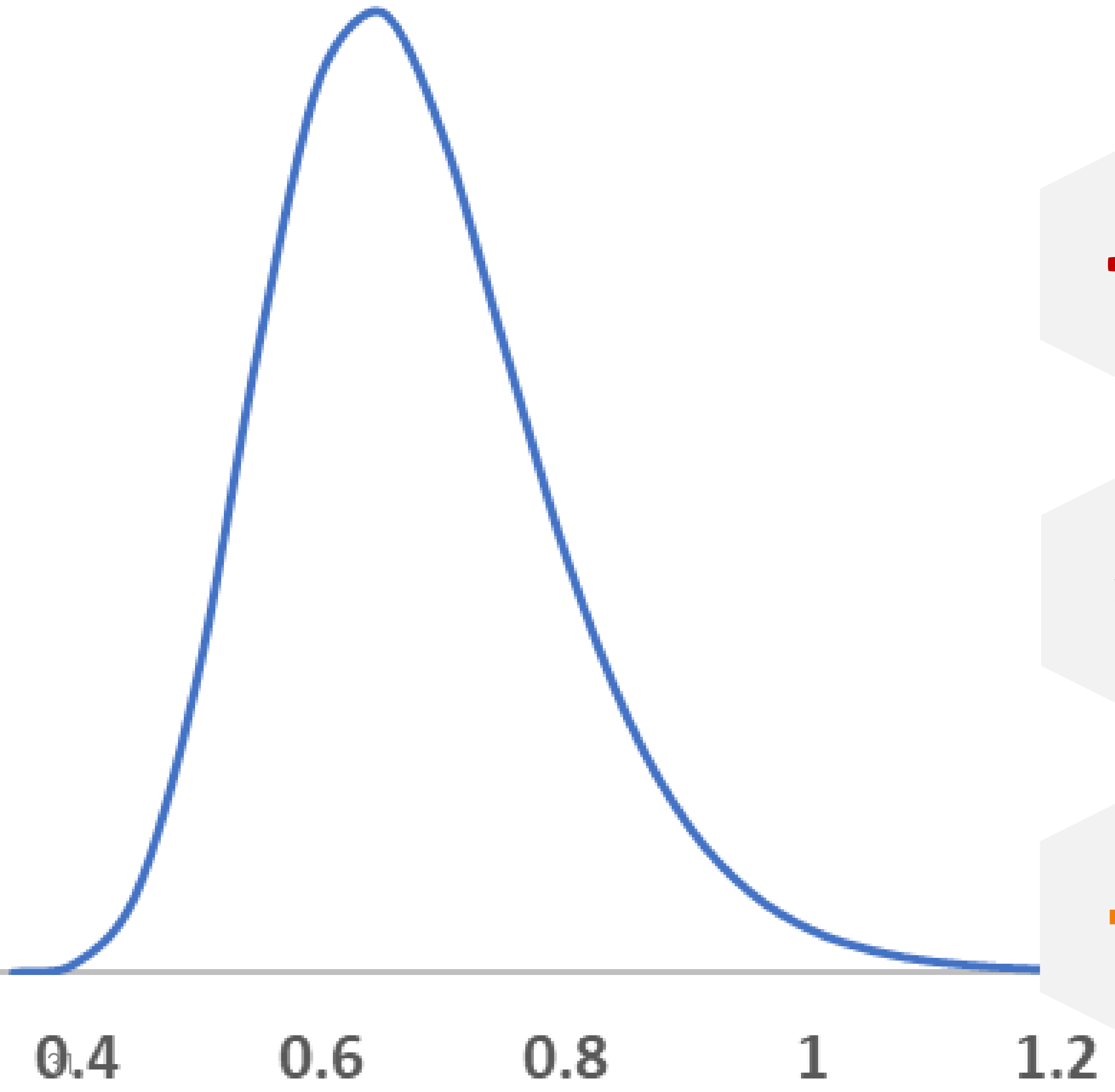
REGRESSION TO THE MEAN

Extremes are typically followed by values closer to the middle

THE BASIC METHOD AND ITS ISSUES

- **Everything we have discussed to this point relies on the use of a basic Bayesian method that has been presented before - for details see:**
  - **Christian Smart, "Bayesian Parametrics: How to Develop a CER with Limited Data and Even without Data," 2014 ICEAA Conference**
  - **Pierre Foussier, "The Benefits of the Bayesian Approach vs. the Frequentist Approach when Dealing with Low Data Sample," presented at the 2008 ISPA-SCEA international conference.**
- **There are some issues with this basic approach:**
  - **For a power equation, it requires the use of log-transformed ordinary least squares (biased low)**
  - **One assumption is that we know that the variance of the CER based on the sample data is known with certainty**
  - **Another assumption is that the residuals are lognormally distributed (normal in log space)**

UNKNOWN VARIANCE



THE MOST PROBLEMATIC ASSUMPTION IS THAT OF KNOW VARIANCE

In our example, the sample data is the small set, the one for which we have the least confidence in knowledge of the population variance

GOOD NEWS THIS CAN BE HANDLED ANALYTICALLY

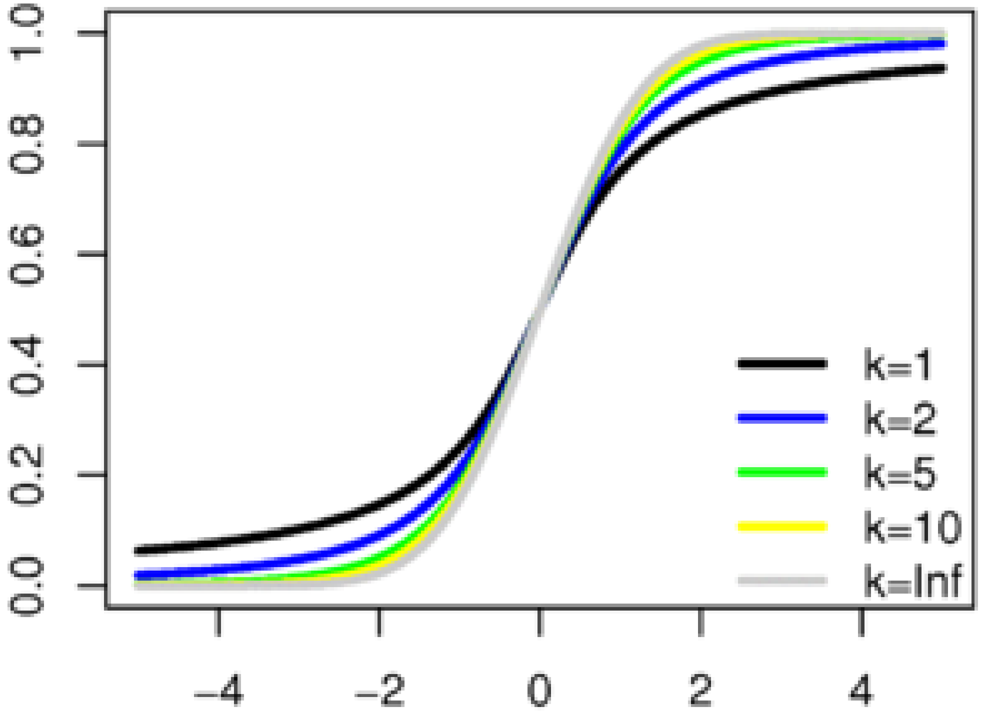
By Cochran's Theorem, the variance can be modeled as a scaled inverse-Chi Square distribution

ESTIMATING PARAMETERS

Degrees of freedom is the same as the degrees of freedom in the regression of the sample data

Sample variance is the scaling factor

(LOG) NORMAL RESIDUALS



Student-t Distribution

GAUSSIAN ASSUMPTION MAKES THE MODEL ANALYTICALLY TRACTABLE

Gaussian likelihood has Gaussian as a conjugate prior

Even with unknown variance model is still tractable

Log-t has been proposed as an alternative to lognormal

Also - nonparametric methods have no distributional assumption

IF RESIDUALS ARE NOT GAUSSIAN, MUST USE SIMULATION

With small samples, residuals may be better modeled with log-t

Markov Chain Monte Carlo simulation can be used to do the Bayesian analysis via simulation

MARKOV CHAIN MONTE CARLO (MCMC)

MCMC uses conditional simulation – each trial depends on the previous

Can implement in R or use WinBUGS

MCMC SIMULATION

THE OTHER CASES

When the model residuals are not distributed according to Gaussian or lognormal distributions, you want to use other CER techniques other than ordinary least squares and log-transformed ordinary least squares then the analytical approaches break down

MARKOV CHAIN MONTE CARLO (MCMC)

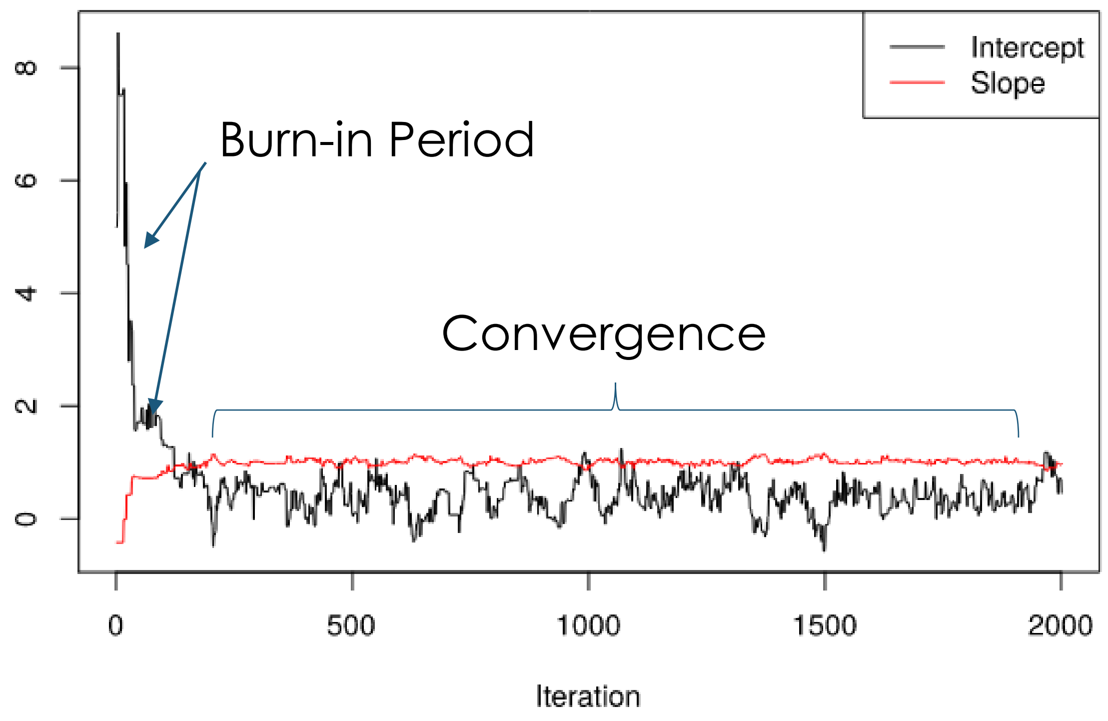
Fortunately, there are simulation-based alternatives such as the Markov Chain Monte Carlo approach for calculating the posterior distribution

Monte Carlo dates back to the Manhattan Project in the 1940s and the concept of MCMC simulation dates back to the 1950s; but MCMC simulation was not widely adopted until the 1990s

CALCULATING MCMC

Two popular options for conducting MCMC calculations are the free platforms WinBUGS (Bayesian inference Using Gibbs Sampling) and R

MCMC Examples



MCMC APPLICATION

COMPARISON WITH OTHER CASES

WinBUGS code:

```
model{for(i in 1:n){
y[i]~dt(mu[i],tau,3)
mu[i]<-beta[1]+beta[2]*(x[i]-mean(x[]))
}
#priors
beta[1]~dnorm(4.6087,109.4299);beta[2]~dnorm(0.8858,152.6058)
tau~dgamma(35,23.02)
}
#Load data. Include the x-value for the prediction, along with "NA" for the y-value, in
#order to calculate statistics for the prediction
list(x=c(8.2938,7.750,7.2235,6.6598,7.6533,8.0956),
y=c(4.9182,4.5276,3.5965,3.68437,3.9526,NA),n=6)

#Initialize parameters for two chains.
list(beta=c(5,0.7),tau=1,yrep=c(0,0,0,0,0,0))
list(beta=c(4,0.9),tau=0.1,yrep=c(0,0,0,0,0,0))
```

ASSUME GAUSSIAN PRIOR WITH STUDENT-T LIKELIHOOD (LOG SPACE)

Cannot be computed analytically but can be implemented in WinBUGS with a few lines of code.

COMPARISON

Case 1: Gaussian prior and likelihood with known variance

Case 2: Gaussian prior and likelihood with unknown variance

Case 3: Gaussian prior and Student-t likelihood

For Cases 2 and 3, the predicted value is $163 million, within 10% of the actual

	Case 1	Case 2	Case 3
log-space intercept	4.4607	4.5800	4.5780
linear intercept	0.8794	0.8846	0.8880
slope	0.1166	0.1263	0.1229
prediction ($ millions)	144	163	163

RSDO - RESULTS COMPARISON

Confidence level	Normal likelihood with known variance	Normal likelihood with unknown variance	Student t likelihood, unknown variance, normal error	Student t likelihood, unknown variance, Student's t error
5%	$83	$44	$15	$23
10%	$94	$59	$25	$41
20%	$109	$83	$47	$72
30%	$121	$107	$76	$100
40%	$132	$133	$112	$129
50%	$144	$163	$163	$163
60%	$157	$199	$236	$205
70%	$171	$248	$352	$266
80%	$190	$319	$560	$370
90%	$220	$453	$1,068	$642
95%	$248	$606	$1,819	$1,169
99%	$311	$1,044	$4,942	$7,294
99.5%	$338	$1,275	$7,125	$21,659
99.9%	$401	$1,922	$15,151	$842,673

FOUR CASES CONSIDERED

1. Base case – Gaussian model (lognormal)
2. (Log) Gaussian residuals with unknown variance
3. Unknown variance, log-t for likelihood, lognormal predictive
4. Unknown variance, log-t for likelihood, log-t predictive

IMPLEMENTED IN BOTH R AND WINBUGS

Unknown variance changes the point estimate

The use of different uncertainty distributions does not change the point estimate (50th percentile)

THE DEVIL IS IN THE TAILS

the use of log-t dramatically effects the tails of the distribution

1 in 1,000 chance that a $163 million cost may grow to tens and hundreds of billions!

COMPARISON COMMENTS (1 OF 2)

- **Changing the assumption results in higher estimates**
  - **The 50th percentile for the Gaussian model is $144 million**
  - **Relaxing the variance assumption increases the 50th percentile to $163 million**
  - **Changing the CER residuals assumption does not change the 50th percentile**
  - **Improves the accuracy – actual cost was $180 million is 10% higher than the estimate with unknown variance**
- **There is also an increase in the heaviness of the right tail when relaxing the Gaussian assumptions**
  - **The unknown variance case has a 99th percentile equal to $1 billion**
  - **Extreme but it happens**

COMPARISON COMMENTS (2 OF 2)

- **When going to the log-t, the extreme right tail explodes, says that a relatively simple earth-orbiting spacecraft could cost as much as the James Webb Space Telescope (or more)!**
- **The relaxation of the assumption of known variance is important and the results are logical**
- **The change of the modeling of the residuals to a Student's t distribution in log space (log-t) is not logical**
- **Need to balance common sense with mathematical correctness – cost modeling is both an art and a science**

SUMMARY - BAYES

AN IDEAL METHOD WHEN YOU HAVE LIMITED DATA

MYRIAD REAL-WORLD APPLICATIONS

Cracking Enigma code

Search-and-Rescue

Property and Casualty Insurance Premium Setting

USES ALL YOUR DATA

Can be objective or subjective

Allows you to use all your data

FILLS A NEED

Ideal for small data sets

GAUSSIAN (BASIC) MODEL

Analytically tractable

Can be done in Excel

ADVANCED TECHNIQUES

The Gaussian model has issues with variance assumptions

If residuals are not lognormal, may require the use of Markov Chain Monte Carlo

LOW-HANGING FRUIT

Currently limited use

High potential