Are you Smarter than an Algorithm?

Andy Prince

Experienced cost analysts know that estimating requires historical data, models, knowledge, and judgment. But over the past few decades, psychological research has shown that people's judgments are faulty and their memories are often colored by their biases. Within the cost community itself, there has been a debate around the impact of biases, with some estimators believing that human judgment is necessary for good cost estimation, and others asserting that human judgment should be minimized, and that estimators should rely more on historical data and models built solely on objective parameters.

To shed some light on this issue I decided to do a test. I decided to test cost analysts to see if they could produce better cost estimates using their judgment than could a simple model using only objective parameters. To do this, I needed an approach, a mechanism, a way to gather the data relatively quickly and easily from a large number of analysts and use it to effectively build a comparison.

Some years ago, NASA paid Joe Hamaker to update a model called "QuickCost." The QuickCost models are designed to provide a ROM (rough order of magnitude) cost estimate for a spaceflight hardware system using very little data. In fact, the spacecraft bus model has only eight input parameters to estimate the development and first unit cost of a robotic spacecraft plus its scientific instrument payload. Six of the parameters are objective and three are subjective (meaning they rely on the judgment of the analyst to assign the correct value). Joe delivered the updated model (version 5.0) in February, 2011.

With QuickCost 5.0 I realized I had a way to gather the data needed to test expert judgment. I could query cost professionals for their estimates of the subjective parameter values. I could perform a regression analysis using the same objective parameters as QuickCost to develop an objective parameter only model. I could compare the estimates generated by both models. Thus, armed with a mechanism to test the theory that algorithms are better than humans at cost estimating, I set forth to conduct a trial.

The remainder of this paper documents my actions and the results. I begin with a more detailed explanation of the arguments for and against subjective parameters followed by an outline of the approach. After that is a short description of QuickCost and the objective model. The fun really begins when I discuss the survey and results. Then it's back to number crunching as I analyze the results. Finally, I try to make sense of it all.

In this corner…

Several years ago, after I began reading books and articles on the subjects of behavioral psychology and behavioral economics, I noticed a consistent theme: human judgment is unreliable. Authors such as (Nobel Prize Winner) Daniel Kahneman, Leonard Mlodinow, and Nate Silver all cited numerous research studies and examples where humans made errors in judgement, and not just random errors. These errors were predictable and easily influenced by environmental factors.

What these and several other authors had to say got my attention. Their explanation for human behavior in the areas of judgment and decision making seemed to fit with my own observations. This led to further research and to a paper I wrote titled "The Dangers of Parametrics." In that paper I argued against the use of subjective parameters in cost models, summing up their impact on our profession thusly:

> 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.

Since then I have had numerous discussions over this issue with several of my dearest friends in the cost business. Most think I am wrong. They strongly believe that judgment is so central to the field of space flight system cost estimating and analysis that the use of subjective parameters is absolutely necessary to effectively do our jobs. The following quote from Don Mackenzie, a winner of both the Frieman Award and the ISPA Parametrician of the Year, succinctly captures their argument.

> 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.

So we have two different and diametrically opposed positions. One based on science and derived from academic studies, but also shown to explain the behavior of people and organizations in the real world. The second based on knowledge and experience gained from decades of estimating the cost of unique space flight systems.

One way to decide who is right would be to do two cost estimates of the same system. One using a model that consists only of objective inputs and the other using a model that contains both objective and subjective parameters. These estimates could then be compared to the actual costs. Whichever approach consistently gave the best results would be the winner. As stated in the Introduction, QuickCost 5.0 gave me the cost model around which I could build my experiment.

The Scientists

In 1954 a psychologist at the University of Minnesota, Paul E. Meehl, published a short book titled "Clinical versus Statistical Prediction." In his book, Meehl examined the case for whether clinical assessments (that is, assessments based on knowledge, experience, and expertise) are better than assessments made using statistical methods, or to use his term "actuarial methods." Meehl performed an assessment of twenty studies that contained an empirical assessment of statistical versus non-statistical methods. The quality of these studies varied considerably, and Meehl raises several concerns with the methods used. However, he believed that the results clearly demonstrated the superiority of the statistical methods.

> In spite of the defects and ambiguities present, let me emphasize the brute fact that we have here, depending upon one's standards for admission as relevant, from 16 to 20 studies involving a comparison of clinical and actuarial methods, *in all but one of which the predictions made actuarially were either approximately equal or superior to those made by a clinician*. (Meehl, page 119, emphasis is that of the author)

Daniel Kahneman, in his book "Thinking, Fast and Slow," discusses Meehl's research, the research of other psychologists, and examples such as algorithms developed to predict the value of vintage Bordeaux to illustrate the better predictive power of statistical methods. Kahneman points out that in approximately 200 studies of statistical versus clinical methods, 60% have shown that the statistical methods are more accurate, and that in the other 40% the two methods tied. Since using an algorithm is generally less expensive than employing an expert, the tie means the algorithm is more cost effective.

So why is expert judgment inferior to an algorithm? As humans we are primed to make rapid judgments based on knowledge, experience, and what we can perceive (or easily recall) at that time. For example, our ancestors walking across the African Savanna did not need (or have time) to do a complex multi-variable regression analysis to decide if that rustle meant there was a lion hiding behind a bush. Evolution tends to favor those who make "good enough" decisions quickly rather than those who take time to try and reach an optimal solution, but end up being eaten.

However, a "good enough" decision making process often does not work very well when faced with more modern (and slower) problems like estimating the cost of a technically advanced, complex space system. Our natural human biases have been studied and written about by several leading thinkers such as Kahneman, Nassim Taleb, Nate Silver, and Leonard Mlodinow. These biases and how they impact our profession have proved useful fodder for several of my ICEAA papers, so I will not elaborate on them here, but rather point out some of the more common ones such as the optimism bias, attractiveness, anchoring, confirmation bias, and substitution. If you wish to know more about these biases I will happily provide you a copy of my paper "The Psychology of Cost Estimating" (for free!) or recommend a book by one of the authors listed above.

Thus the problem comes down to this: biased human judgment overrules any attempt we make at objectivity and introduces additional error into the estimate. Our biased judgement is so pervasive that it even affects our ability to properly adjust objective models. Kahneman makes the point that studies have shown that attempts to improve on objective models with human judgment results in worse performance, not better.

> Several studies have shown that human decision makers are inferior to a prediction formula when they are given the score suggested by the formula! They feel that they can overrule the formula because they have additional information about the case, but they are wrong more often than not. (Kahneman, "Thinking, Fast and Slow," page 224)

The Estimators

When I began thinking about doing this experiment, several of my colleagues in the profession were only too happy to share their opinion with me. In fact, they were eager to make sure I knew the importance of having subjective variables in models. Several of them provided me with their rationale, which I will use as the basis for the argument that subjective variables are necessary.

Christian Smart, while acknowledging that subjective variables could be misused, was kind enough to give me a summary of his position why he thinks there is value in subjective variables.

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. This hearkens back to a presentation that Neal Hulkower gave about 10 years ago called "Estimating in the 'Blink' of an Eye," which is the application of Malcolm Gladwell's book Blink to cost estimating.
3. We have small data sets. This makes the use of experience, such as Bayesian methods, very important in producing accurate estimates.
   (Email from Christian Smart to the author, 12/6/2019)

Christian makes three key points that were reinforced by comments provided by others, such as Joe Hamaker, Don MacKenzie and Ron Larson: cost estimating (and cost modeling) is subjective; the data sets for estimating space flight systems are small; and the estimator brings valuable knowledge and experience to the estimate.

Another point made by the estimators focuses on the uniqueness of space flight systems developed for the government. Both NASA and the Department of Defense (DoD) develop cutting edge space systems that push the boundaries of technology. We also rely heavily on a relatively small and specialized industrial base. NASA space science systems are almost always unique, one-of-a-kind systems that may or may not have historical precedence. The estimators uniformly agree that these factors cannot be fully captured by objective parameters and can only be adequately addressed by human judgment.

Joe Hamaker (another Frieman winner) suggests that subjective variables enable the analyst to "tease out" (in his words) important factors that cannot be discerned due to the noise in the data, yet are important to achieving a credible estimate. Joe also believes that subjective variables give the analyst a necessary tool to incorporate system characteristics that are important to a customer, such as complexity or inheritance. Addressing these characteristics in an estimate can make it easier for the customer to accept the results.

In conclusion, the argument of the estimators comes down to this: space systems cost estimating requires both art and science. If you accept that it involves art, then you have to accept the human element, and the best way to capture that human element is through subjective parameters.

Design of the Experiment

The design of this experiment was straightforward and seemed simple enough, at least to my way of thinking. Step one was to find or develop two space system cost models: one with subjective and objective parameters; and the other with only objective parameters. This step was facilitated by Joe Hamaker's QuickCost Model, version 5.0, as discussed in the Introduction.

Step two was to identify the space missions to use in the survey. Since QuickCost was released in 2011, I had a large number of completed NASA science missions and a ready set of unbiased data (the NASA CADRes) to draw on for my test data set.

Step three was to solicit inputs for the subjective parameters for the test data set from fellow space cost professionals. I shamelessly solicited everyone I could think of, and even from people I really didn't know.

Step four was to evaluate the performance of both models against the test data set. At this point I ran wild and indulged in spreadsheet nerdvana. Later we will swim in the sea of statistics together.

The fifth and final step: Determine the winner.

Developing the Models

The advantage of choosing the QuickCost model is that it was already developed. Joe had done all the hard work. QuickCost 5.0 is a single CER built on 131 data points, all robotic space science missions. The model has eight input parameters, including 3 subjective parameters. The QuickCost 5.0 equation form and the input variables are shown in Exhibit 1.

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%

Exhibit 1. QuickCost 5.0.

A second advantage to choosing the QuickCost model is that I could continue to pilfer Joe's labor by using his data set to create the objective parameter model called, imaginatively, the Objective Model. To ensure consistency I used the same objective parameters in the Objective Model (OM) that Joe used in QuickCost. After all, I don't want anyone to accuse me biasing the outcome by using a different parameter set. The Objective Model is shown in Exhibit 2.

LnCost = 1.0186 + 0.6031 LnTotDryMass + 0.1294 LnPower – 0.4970 LnATP + 0.3501 LnLifeMonths + 0.4504 Destination

Exhibit 2. The Objective Model (OM).

There is one small difference between the Objective Model and QuickCost 5.0. When I did the regression for the Objective Model *I did not catch that Joe did not perform a log transform on ATP* (Authority to Proceed). Therefore, the Objective Model uses the log transform of ATP rather than ATP as an input. I thought about rerunning the regression and revising the analysis but decided that I did not have the time to fully investigate the potential differences. If I should do an update to this analysis I can revisit the issue.

I evaluated the quality of the models using a variety of statistics. These are shown in Exhibit 3. By all measures QuickCost 5.0 outperforms the Objective Model. Having more input parameters (8 versus 5) helps. But I also believe that as Joe was developing the model he was using the subjective parameters to account for random noise in the data. I explored this topic in detail in my paper "The Dangers of Parametrics" so will not revisit it here, except to say that subjective parameters have been proven to enable human biases to easily influence the modeling process.

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

Exhibit 3. Model Statistics.

Formulas for selected statistics are as follows:

- 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

To conduct the test I needed a set of robotic space science missions that had been completed and launched after the development of QuickCost 5.0. A true test of the analyst's subjective capabilities, versus the capability of the Objective Model, could only be determined using data from missions that were not in the model data set. Knowledge of a model's ability to reproduce the underlying data set is useful, but is not indicative of its ability to predict the cost of future missions.

Fortunately, NASA has launch several robotic space science missions since 2011. In my zeal to create as much data as possible for my analysis, I selected as many as possible, a total of 15, for the analysis. This would come back to bite me later. The 15 missions are shown in Exhibit 4.

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
WISE	Wide-field Infrared Survey Explorer

Exhibit 4. NASA Robotic Science Missions Selected for the Test Data Set.

To make my job even easier, NASA has created a standard data collection instrument called the CADRe (Cost Analysis Data Requirements). CADRes are developed at each major project milestone, such as PDR, CDR, etc. I used the CADRe data to provide the objective inputs for both models as well as the actual costs for the spacecraft and instrument payloads. No normalization was performed on the cost data except for adjusting for inflation. Therefore, the CADRes gave me a neutral, unbiased data set for comparing the results of the experiment.

The Survey Instrument

A lot of thought and effort went into designing the survey instrument. Since this was the first time I had attempted to do this kind of survey, I wanted to make it as clear and self-explanatory as possible. The survey needed to capture an analyst's estimate of Bus New Design, Instrument New Design, and Instrument Complexity for each of the 15 missions in the test data set. To maintain the connection to QuickCost 5.0 as well as avoid biasing the experiment, I used the definitions for the subjective parameters that were provided in the model documentation.

I developed a preliminary survey instrument in Excel and sent it to Joe and Christian and a few other cost analysts. Based on their feedback, I realized I needed to provide more detailed instructions and to emphasize the need to provide inputs for all of the new design and complexity factors. Finally, I decided to add a few demographic questions so that I could characterize the respondents in terms of age, gender, and experience.

The final survey instrument was three tabs in an Excel workbook. The first tab was instructions. The second tab was the survey. The third tab was the demographic questions. The complete survey instrument is provided in Appendix A.

Conducting the Survey

The next step in the process was to distribute the survey to space cost professionals throughout the world and gather their responses. I sent the survey instrument to fellow NASA cost professionals, cost engineers at other government agencies, contractors, and members of the international space cost community. Some of the people I sent it to were longtime friends and colleagues. Others were contacts gathered from the organizational meetings of the ICEAA Space Special Interest Group (SIG). All in all, I distributed the survey to approximately 50 cost professionals. As the author of the study and the survey instrument, I did not participate.

After distributing the survey I sat back and waited for the responses to pour in. I gave people five or six weeks to respond (plenty of time, or so I thought). After a few responses trickled in not much happened. After a few weeks I sent out a reminder. I got a few responses from people letting me know they would not be able to fill out the survey. Nice of them to tell me but that did not solve my need for data. I began to send out targeted emails to people I know well asking them more pointedly (begging them) to complete a survey. I extended the deadline. And I extended it again. I began to gently remind (beg) members of my staff to take the survey. Eventually I got 10 full responses and a partial response.

So what happened? In retrospect, I realize that the survey was too long. Asking people to provide new design and complexity assessments on 15 robotic space flight missions was just too much. Especially because most of the people I sent the survey too were unfamiliar with these NASA missions. Several respondents told me directly that they could not take the survey because they did not have sufficient knowledge or data. In my naivety I assumed that people would just Google the missions or use other publicly available information sources to learn what they could and take their best guess. What I did not anticipate is that professional integrity would prevent some from completing the survey. They were just not willing to do what they considered to be a mediocre job.

I also suspect that some people did not have the time. 15 missions was probably about 10 too many, maybe even 12 too many. Money may have also played a part. Contractors need to be paid for their time, and this was obviously voluntary. While many of us love our jobs, most of us aren't willing to work nights or weekends on a task that has no direct benefit. It probably goes without saying (but I will say it anyway) that some folks were just not interested.

Interestingly, none of the major proponents of using subjective variables provided me a usable survey. Some forgot, some did not have time, some did not believe they had sufficient data. All of their reasons were valid. Still, it kind of makes me wonder…

Characteristics of the Respondents

As stated previously, I got 10 full responses and 1 partial response. Everyone filled out the demographic part of the survey, so I do have some information about the participants. In terms of location, everyone is a citizen of the United States living in the continental US. Their place of employment breaks down as follows: 3 work for NASA; 7 work for a government services contractor; and one works for a hardware prime contractor. Gender was split nine male and two female.

The age of the participants was fairly evenly spread, as shown in Exhibit 5. No one age group predominated. But there might be a very slight skew towards the higher end of the scale (older).

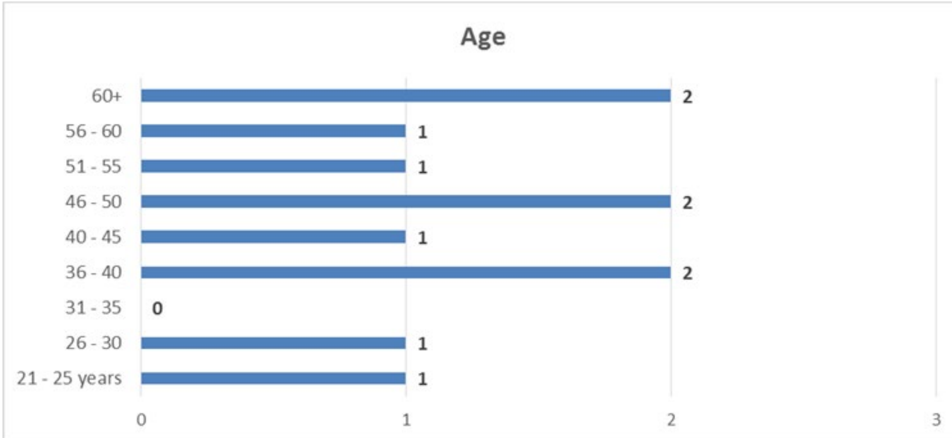

Exhibit 5. Age Distribution of the Survey Respondents

Years in the space business was also fairly evenly spread, as can be seen in Exhibit 6. The largest category was 20+ years, with 3 respondents. However, three other age ranges had 2 respondents and two others had 1 each. So I doubt that experience in the space business had any meaningful impact on the results.

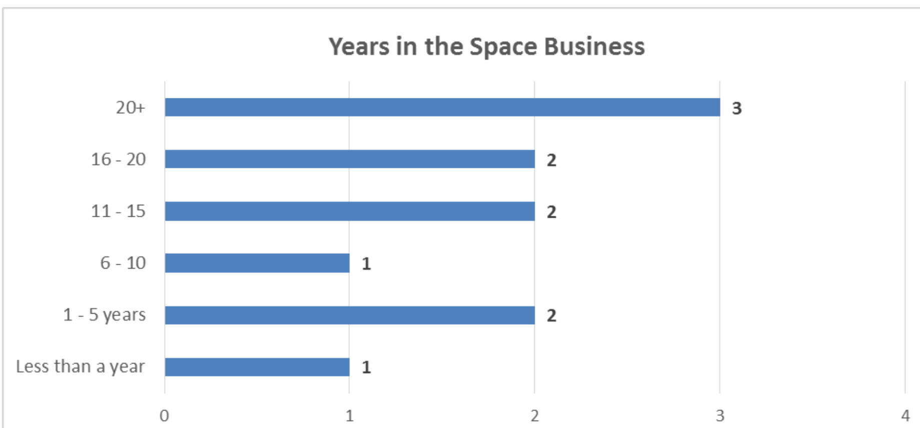

Exhibit 6. Respondents Experience in the Space Business.

One demographic category that might have had some influence is number of years of experience as a cost estimator. Exhibit 7 shows the distribution of the responses. Four out of the eleven participants have 5 or fewer years as a cost estimator, a substantial percentage. That is offset to some degree by the fact that three participants have 20 or more years of estimating experience. The other four fell somewhere in between.

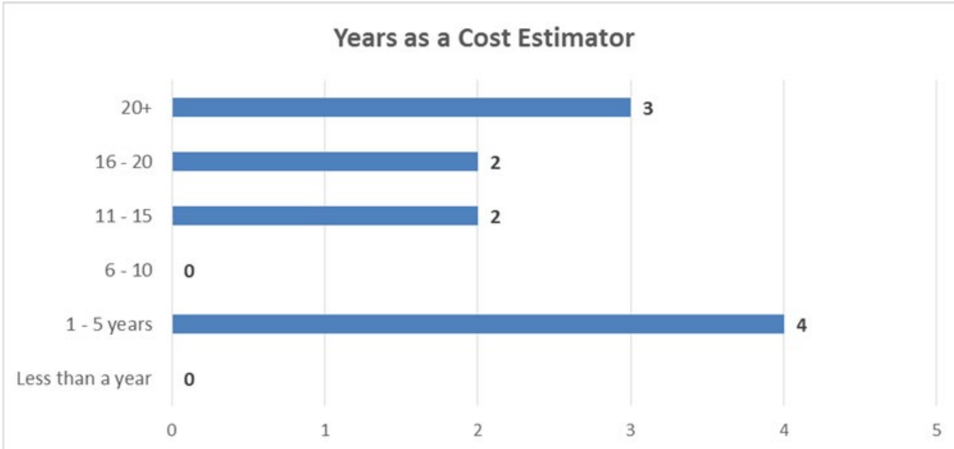

Exhibit 7. Experience Level of the Respondents.

Later I will discuss correlations between the demographic data and the relative accuracy of the estimates.

Organizing the Results

I used the survey responses (the subjective parameter values) as inputs to QuickCost. Thus every survey generated 15 estimates, one for each mission (except for the one survey that only had usable inputs for one mission). I generated statistics measuring the performance of each respondent by copying tabs pre-loaded with formulas from a workbook that I used as a template. The results are saved in Excel workbooks, one for each respondent.

The process was not totally automatic. Some manual steps were required, such as using the Excel Data Analysis add-in to perform the Paired t-Test. Also, because I continued to explore and experiment with different statistics and different ways of analyzing the data, I would periodically have to make changes to each respondent's workbook. But because I had used a standard template I could make changes relatively easily by copying and pasting.

Partially out of curiosity and partially because I had so few responses, I created an artificial response by setting the QuickCost subjective parameters to their average or median values. I say average or median because the guidance provided with QuickCost 5.0 for Satellite Bus and Instrument Suite New Design states that 60% represents an average amount of new design, but; for Instrument Suite Complexity a median value of 50% is recommended as representative of a 50th percentile level of complexity.

For the purposes of the analysis, I treated the QuickCost model with subjective inputs set to their average or median values and the results obtained from running the Objective Model as responses, calculating the same statistics and performing the same statistical tests as I did for the survey participants.

To compare the results by mission I aggregated the subjective estimates, sorted them by mission, and compared the results to the Objective Model results. Because years ago I had read a book years called "The Wisdom of Crowds," I decided to average the estimates from the 11 respondents to see if those averages would provide the best estimates.

Since I had so few participants, I decided to compare each response against the Objective Model. Because I promised anonymity, I have used a coding system to track the responses for those who provided me inputs and for the estimates I developed using the QuickCost averages/medians and the Objective Model. Exhibit 8 gives the coding system.

Participant Code	Key
AA	Responses Provided for all 15 Missions
AB	Responses Provided for all 15 Missions
AC	Responses Provided for all 15 Missions
AD	Responses Provided for all 15 Missions
AE	Responses Provided for all 15 Missions
AF	Responses Provided for all 15 Missions
AG	Responses Provided for all 15 Missions
AH	Responses Provided for all 15 Missions
AI	Responses Provided for all 15 Missions
AJ	Responses Provided for all 15 Missions
BA	Response Provided for 1 Mission
QM	QuickCost Model with Average/Median Subjective Inputs
OM	Objective Model

Exhibit 8. Coding System for Results.

Participant codes beginning with an “A” are for respondents who provided subjective inputs for all 15 missions. The participant coded BA provided inputs for one mission. QM is the code for QuickCost 5.0 using the average/median values for the subjective parameter inputs. OM is the code for the Objective Model.

Defining Victory

In every contest there is some metric used to determine a winner. Runners race to see who will cross the finish line first. Basketball teams vie to score the most points. Golfers seek to complete 72 holes in the fewest number of strokes. But in the world of statistics, it is not the score that matters, it is a method called statistical inference that creates a somewhat hazy metric called *statistical significance*.

Perhaps only in statistical inference does the absolute score not determine the winner. It is not the absolute score that matters, but the magnitude of the difference and the variation in the scores that determine if a difference is truly statistically significant.

Since we live in an age of computerized spreadsheets, generating statistics is easy. All I needed was to find the right statistic, calculate it using the data generated from my survey, and declare a winner. What I wanted was the metric that would provide definitive proof that subjective parameters consistently produce estimates that are Better Than the Objective Model (BTOM).

If there is such a metric, such a statistic, I could not find it.

Ultimately I settled on a hodge-podge of statistics and statistical tests. I calculated the usual descriptive statistics: mean, median, standard deviation, minimum, maximum. I calculated statistics that measured the accuracy of the estimated cost relative to the actual, such as the residual, absolute deviation, deviation ratio, absolute deviation ratio, and root mean squared error (RMSE). I performed three different statistical tests: the Paired t-Test; the Wilcoxon Signed Rank Test; and the van der Waerden Signed Rank Test. I even calculated prediction intervals on the Objective Model estimates to see if I could learn anything from them (I didn't).

I analyzed the data by estimator and by mission. I looked at the effect of averaging the estimates calculated from the subjective inputs. I ran a simulation to see if a room full of monkeys randomly entering subjective values would do better than the Objective Model (just kidding about the monkeys, no animals were harmed in the performance of this analysis).

Now let's look at the results and see what we can learn.

A Sea of Statistics

Honestly, I generated way more statistics than I needed or could evaluate. To give you an example, look at Exhibit 9, which is the basic calculation worksheet for one of the participants.

	FY2020$								
Mission	**Estimated Cost ($M)**	**CADRe Values**	Act-Est	ABS(Act-Est)	(Act-Est)/Act	ABS((Act-Est)/Act)	Within 30%		Res^2
CYGNSS	$66.1	$62.8	-$3.3	$3.3	-5.2%	5.2%	1		$10.62
GPM	$1,180.5	$715.9	-$464.6	$464.6	-64.9%	64.9%	0		$215,831.76
GRAIL	$159.2	$264.4	$105.2	$105.2	39.8%	39.8%	0		$11,074.38
IRIS	$112.7	$83.9	-$28.8	$28.8	-34.3%	34.3%	0		$826.80
JUNO	$1,723.6	$624.2	-$1,099.4	$1,099.4	-176.1%	176.1%	0		$1,208,741.35
LADEE	$79.3	$196.9	$117.6	$117.6	59.7%	59.7%	0		$13,832.10
MAVEN	$461.1	$339.6	-$121.5	$121.5	-35.8%	35.8%	0		$14,753.11
MMS	$714.4	$770.3	$55.9	$55.9	7.3%	7.3%	1		$3,122.07
MSL	$3,776.2	$1,921.2	-$1,855.0	$1,855.0	-96.6%	96.6%	0		$3,441,120.95
NuSTAR	$120.8	$118.5	-$2.3	$2.3	-1.9%	1.9%	1		$5.28
OCO-2	$148.5	$218.3	$69.8	$69.8	32.0%	32.0%	0		$4,867.89
OSIRIS-REx	$805.8	$478.1	-$327.7	$327.7	-68.5%	68.5%	0		$107,380.03
THEMIS	$256.7	$110.5	-$146.2	$146.2	-132.3%	132.3%	0		$21,380.48
VAP	$510.1	$384.0	-$126.1	$126.1	-32.8%	32.8%	0		$15,907.51
WISE	$281.1	$198.7	-$82.4	$82.4	-41.4%	41.4%	0		$6,783.22
Total	$10,396.0	$6,487.3	-$3,908.7	$4,605.7			3		
Average	$693.1	$432.5	-$260.6	$307.0	-36.7%	55.2%		RMSE =	$581.13
Median	$281.1	$264.4	-$82.4	$117.6	-34.3%	39.8%			
Standard Deviation	$941.0	$455.0	$519.4	$493.4	62.2%	46.5%			
Minimum	$66.1	$62.8	-$1,855.0	$2.3	-176.1%	1.9%			
Maximum	$3,776.2	$1,921.2	$117.6	$1,855.0	59.7%	176.1%			

Exhibit 9. Basic Calculation Worksheet.

Starting with the left-hand column are the 15 missions used for the survey. The next two columns are the estimated costs using the participant's subjective inputs and the actual costs derived from the CADRes. These are followed by various calculations: residuals (actual cost minus the estimated cost); absolute value of the residuals; the deviation ratio ((actual – estimate)/actual); and the absolute value of the deviation ratio. The "Within 30%" column determines the number of estimates that fall within plus/minus 30%, a somewhat arbitrary metric. The final column is for the calculation of the root mean squared error (RMSE). These same calculations were done for all 10 responses, the Objective Model, and QM.

I calculated the mean (average), median, standard deviation, minimum, and maximum values for all of the residual-based statistics. I also calculated totals for the estimated cost, CADRe cost values, residuals, and absolute value of the residuals. Add to this the three statistical tests I did, and you can begin to understand the challenge of figuring out what is meaningful and what is not.

To keep the assessment manageable, I looked at the mean values of the descriptive statistics and two of the statistical tests. The summary results are shown in Exhibit 10.

Code	Mean Residual	Mean Absolute Deviation	Mean Absolute Deviation Ratio (%)	RMSE	Paired t-Test	Wilcoxon Signed-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
OM	**-$2.2**	**$162.0**	**37.3%**	**$261.4**	**0.9748**	**0.305**

Exhibit 10. Summary of Descriptive Statistics and Inference Tests.

The bottom row of Exhibit 10 gives the statistics for the Objective Model. The mean of the residuals for the Objective Model is -$2.2, meaning that for this data set the model is close to an unbiased estimator. None of the QuickCost estimates performed better, though participant AH came close. Looking at the next three statistics (Mean Absolute Deviation, Mean Absolute Deviation Ratio, and RMSE) one participant (AH) and QM both performed better than the Objective Model. Participant AJ had a better Mean Absolute Deviation and RMSE than OM. Note that statistics that are better than OM are highlighted in green.

The statistical inference tests measure the behavior of the estimated cost relative to the actual cost. In other words, if you were to rank the missions using their actual cost from low to high, then rank the missions using the estimated cost from low to high, how closely would those rankings agree? They are a measure of consistency. There results for the Objective Model are significant. The values highlighted in red are participant responses that fail to meet a 5% threshold for significance. Five of the 11 participants (counting QM as participant) fail to meet the 5% threshold for the Paired t-Test and seven fall below that threshold using the Wilcoxon Signed-Rank Test.

Wanting to put to use the demographic data I collected, I calculated correlations between the RMSE for each participant and years as an estimator, years in the space business, and age. Age turned out to have the highest correlation at -0.40, not very strong but it could mean that as we get older we become slightly better estimators. Years as an estimator and years in the space business both had weak correlations, -0.33 and -0.25 respectively.

A different way of analyzing the data is to evaluate the results mission by mission. Exhibit 11 is an example of how I did that, using the CYGNSS mission as an example.

Code	Mission	Estimated Cost ($M)	CADRe Values	Act-Est	ABS(Act-Est)	(Act-Est)/Act	ABS((Act-Est) /Act)	Better than OM
OM	CYGNSS	$95.5	$62.8	-$32.7	$32.7	-52.1%	52.1%	
AA	CYGNSS	$105.9	$62.8	-$43.1	$43.1	-68.7%	68.7%	0
AB	CYGNSS	$85.8	$62.8	-$23.0	$23.0	-36.6%	36.6%	1
AC	CYGNSS	$117.7	$62.8	-$54.9	$54.9	-87.4%	87.4%	0
AD	CYGNSS	$55.3	$62.8	$7.5	$7.5	11.9%	11.9%	1
AE	CYGNSS	$66.1	$62.8	-$3.3	$3.3	-5.2%	5.2%	1
AF	CYGNSS	$118.1	$62.8	-$55.3	$55.3	-88.1%	88.1%	0
AG	CYGNSS	$107.9	$62.8	-$45.1	$45.1	-71.8%	71.8%	0
AH	CYGNSS	$55.8	$62.8	$7.0	$7.0	11.1%	11.1%	1
AJ	CYGNSS	$110.2	$62.8	-$47.4	$47.4	-75.5%	75.5%	0
AI	CYGNSS	$77.6	$62.8	-$14.8	$14.8	-23.5%	23.5%	1
Average		$90.0		-$27.2	$30.1	-43.4%	48.0%	
Median		$95.9		-$33.1	$33.1	-52.7%	52.7%	
Standard Deviation		$23.8		$23.8	$20.0	37.9%	31.9%	
Minimum		$55.3		-$55.3	$3.3	-88.1%	5.2%	
Maximum		$118.1		$7.5	$55.3	11.9%	88.1%	
Total								5
AVG	CYGNSS	$90.0	$62.8	-$27.2	$27.2	-43.4%	43.4%	1

Exhibit 11. Analysis by Mission Example.

You can see that I calculated many of the same descriptive statistics as was used for the participants. One addition is a count of how many participants performed better than the OM based on the absolute deviation ratio. I also took the average of the participants and performed the same statistical tests that I did for the individual statistics. Those results will be discussed below.

Exhibit 12 summarizes the residuals by mission by participant. Results highlighted in green are where the QuickCost based estimate was closer to the actuals than the Objective Model. While the table shows residuals, BTOM was calculated on the absolute deviation. Notice that participant BA now makes an appearance.

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
OM	**-$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**

Exhibit 12. Results by Mission and by Participant.

If you look at the results in Exhibit 12 column by column, you will see missions where most of the participants did better than OM. Looking row by row you may notice that there are participants that did better than OM across most of the missions. To get a better sense of how the QuickCost estimates were faring against the Objective Model, I summed the number of BTOM estimates by participant, by mission, and in total. Exhibit 13 shows the results by participant and by mission.

Code	BTOM
AA	6
AB	5
AC	2
AD	8
AE	4
AF	4
AG	4
AH	8
AI	8
AJ	5
BA	0
QM	7

Mission	BTOM
CYGNSS	5
GPM	1
GRAIL	3
IRIS	1
JUNO	4
LADEE	4
MAVEN	4
MMS	4
MSL	7
NuSTAR	3
OCO-2	3
OSIRIS-REx	7
THEMIS	2
VAP	6
WISE	7

Exhibit 13. BTOM by Participant and By Mission.

Looking at the results by participant, 3 participants (AD, AH, AI) outperformed OM slightly more than half the time (8 out of 15). QM did better than OM for 7 missions out of 15 or slightly below 50%. The rest were all well below 50%. Examining the results by mission, there are four where QuickCost stands out as getting BTOM greater than 50% (7 out of 11): MSL; OSIRIS-REx; VAP; and WISE. One other mission, CYGNSS, comes close to 50%. However, there are others (GPM, IRIS, THEMIS) where the QuickCost based estimates did poorly.

The takeaway here is that some estimators did beat the Objective Model about half the time and that for some missions the estimators got overall better results with QuickCost than with the Objective Model. But when looking at the results in aggregate it is apparent that the Objective Model yields consistently better estimates. Look below at Exhibit 14. Summing the BTOM results across all the participant estimates give a BTOM rate of 36%, which means that 64% of the time the Objective Model gave the better estimate. Adding in the results from the QM estimates increases the BTOM rate slightly to 37%.

	11 Responses	Plus QM
Total Number of Responses:	151	166
Better Than Objective Model	54	61
BTOM Rate:	36%	37%

Exhibit 14. BTOM Results for All Responses.

At this point you might be asking "What about the averaging of the estimates across all the participants, how did that turn out?" Well, it turns out that averaging the responses by mission yielded 7 BTOM estimates but poor statistics. For example, the mean residual was -$191.9M versus -$2.2M for the Objective model, and both the Paired t-Test and the Wilcoxon Signed-Rank tests were less than 5% significant.

Are Monkey's Better Cost Estimators?

There is a mathematical postulate known as the Infinite Monkey Theorem. The Infinite Monkey Theorem states that given a near infinite number of monkeys armed with typewriters and typing for a near infinite period of time, the monkeys would produce lots of gibberish, but they would also reproduce great works of literature like Shakespeare's *Hamlet*. Knowing this theorem I could not help but put it to the test, except in this case the monkeys would be producing subjective input values for QuickCost. Could monkeys outperform the humans or maybe even the Objective Model?

By the way, before you mathematical purists out there get all riled up, the Infinite Monkey Theorem as applied to great works of literature (or even romance novels) has been totally debunked. While you can show that given enough monkeys and enough time, there is a reasonable probability that they will reproduce recognizable words, the probability of reproducing an entire work of Shakespeare is so small as to be practically zero. However, imagining monkeys producing QuickCost inputs is more fun than trying to imagine a random number generator doing the same. For more, see https://en.wikipedia.org/wiki/Infinite_monkey_theorem.

I constructed a Monte Carlo simulation model using uniform distributions for the subjective variable inputs. The minimum and maximum values are based on the QuickCost 5.0 instructions: 20% to 130% for Spacecraft and Instrument Suite New Design; and 0% to 100% for Instrument Complexity. I correlated the subjective variable inputs using correlation coefficients calculated from the values Hamaker assigned to the mission in the QuickCost data set. These are shown below:

- 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

Since I wanted to know if the monkeys could produce BTOM results, I calculated the absolute value of the residual for each mission for each trial. I set the software to run 5000 trials (because 1000 is never enough and 10,000 seemed like too many) and hit start. After I ran the simulation, I used the Objective Model absolute residual for each mission to determine the percentage of monkeys that were BTOM. The results are shown in Exhibit 15.

Mission	Objective Model ABS(Residual)	Monkey BTOM 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%**

Exhibit 15. Performance of the Monkeys Relative to the Objective Model.

Bottom line: the monkeys did better than the Objective Model one-third of the time. There were three missions were the monkeys outperformed the Objective Model more than half the time: JUNO; LADEE; and OSIRIS-REx. This may be due to particularly poor estimating performance by the Objective Model for those missions or it may be the QuickCost is just well suited for them. For two of the three missions (JUNO and OSIRIS-REx) the QM model also performed BTOM. But QM performed better than OM on five other missions, while the monkeys didn't. Interesting, but not enlightening.

Note that the monkeys did only slightly worse than the humans on average (33% versus 36% BTOM). I considered doing a significance test to determine if the results are truly different, but I did not want to embarrass my participants. So for now, let's just say that humans are better estimators than monkeys.

Picking the Winner

By almost any measure, the Objective Model yielded estimates that were consistently better than the estimators on the 15 missions in our data set (I will address the *almost* a little later). Only one participant had portfolio statistics that were better than the Objective Model, and one other could be considered a tie (see Exhibit 10). Of the 151 estimates performed, the estimators were able to improve on the Objective Model only 36% of the time. Which means that 64% of the time the Objective Model gave a better estimate with less work.

The monkey test was illuminating. 33 percent of their estimates bettered the Objective Model, which could be treated as a baseline against which we could evaluate human performance. By that measure, of the 10 participants who provided inputs all 15 missions, 4 did worse than the monkeys and two tied. That means only 40% of the participants could do better than the monkeys at bettering the Objective Model.

QuickCost 5.0 with average/median settings did very well. The overall results were good enough that I consider it to have tied with the Objective Model. This illustrates the point made by Kahneman and others that sometimes adding human judgment to an objective assessment actually degrades the results.

So given these results, are there situations where an estimator can use subjective variables to improve an estimate? The data is inconclusive, but it does offer some hints. Three of the participants achieved BTOM results on 8 of the 15 missions. One of these participants, AH, also had portfolio statistics that were better overall than the Objective Model. I have it on good authority that AH is actually a composite participant. In other words, AH is really several individuals, each one a highly experienced cost estimator, all of whom had deep knowledge and insight into the missions. As Malcolm Gladwell points out in his book "Blink," there is value to expertise. The performance of AH shows that a model with subjective inputs could be valuable when used by knowledgeable experts.

There also can be situations where subjective parameters can overcome model inadequacies. Because of the unique hardware requirements of some NASA missions, our historical data can be a less than perfect analog for the system we are estimating.

The best illustration of this point is the Mars Science Lander (MSL), later renamed Curiosity. MSL is the largest rover ever landed on a planetary body. It is powered by a nuclear power source. Landing MSL on Mars required the development of a new system called the "Sky Crane," which had never been used before. The Objective Model had no direct means to address MSL's complexity, but 7 of the 10 participants were able to use the subjective parameters in QuickCost to estimate a cost closer to the actuals.

There are two other missions where the participants did noticeably better than the Objective Model. In the case of one mission, OSIRIS-REx, the Objective Model missed it so badly there was a lot of room for improvement. For this mission, the better participant performance may be a case of random luck rather than skill. The other mission where overall the participants did better was WISE. WISE was an infrared (IR) telescope survey mission with a hydrogen-filled cryostat. The Objective Model estimate was reasonably close, a 30.4% absolute value ratio, yet 6 of the 10 participants were closer to the actual cost. I can find not sound explanation for why the participants did better. It might be random noise, or it might be that the participants were better able to adjust for the complexities of a cryogenically cooled IR telescope. I just don't know.

Guidance on Using Subjective Parameters

I want to set down rules or guidelines for using subjective parameters. As with any rule or guideline in cost estimating, these should be applied using the analyst's best judgment.

First, when in doubt (and you should be in doubt 90% of the time), don't use subjective parameters in a cost model. These parameters have a high potential for abuse. As the manager of a cost office with a staff of estimators, I have seen numerous estimates where the analyst used subjective parameters (in some cases a simple scalar multiplier) to adjust an estimate to meet customer expectations. Sadly, I must admit that I am also guilty of having done the same thing.

However, subjective parameters can add value if the estimator has in-depth knowledge of the system being estimated and significant estimating experience. Just beware, *most people overestimate their knowledge and experience.*

Second, you should avoid introducing overt subjectivity into an estimate without credible, supportable, and defendable logic and justification. Subjectivity in cost modeling and cost estimating is impossible to avoid. After all, analyzing data for modeling often requires making judgments, and choosing to use a model without subjective parameters is also a subjective assessment. If you choose to incorporate subjectivity more directly, you must recognize the potential for increasing error. See the paragraph in the Conclusions section on being humble.

Be aware that a subjective parameter can be used against you. Smart managers will get the right people in the room to argue with your subjective assessments. I have had that experience, and I suspect you have too. Having an estimate using only objective parameters is an easy way to give an alternative answer that is more difficult to criticize.

If you decide to use subjective parameters, you should calibrate and validate whenever possible. Rely on historical experience to establish reasonable values for subjective parameters by calibrating a model to actual costs. This gives you an anchor that is defendable. And always validate your estimate against historical experience. I have had my greatest impact on NASA decision-makers by showing how my estimates were closely aligned with historical experience whereas the project's estimates were not.

Third, subjective parameters might be valuable when estimating a system outside the model's underlying experience base. In the space business historical data is limited. It is not unusual for a new system to be unlike anything we have flown before or for the analogs to be limited to one or two missions. Subjective adjustments could overcome a model's shortcomings an enable an estimate. MSL is more complex than the missions in the QuickCost database and the estimators generally were BTOM.

Fourth, subjective parameters could be useful when doing a comparative analysis, where relative cost is more important than absolute cost. I have not addressed the subject previously in this paper, but I have anecdotal evidence from experienced cost estimators where they were able to use subjective parameters as part of a comparative cost analysis for good effect. If you are using subjective parameters as part of a comparative analysis, be sure that your knowledge of each system is equivalent and that you are applying the subjective criteria consistently.

Conclusions

I get it. I have been in this business for over 35 years. Subjective parameters make it easier to tell our story, they make it easier to show our customers that we are listening to them, and they assuage our egos by making us look like experts. *Subjective parameters may make our job easier, but they do not automatically make our estimates better.* Interestingly, they also increase our confidence in our estimates. But the fact is that except under certain special conditions, they do not increase accuracy.

The less information you have, the more you must rely on objective models. One of the surprising things about human thinking is that the less we know, the more we think we know. Our minds do a great job of filling in the holes in our knowledge with relatable, but ultimately extraneous information. Thus, we think we know more than we really do. Recognize when you are in a low information situation and avoid using subjective parameters; or if you cannot avoid them, choose the median or average values. You will get more reliable results.

Be humble about your knowledge, expertise, and ability to be objective. Research shows that most of us overestimate our knowledge and experience and underestimate our biases. It is human to believe we are above average and capable. It is also human to be swayed by appearances or a good story or simply being told the cost needs to fit within a certain budget allocation. Be aware, be very aware that you are human.

Doing a survey is hard. Two big mistakes I made was in asking for too much (as in too many missions) and in not recognizing the integrity of my fellow cost professionals (do it right or don't do it at all). I understand now why so many behavioral studies have been done using college students. They are readily accessible to academics, they respond to small inducements (Hershey's Kisses for example), and being college students, they don't know what they don't know.

Finally, let me close by saying that I know there is subjectivity in cost estimating. It cannot be avoided. However, we must understand that subjective parameters are a tool that can all too easily give us less reliable estimates while simultaneously making us feel better about the results. Recognize this and use them always with great care and caution.

Bibliography

Ariely, Dan, *Predictably Irrational*, Revised and Expanded Edition, New York: Harper Perennial, 2009

Aschwanden, Christie, "Your Brain is Primed to Reach False Conclusions." *fivethirtyeight.* February 17, 2015. http://fivethirtyeight.com/features/your-brain-is-primed-to-reach-false-conclusions/

Hamaker, Joseph W., "QuickCost 5.0 Final Report," Huntsville, Alabama: SAIC: May 2011

Hubbard, Douglas W., *How to Measure Anything*, New Jersey: John Wiley & Sons, 2010

Hubbard, Douglas W., *The Failure of Risk Management,* New Jersey: John Wiley & Sons, 2009

Kahneman, Daniel, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011

Meehl, Paul E., *Clinical versus Statistical Prediction,* Reprint published by Echo Point Books & Media, 2013

Mlodinow, Leonard, *The Drunkards Walk: How Randomness Rules Our Lives*, New York: Pantheon Books, 2008

Mooney, Chris, "The Science of Why We Don't Believe Science." *Mother Jones.* May/June 2011. http://www.motherjones.com/politics/2011/03/denial-science-chris-mooney

Nuzzo, Regina, "How scientists fool themselves – and how they can stop." Nature. October 7, 2015. http://www.nature.com/news/how-scientists-fool-themselves-and-how-they-can-stop/

Prince, Frank, "The Dangers of Parametrics," *Proceedings of the 2016 International Cost Estimating and Analysis Association Professional Development and Training Workshop,* Atlanta, GA, June, 2016

Prince, Frank, "The Psychology of Cost Estimating," *Proceedings of the 2015 International Cost Estimating and Analysis Association Professional Development and Training Workshop*, San Diego, CA, June, 2015

Siegel, Eric, "The One Rule Every Data Scientist (and Manager) Should Know By Heart," GovExec.com, December 21, 2015. <http://www.govexec.com/technology/2015/12/oneruleeverydatascientistandmanagershouldknowheart/124803/print/>

Silver, Nate, *The Signal and the Noise: Why most Predictions Fail but some Don't*, New York: The Penguin Press, 2012

Smart, Christian, private correspondence, December 2019

Taleb, Nassim Nicholas, *Antifragile, Things that Gain from Disorder,* New York: Random House, 2012

Taleb, Nassim Nicholas, *The Black Swan: The Impact of the Highly Improbable,* New York: Random House, 2010

Appendix A
QuickCost 5.0 Subjective Parameter Survey Instrument

Survey Instructions

Instructions for the New Design and Complexity Survey

The purpose of the survey is to compare two cost models. Both models use the same set of objective parameters which include mass, power, date development started, and design life. What distinguishes the two models is that one uses expert opinion inputs in the form of a Satellite Bus New Design factor, an Instrument Suite New Design factor, and an Instrument Suite Complexity factor. The other model only uses the objective parameters. Your job is provide the expert opinion inputs per the instructions below. If you have any questions you can call me at 256-544-8360.

Thank you for taking the time to complete this survey!

Sincerely,
Andy Prince

How your data will be used:
Your data will be used in a study to determine if a model that captures expert opinion in the form of subjective input parameters can outperform a simple model using only objective parameters. The aggregated results of the study will be reported in public forums. **Your participation in this study and your results will be considered confidential and not for public release without your permission.** Specific results may published using a code in place of an individual's name (i.e. Participant A1).

Important Notes:
The goal of the study is to advance the cost estimating and analysis profession. Your participation is voluntary. This study is not endorsed by the National Aeronautics and Space Administration (NASA).

Instructions for the Expert Opinion Factors

Provide an assessment of Satellite Bus New Design, Instrument Suite New Design, and Instrument Suite Complexity for 15 NASA science missions. You are to rely on your judgment as a cost estimator and whatever data sources you have available to you. ***Please provide an input for all factors!*** The 15 science missions are as follows:

Survey Instructions (Continued)

Mission	
CYGNSS	Cyclone Global Navigation Satellite System
GPM	Global Percipitation 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
WISE	Wide-field Infrared Survey Explorer

The specific scales for each factor are as follows:

Satellite **Bus New Design** factor in percentile terms. Consider the following guidelines:

- 20% of totally off-the-shelf
- 60% average
- 100% all new
- 130% (or more) for all new and pushing state-of-the-art

Instrument suite New Design factor in percentile terms with the same scale as bus (but instruments typically have less heritage or higher new design factors than buses)

Instrument Complexity in percentile terms and representing a weighted average of the entire instrument suite.
For example if the instruments are of median complexity, 50% is entered.
Instruments that are judged to be around the 75th percentile of complexity would be entered as 75%.
Scale is 0% to 100%

Subjective Parameter 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 provide an input for all factors! Use your best judgment, nobody is being graded.						
	When finished please answer the questions in the "Demographics" tab						
Mission	**Bus New Design**	**Instrument New Design**	**Instrument 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						

Demographic Questions

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	