Are You Smarter than an Algorithm?

Andy Prince Manager, Engineering Cost Office NASA/Marshall Space Flight Center May 14, 2020



Outline



- Introduction and Problem Statement
 - The scientists versus the cost estimators
- Approach
 - Develop two models, one with and one without subjective parameters
 - Solicit subjective inputs from fellow cost professionals
 - Compare results, see who wins
 - Figure out this is harder than it looks
- Analysis of Results
 - Statistics and more statistics
 - Signal and noise: what is useful and what is not
 - Do monkeys make better estimators?
- Winners and Losers
 - What did we learn?
 - Lessons for the future



Introduction



- 2016 Paper: The Dangers of Parametrics
 - "Much like the "Dark Side of the Force" from "Star Wars" mythology, subjective parameters seduce the cost model developer. This seduction comes from their power to explain the random noise in our data, to improve the model statistics, and to enable the estimator to fine-tune the estimate to reflect their evaluation of a new system."
- Several Luminaries in the Field of Space Systems Cost Analysis Disagree
 - Don MacKenzie: "I can offer this: having been a long-time PRICE H user, I believe that employing subjective cost drivers is superior to models with only measurable inputs. Thus percent new design, percent unique, fractional development unit counts and relative complexity of hardware are all important subjective inputs. "





- Daniel Kahneman, Nassem Talib, Nate Silver, Leonard Mlodinow, etc.
- The Gist of the Argument
 - Humans are primed to make judgements on what we know and what we perceive
 - Human judgement works very fast (is there a tiger behind that bush?)
 - Human judgement is subject to biases (Anchoring, Confirmation Bias, Attractiveness, Optimism, Frequency, etc.)
 - Biased judgement introduces additional error into an assessment process
 - Numerous studies and trials have shown that simple algorithms are superior to human judgement





- Joe Hamaker, Don MacKenzie, Ron Larson, Christian Smart, etc.
- The Gist of the Argument
 - Each spaceflight hardware system is unique, with it's own challenges and peculiarities
 - No parametric model, no matter how good, can account for all unique system attributes (or totally unique systems)
 - Our data is limited and imperfect
 - Cost estimating has a high degree of subjectivity, so go ahead and be upfront about it
 - Subjective variables allow the estimator the ability to address issues important to the customer
 - You cannot replace in-depth knowledge and years of experience with an algorithm





6

- The Question: Would a Cost Model Consisting Solely of Objective Parameters Outperform a Cost Model with both Objective and Subjective Parameters?
- The Challenge: Design an Experiment that would Test the Hypothesis and Provide Definitive Proof





7

- 1. Find or Development Two Space System Cost Models
 - Model "A" would have both Subjective and Objective Inputs
 - Model "B" would have only Objective Inputs
- 2. Use Completed (Launched) Space Missions to Test both Models (Test Data Set)
 - Missions cannot be in the Model Development Data Set
 - Rely on NASA CADRe Data unbiased source
- 3. Solicit Inputs for the Subjective Parameters for the Test Data Set from Space Cost Professionals
 - Did Not Provide the Model
- 4. Evaluate the Performance of both Models on the Test Data Set
 - Generate Lots and Lots of Statistics

5. Determine the Winner!





- Estimate Space Flight Hardware
- Have Subjective Input Parameters
- Not too Recent, Not too Old
- Not too Complex, Not too Simple
- The Solution: QuickCost 5.0
 - Developed by Joe Hamaker in 2011
 - Estimates Robotic Spacecraft Bus + Science Instruments for NASA Missions
 - Single CER Built on 131 Data Points
 - 8 Input Parameters including 3 Subjective Parameters





- Developed to Give NASA a Capability to Quickly Estimate the Cost of a New Space Science Mission
- Equation Form:
 - LnCost = 1.17 + 0.135 Destination 0.0179 ATP + 0.170 LnLifeMonths + 1.09 InstrComp% + 0.682 LnTotDryMass + 0.118 LnPower + 0.926 BusNew + 0.462 InstrNew
 - Where
 - LnCost: Estimate in FY2004 Ln Dollars
 - Destination: Earth Orbital (0) or Planetary (1)
 - ATP: Start of Preliminary Design, ATP Year 1960
 - LnLifeMonths: Ln of the Planned Mission Lifetime in Months
 - InstrComp%: Complexity of the Instrument Suite on a Scale of 0% to 100%
 - LnTotDryMass: Ln of the Total Dry Mass of the Flight System in Kilograms
 - LnPower: Ln of the BOL Output Power in Watts of the System Normalized to LEO Equivalent
 - BusNew: Spacecraft Bus New Design on a Scale of 0% to 130% (or beyond)
 - InstrNew: Instrument Suite New Design on a Scale of 0% to 100%





- To Ensure Consistency, Used QuickCost 5.0 Data Set and the same Objective Parameters to Construct the Objective Model
- Equation Form
 - LnCost = 1.0186 + 0.6031 LnTotDryMass + 0.1294 LnPower 0.4970
 LnATP + 0.3501 LnLifeMonths + 0.4504 Destination
 - All the Parameters are the same with the Exception of ATP, Ln of ATP is used in the Objective Model





Model Statistics



11

Statistic	QuickCost 5.0	Objective Model
R-Squared	0.8694	0.7536
Standard Error	0.3645	0.4947
Mean Residuals	-\$30.2	\$58.7
Standard Deviation Residuals	\$226.3	\$289.3
Mean Absolute Deviation (MAD)	\$28.4	\$170.5
Standard Deviation MAD	\$191.7	\$240.9
Mean Absolute Deviation Ratio %	28.4%	41.9%
Standard Deviation Absolute Deviation Ratio %	28.4%	41.1%
Root Mean Squared Error	\$228.3	\$295.2

- **Residual =** *Estimated Cost minus Actual Cost*
- Mean Absolute Deviation = $\Sigma(Absolute Value(Residual))/n$
- Mean Absolute Deviation Ratio = $\Sigma(Absolute Value(Residual/Actual))/n$
- Root Mean Squared Error = $\sqrt{\sum (Estimate Actual)^2/n}$



The Test Data Set



<u>15</u>	5 NASA Science Missions Launched Since 2011
CYGNSS	Cyclone Global Navigation Satellite System
GPM	Global Precipitation Measurement
GRAIL	Gravity Recovery and Interior Laboratory
IRIS	Interface Region Imaging Spectrograph
JUNO	
LADEE	Lunar Atmospheric and Dust Environment Explorer
MAVEN	Mars Atmosphere and Volatile Evolution
MMS	Magnetospheric Multiscale Mission
MSL	Mars Science Lander (aka Curiosity)
NuSTAR	Nuclear Spectroscopic Telescope Array
OCO-2	Orbiting Carbon Observatory - 2
OSIRIS-Rex	Origins - Spectral Interpretation - Resource Identification - Security Regolith Explorer
THEMIS	Time History of Events and Macroscale Interactions during Substorms
VAP	Van Allen Probes
WISEresented f	oWidentialdinfraredsSesterategeoAnalysis Association - www.iceaaonline.com



The Survey Instrument



Input Sheet

Instructions:	Assign New Design and Complexity Factors to each of the 15 NASA science missions listed below							
	Scale is 0% to 130% (or greater) for New Design, 0% to 100% for Instrument Complexity							
	You may use whatever data sources are available to you							
	Please prov	vide an input f	for all factors! ເ	Jse your best jud	dgment, nobody is being graded.			
	When finishe	ed please ansv	ver the questions	in the "Demogr	aphics" tab			
	Bus New	Instrument	Instrument					
Mission	Design	New Design	Complexity		Definitions			
CYGNSS					Satellite Bus New Design factor in percentile terms. Consider the following guidelines:			
GPM					o 20% of totally off-the-shelf			
GRAIL					o 60% average			
IRIS					o 100% all new			
JUNO					o 130% (or more) for all new and pushing state-of-the-art			
LADEE								
MAVEN					Instrument suite New Design factor in percentile terms with the same scale as bus			
MMS					(but instruments typically have less heritage or higher new design factors than buses)			
MSL								
NuSTAR					Instrument Complexity in percentile terms and representing a weighted average of the entire instrument suite.			
OCO-2					For example if the instruments are of median complexity, 50% is entered.			
OSIRIS-Rex					Instruments that are judged to be around the 75th percentile of complexity would be entered as 75%.			
THEMIS					Scale is 0% to 100%			
VAP								
WISE								

Used Full and Complete Definitions from QuickCost





Please answer the following questions	
	Answers Here
Who is your employer?	
How long have you been a cost estimator?	
How many years have you worked in the	
space business?	
What is your gender?	
What is your age?	
Citizenship	





- Sent survey to approximately 50 professionals in the NASA, US government, business, and international cost communities
- Sat back and waited for responses

...and waited...

...and waited...

...and waited...

Sent out reminders!!!

...and waited...

...and waited...

...and waited...

15





- Survey was too long
- People did not feel qualified (professional integrity)
- Lack of time
- No money
- Lack of interest?
- Ultimately got 10 full responses and one partial response
 - Browbeat staff and contractors



Who are these People?









- Easy to generate statistics, but what is of value, and what is not?
- How much Better Than the Objective Model (BTOM) is proof that subjective judgement adds value?
- Is there a killer statistic that will irrefutably demonstrate the superiority of either the objective or subjective approach?









19

- Means, Medians, Standard Deviations, Mins Maxes, …
 - Residual = Estimated Cost minus Actual Cost
 - Absolute Deviation = Absolute Value(Residual)
 - Deviation Ratio = Residual/Actual
 - Absolute Deviation Ratio = Absolute Value(Residual/Actual)
- Root Mean Squared Error = $\sqrt{\sum (Estimate Actual)^2/n}$
- Prediction Intervals
- Statistical Tests
 - Paired t-Test
 - Wilcoxon Signed Rank Test
 - van der Waerden Signed Rank Test
- Analysis of 15 Mission Portfolio vs. Individual Mission
 - Individual Performance or Averages?



Roadmap



20

A guide to reading future charts...







Results for the 15 Mission Portfolio

	Mean	Mean Absolute	Mean Absolute			Wilcoxon Signed-
Code	Residual	Deviation	Deviation Ratio (%)	RMSE	Paired t-Test	Rank Test
AA	-\$232.9	\$244.7	79.5%	\$361.7	0.0071	0.003
AB	-\$156.4	\$194.2	52.1%	\$269.0	0.0182	0.020
AC	-\$344.5	\$358.7	95.8%	\$640.6	0.0317	0.002
AD	-\$139.0	\$182.9	41.1%	\$312.7	0.0844	0.140
AE	-\$260.6	\$307.0	55.2%	\$581.1	0.0815	0.023
AF	-\$132.1	\$220.4	62.9%	\$342.5	0.1401	0.039
AG	-\$306.3	\$351.4	81.3%	\$613.2	0.0488	0.012
AH	\$9.1	\$128.6	36.6%	\$167.0	0.8407	0.860
AI	-\$254.6	\$315.2	60.6%	\$655.5	0.1370	0.128
AJ	-\$103.2	\$130.6	43.8%	\$178.3	0.0189	0.009
QM	\$31.9	\$118.7	27.3%	\$242.3	0.6266	0.140
ОМ	-\$2.2	\$162.0	37.3%	\$261.4	0.9748	0.305

Respondent AH and QuickCost with median subjective variable settings (QM) performed as well or better than the Objective Model





= Better Than Objective Model (BTOM)

Code	CYGNSS	GPM	GRAIL	IRIS	JUNO	LADEE	MAVEN	MMS	MSL	NuSTAR	OCO-2	OSIRIS-REx	THEMIS	VAP	WISE
AA	-\$43.1	-\$881.6	\$33.9	-\$83.6	-\$496.0	\$31.3	-\$329.1	-\$26.4	-\$65.4	-\$356.7	\$23.5	-\$748.3	-\$134.6	-\$312.8	-\$104.7
AB	-\$23.0	-\$448.9	\$84.5	-\$36.9	-\$588.0	\$110.5	-\$43.0	-\$458.0	-\$278.9	-\$27.1	\$88.8	-\$407.4	-\$208.2	-\$95.9	-\$14.0
AC	-\$54.9	-\$311.8	-\$338.4	\$22.8	-\$2,171.7	-\$101.4	-\$740.0	-\$125.0	-\$245.8	-\$55.0	\$83.8	-\$708.2	-\$236.7	-\$146.6	-\$38.8
AD	\$7.5	-\$627.8	\$52.6	\$10.2	-\$943.0	\$94.6	-\$79.5	-\$307.5	-\$128.6	-\$8.2	\$133.6	-\$127.3	-\$86.2	-\$106.1	\$30.2
AE	-\$3.3	-\$464.6	\$105.2	-\$28.8	-\$1,099.4	\$117.6	-\$121.5	\$55.9	-\$1,855.0	-\$2.3	\$69.8	-\$327.7	-\$146.2	-\$126.1	-\$82.4
AF	-\$55.3	-\$49.9	\$78.5	\$17.9	-\$1,029.4	\$31.3	-\$318.1	-\$204.2	\$378.8	-\$149.8	\$97.7	-\$571.6	-\$93.1	\$58.0	-\$172.5
AG	-\$45.1	-\$709.0	\$132.1	\$29.4	-\$928.7	\$83.4	-\$242.3	\$44.1	-\$1,981.7	-\$235.7	\$48.8	-\$202.8	-\$262.3	-\$301.6	-\$23.7
AH	\$7.0	-\$411.5	\$145.8	\$9.3	-\$264.6	\$120.7	\$145.9	-\$130.3	\$74.5	-\$3.3	\$146.3	\$244.6	-\$86.6	\$91.6	\$47.6
AI	-\$47.4	-\$242.9	\$60.5	\$19.9	-\$497.5	\$77.3	-\$128.3	-\$174.0	\$68.3	-\$89.2	-\$27.5	-\$245.7	-\$171.5	-\$74.2	-\$55.2
AJ	-\$14.8	-\$378.2	\$146.7	-\$0.2	-\$2,149.1	\$117.1	-\$52.5	-\$110.8	-\$1,259.9	-\$31.5	\$123.5	\$47.3	-\$104.0	-\$84.8	-\$88.7
BA				\$15.2											
QM	-\$36.1	-\$205.7	\$63.8	-\$4.6	-\$185.9	\$97.6	-\$19.7	-\$50.5	\$876.7	-\$25.7	\$43.6	-\$12.9	-\$109.3	\$48.1	-\$0.8
ОМ	-\$32.7	-\$82.4	\$60.9	-\$3.8	-\$540.4	\$94.4	-\$97.0	\$105.9	\$730.4	-\$18.7	\$48.5	-\$365.7	-\$91.5	\$98.0	\$60.5

Averaging the estimates resulted in 7 BTOM estimates but poor aggregate statistics



Closest to the Target



			11 Respor	ises	Plus QM
Total Number	of Respo	nses:	151		166
Better Than C	Dbjective N	/lodel	54		61
"Win" Rate:	2		36%		37%
Mission	BTOM			Code	BTOM
CYGNSS	5			AA	6
GPM	1			AB	5
GRAIL	3			AC	2
IRIS	1			AD	8
JUNO	4			AE	4
LADEE	4			AF	4
MAVEN	4			AG	4
MMS	4			ΔН	8
MSL	7				8
NuSTAR	3				5
OCO-2	3			AJ	5
OSIRIS-REx	7			BA	0
THEMIS	2			_	
VAP	6			QM	7
WISE	7				





- Ran QuickCost 5.0 using Uniform Distributions to simulate subjective variable input values
 - PDF minimum and maximum based on recommended inputs
 - Correlated inputs based on Hamaker's QuickCost values
 - Bus New Design to Instrument Suite New Design: 0.43
 - Instrument Complexity to Instrument Suite New Design: 0.33
 - Bus New Design to Instrument Suite Complexity: 0.32
 - Calculated the absolute value of the residual for each trial
 - Ran 5000 trials

	Objective Model	Monkey BTOM
Mission	ABS(Residual)	Percentile
CYGNSS	\$32.7	35.7%
GPM	\$82.4	11.2%
GRAIL	\$60.9	27.9%
IRIS	\$3.8	4.9%
JUNO	\$540.4	59.3%
LADEE	\$94.4	53.8%
MAVEN	\$97.0	31.4%
MMS	\$105.9	15.8%
MSL	\$730.4	43.2%
NuSTAR	\$18.7	17.0%
OCO-2	\$48.5	26.9%
OSIRIS-Rex	\$365.7	69.4%
THEMIS	\$91.5	32.2%
VAP	\$98.0	31.4%
WISE	\$60.5	35.1%
Average		33.0%







	Result	Winner
Participant Statistics (W - L - T)	1 - 9 - 1	OM
Participant BTOM	36%	OM
QM Statistics	Tie	Tie
QM BTOM	47%	Tie
Monkeys	33%	OM

OM - Objective Model

QM - Quickcost with Average/Median Subjective Inputs

- <u>The Objective Model was a clear winner</u> over the survey participants and the monkeys
- QuickCost with average/median settings did well enough that it could be called a tie
- Given these results are there any situations where the estimator should use subjective input variables?



Another Point of View

Thanks to Christian Smart



- I agree there are hazards with using subjective parameters. Two that spring to mind, and I'm sure there are more, are:
 - 1. It is prone to the biases that you mention, such as the planning fallacy.
 - 2. It can be manipulated to produce lower cost estimates. This applies more to program office estimates than independent ones.
- However, percent new design is one of my favorite parameters, and I think it is an important cost driver.
- Even with our biases, subjectivity is important in cost estimating:
 - 1. Unless we completely automate the model development process, there will be subjectivity in our estimates and our models.
 - 2. Subjectivity is an important part of the experience that an estimator brings to the development of cost estimates. A parametric model, even a black-box one, is just a framework for codifying experience in developing a cost estimate. Sometimes too few parameters in a model can make it hard for an estimator to use their experience in applying a parametric model.
 - 3. We have small data sets. This makes the use of experience, such as Bayesian methods, very important in producing accurate estimates. Presented for the International Cost Estimating & Analysis Association - www.iceaaonline.com







- When in doubt (90% of the time), don't use subjective parameters
 - High potential for abuse
 - Can add value if the estimator has in-depth knowledge of the system and significant estimating experience
 - Most people overestimate their knowledge and experience
- You should not overtly introduce subjectivity into an estimate without credible, supportable, and defendable logic and justification
 - Recognize that you could be adding error
 - Your subjectivity can be used against you
 - Calibrate and validate whenever possible
- Subjective inputs may be useful when estimating systems outside the model's experience base
 - MSL was more complex than the typical mission and estimates using the subjective parameters generally were BTOM
- A subjective assessment could add value when doing a comparative analysis
 - Requires equivalent knowledge and a systematic application of subjective criteria
 Presented for the International Cost Estimating & Analysis Association www.iceaaonline.com





- Subject parameters may make our job easier, but they do not automatically make our estimates better
- The less information you have the more you should rely on objective models
- Don't assume that you are an expert or that you are unbiased
- Doing a survey is harder than it looks







Objective Model

Respondent AB



8 Missions Overestimated Range of Residuals: -\$540.4 to \$730.4

12 Missions Overestimated Range of Residuals: -\$588.0 to \$110.5



But not Everyone



Respondent AH



5 Missions Overestimated Range of Residuals: -\$411.5 to \$244.6

QuickCost Mean Settings



10 Missions Overestimated Range of Residuals: -\$205.7 to \$876.7



Contact Information





Andy Prince Manager, Engineering Cost Office NASA, Marshall Space Flight Center 256-544-8360 andy.prince@nasa.gov