Blind Study Validating Parametric Costing Tools PRICE TruePlanning and SEER-H for NASA Science Missions

Paul D. Friz* and Jordan J. Klovstad†
NASA Langley Research Center, Hampton, VA, 23681, USA

Benjamin B. Leser‡
University of Michigan, Ann Arbor, MI, 48109, USA

Benjamin C. Towle§
University of Michigan, Dearborn, MI, 48128, USA

Serhat Hosder¶
Missouri University of Science and Technology, Rolla, MO 65401, USA

Two of the primary parametric costing tools used to estimate the development and production cost of future spacecraft hardware are PRICE TruePlanning - Space Missions Catalog by PRICE Systems, and SEER-H by Galorath. These are standard tools used by NASA and industry to estimate the cost of new aerospace hardware. However, no independent verification of the accuracy of these tools is publicly available. Both PRICE Systems and Galorath have completed internal validation studies of their parametric cost estimating tools; however, they only provided the results of the studies and did not detail the exact methods used to perform the validation. In the present study, cost estimators used PRICE TruePlanning and SEER-H to estimate the cost of twelve different past NASA science missions. The estimators were prevented from knowing the actual cost of the missions in an effort to minimize cognitive biases. In the present study, SEER had an average error of 23%, median error of -0.3%, with a standard deviation of 43%. PRICE had an average error of 52%, median error of 50%, and standard deviation of 45%. There were several factors independent of PRICE and SEER which may have affected the accuracy of the results in the present study including: uncertainty in the technical data used for the estimates, the methods used to estimate uncertainty in spacecraft component mass and numbers of prototypes, and the experience of the estimators.

Nomenclature

C_A	=	Actual Cost
C_E	=	Estimated Cost
C&DH	=	Command and Data Handling
CADRe	=	Cost Analysis Data Requirement
CDF	=	Cumulative Distribution Function
CDR	=	Critical Design Review
CER	=	Cost Estimating Relationship
EOM	=	End Of Mission
EOS	=	Electro-Optical System
FPGA	=	Field Programmable Gate Array
ε	=	Percent Error
GN&C	=	Guidance, Navigation, and Control

*Student Trainee (Engineering), Systems Analysis and Concepts Directorate, AIAA Student Member.
†Aerospace Engineer, Systems Analysis and Concepts Directorate.
‡Student Engineer, Systems Analysis and Concepts Directorate.
§Student Engineer, Systems Analysis and Concepts Directorate.
¶Associate Professor of Aerospace Engineering, Department of Mechanical & Aerospace Engineering, Associate Fellow AIAA.

IAT	=	Integration, Assembly, and Test
IC	=	Integrated Circuit
LRR	=	Launch Readiness Review
MCR	=	Mission Concept Review
MEL	=	Master Equipment List
ONCE	=	One NASA Cost Engineering
PDF	=	Probability Density Function
PM	=	Project Management
RCS	=	Reaction Control System
RTG	=	Radioisotope Thermoelectric Generator
S&MA	=	Safety and Mission Assurance
SE	=	Systems Engineering
SEER-H	=	System Estimation and Evaluation of Resources-Hardware
TCS	=	Thermal Control System
WBS	=	Work Breakdown Structure

I. Introduction

PARAMETRIC costing tools are commonly used both at NASA and in industry to estimate life cycle costs of space systems. Parametric tools allow users to quickly estimate the costs of a mission concept before detailed designs have been completed. Frequently, they are used during Phase A or Pre-Phase A of a mission to perform trade studies and predict costs of design variations. They are also frequently used for independent cost estimates to evaluate mission proposals. Two of the primary costing tools used by NASA to evaluate mission, spacecraft, and instrument proposals are PRICE TruePlanning - Space Missions Catalog by PRICE Systems and SEER-H by Galorath, henceforth referred to as PRICE and SEER[1]. These are standard tools used both by NASA and industry, however, there is little to no independent verification of the accuracy of these tools publicly available. Both PRICE Systems and Galorath have completed internal validation studies of their parametric cost estimating tools; however, they only provide the results of the studies and do not detail the exact methods used to perform the validations. It is not known if the estimators from Galorath and PRICE Systems performed these studies in a blind fashion, or if they had access to the final cost information before making their estimates. The primary goal of the validation studies performed by PRICE Systems and Galorath appears to have been to tune PRICE and SEER, and provide users with standard inputs and assumptions which improve the accuracy of estimates.

Galorath's validation study of SEER consisted of fifteen case studies of various robotic NASA spacecraft, and included modeling 46 instruments[2]. Both Galorath's study and the present study make use of the Electro-Optical System (EOS) and Integrated Circuit (IC) plug-ins for SEER. Galorath provides little information on how the study was conducted other than the names of the missions used and that "A standardized modeling approach was utilized, which formed the basis of a Space Guidance document to be released in the future." The SEER Space Guidance document has been released and was used to form the SEER estimates in the present study[3]. Of the fifteen missions in the SEER study, seven were "Discovery," six were "Explorer," and two were "New Frontiers." "Discovery" class missions are low-cost solar system exploration missions typically cost capped at around $300 million. "Explorer" class missions are even lower cost and are typically capped at $200 million. "New Frontiers" missions are larger than "Discovery" and "Explorer" missions, but are not as expensive as flagship missions. "New Frontiers" missions typically are cost capped under $1 billion. Galorath's study found that over the fifteen missions SEER's average error in predicting cost was -1% with a standard deviation of 19%. Therefore, roughly 68% of all SEER estimates will be from -20% to +18% of the actual costs and 95% of estimates will be from -39% to +37%, assuming the mean values of SEER estimates follow a Gaussian distribution. These relatively low error levels may imply that the estimators might have known the mission costs before performing their estimates so that they could provide guidance to customers on how to adjust inputs.

PRICE Systems has also performed an internal validation study of their tool PRICE TruePlanning[4]. Their study included thirteen NASA robotic spacecraft, ten of which were also in the SEER study. Of the thirteen missions included in the study, four are "Discovery," five are "Explorer," three are "New Frontiers," and one is a heliophysics satellite that does not fall into the other classes. PRICE Systems validation study found that PRICE TruePlanning's average error was +1% with a standard deviation of 13%, meaning roughly 68% of all TruePlanning estimates will be from -12% to +14% of the actual costs and 95% of all estimates will be from -25% to +27% of the actual costs, assuming the mean values follow a Gaussian distribution. As with Galorath's study, it is not clear if the estimators at PRICE Systems knew the mission costs before using the tool to estimate the mission costs.

Studies, like those produced by Galorath and PRICE Systems, to determine the tool's inputs and settings to obtain the most accurate results are very useful to the user. The goal of the present study was to independently assess the accuracy of PRICE and SEER in an environment that matches that of an independent cost estimate as closely as possible. This required that the estimators have no prior information of the mission cost, so that cognitive biases could be minimized. Section II describes the methodology of the present study to evaluate PRICE and SEER. Section III presents the results of the present study, Sec. IV discusses factors which may have affected the results of the present study that are independent of PRICE and SEER, and Sec. V provides the conclusions of the present study.

II. Methodology

For the present study, twelve missions were selected from the One NASA Cost Engineering (ONCE) database. ONCE is a database maintained by NASA, which stores technical and managerial documents presented at reviews for NASA missions. These documents include Master Equipment Lists (MELs) for each spacecraft, technical descriptions of the spacecraft components, and records of how much money was spent on the mission. This information is captured in documents know as Cost Analysis Data Requirements (CADRes). CADRes are a set of three documents recording:

1) An overview of the mission.
2) A mass and power breakdown.
3) A cost breakdown by subsystem and year.

They are usually generated at each major review for a mission; such as, the Mission Concept Review (MCR), Critical Design Review (CDR), Launch Readiness Review (LRR), End of Mission (EOM), and potentially others. The present study used the technical data, MELs hardware descriptions etc., from the CDR to generate cost estimates and compared them with their actual costs at either LRR or EOM, depending on which data set was available. This is significantly different from the validation studies performed by PRICE Systems and Galorath, where LRR/EOM technical data was used to generate estimates that were compared to the same LRR/EOM costs. The present study used the data from CDR instead of LLR/EOM in order to test the predictive capabilities of PRICE and SEER. Additionally, CDR technical data typically goes into more detail than any other review during the mission. This was also done to review the standard mass margin estimation assumptions discussed in Sec. IV.A.

The present study was conducted in the following manner. Cost estimator A removed all references to cost from the CADREs and supporting technical documents. These cleansed documents were then given to estimators B and C who used them to produced PRICE TruePlanning and SEER cost estimates. Estimators B and C were blind to (i.e., unaware of) the actual mission costs in order to prevent cognitive biases, such as anchoring, from influencing their cost estimates. Anchoring is a cognative bias that can cause an individual to rely too heavily on an initial piece of information when making decisions. The estimates included the costs of Project Management (PM), Systems Engineering (SE), Safety and Mission Assurance (S&MA), Payload, Spacecraft Bus, and Systems Integration and Test (IAT), which correspond to the standard NASA WBS items 1-3, 5, 6, and 10, respectively. Estimates of WBS 4, 7-9, and 11: Science/Technology, Mission Operations, Launch Vehicle, Ground Systems, and Education and Public Outreach respectively, were not included in this study. Additionally, no software costs were included in this study. The costs of the Spacecraft Bus were broken down into the following subsystems: structures, thermal control, propulsion, Guidance, Navigation, and Control (GN&C), communications, electrical power, harness, and Command, and Data Handling (C&DH). The majority, but not all, of the missions have their costs broken down in this manner, with the exceptions being attributed to data limitations.

The cases were selected from recent robotic NASA science missions which have complete CADRes on the NASA ONCE Database and sufficient supporting technical documentation to build a credible estimate. Not all CADRes include technical details about spacecraft instruments, sometimes instruments are donated, paid for by other space agencies, or by universities. All the spacecraft included in this study carried multiple instruments but not all missions had enough supporting documentation to build credible estimates of the instrument costs. For missions where insufficient technical detail existed to model the instruments in PRICE and SEER the instrument cost estimates were omitted. The selected missions are DAWN, MESSENGER, MAVEN, GRAIL, New Horizons, SMAP, CONTOUR, WISE, IBEX Juno, Deep Impact, and Kepler.

To ensure that proper comparisons were being made with the actual mission costs all cost estimates were converted to real-year dollars using the 2017 NASA New Start Inflation Index. As the costs of various payloads/subsystems are sometimes sensitive/proprietary, all costs presented in the present paper have been normalized.

There are three main factors that determine the accuracy of a spacecraft hardware cost estimate using parametric cost estimating tools:

1) The accuracy and precision of the tool used to model the cost,
2) The quality and quantity of the technical data describing the hardware,
3) The knowledge, experience, and skill of the cost estimator using the cost modeling tool and evaluating the data.

The goal of the present study is to test the first factor, while minimizing the contributions of the second and third factors. Both the present study, and the validation studies performed by Galorath Inc. and PRICE Systems LLC, use CADRes and supporting data from ONCE. However, for many missions there are a large number of technical documents with varying detail so there may be some slight differences in the exact documents used to build the estimates. It is worth noting that although the CADRes contain significant quantities of data they do not contain every single piece required to complete a parametric cost estimate. Inevitably, the estimators were required to interpret limited data or make assumptions where data was missing.

III. Results

Since no actual costs are presented all results are expressed as a percent error. The percent error of each estimate is calculated using Eq. 1.

$$\varepsilon = \frac{(C_E - C_A)}{C_A} \tag{1}$$

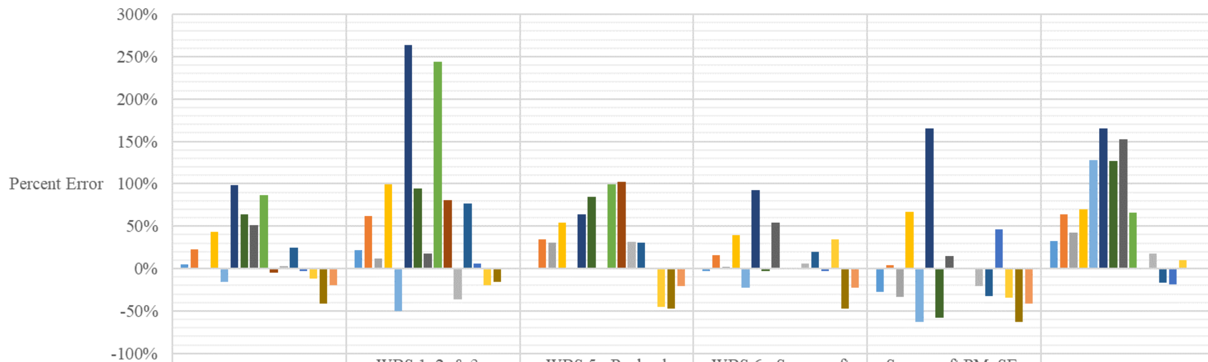
Where ε is the percent error, C_E is the estimated cost, and C_A is the actual cost from the CADRe data. For each mission the percent error for the mission total, as well as each of the spacecraft bus subsystems is presented. The mission total was defined as all of the costs that were estimated in the present study for a particular mission, namely WBS 1, 2, & 3, payload total, spacecraft bus total, and WBS 10 cost. The mission total cost does not include Science/Technology, Mission Operations, Launch Vehicle, Ground Systems, Education and Public Outreach, software development, or any other aspects of the mission that were not included in the estimate. There were several instruments without enough supporting technical documentation to estimate their cost. Such instruments were omitted from the estimates presented in this study. Similarly, there were certain spacecraft components such as the RTG on New Horizons that PRICE and SEER could not estimate and were not included in the cost estimates. WBS 1, 2, & 3 is the combined cost of NASAs project management, systems engineering, and safety and mission assurance. Payload Total is the combined costs of all the instruments which were included in the estimate, including any PM, SE, or IAT costs associated with the instruments. The Spacecraft Bus Total is the total cost paid to the contractor for developing and producing all the spacecraft bus's subsystems as well as the associated PM, SE, and IAT costs. Spacecraft PM, SE, and IAT are the costs associated with the project management, systems engineering and integration assembly and test performed by the prime contractor building the spacecraft bus. WBS 10 IAT is the final integration assembly and test of the spacecraft bus to the instrument and the launch vehicle. PRICE and SEER's errors in estimating each of these systems and subsystems are presented except in cases where the CADRe data did not provide a detailed enough breakdown.

In addition to the errors of all the systems and subsystems costs for each mission, the average, weighted average, and median errors are presented for each system and subsystem, respectively. The average and median errors are simply the mean and median value of the errors for a given system or subsystem. The weighted average error is the mean value of the errors weighted by the actual costs, which were converted from real-year dollars to a common base year for comparison.

A. SEER

The results for SEER in estimating mission total costs and systems costs are given in Fig. 1. The missions are ordered by the magnitude of the actual cost for each mission included in this study; where IBEX was the least expensive mission in the study and Juno being the most expensive. Note that this ordering only takes into account the costs estimated by this study. In reality, New Horizons cost more than MESSENGER, but it is ordered between WISE and MESSENGER because New Horizons payload, RTG, and operations costs were not included.

The first column of Fig 1 shows SEER's error in estimating total mission cost for each of the twelve missions. Of the twelve missions in the present study, SEER over estimated the cost of six and under estimated the cost of six, resulting in a median error of -0.3%. Therefore, SEER was shown to be very accurate in that it was just as likely to over predict the cost as under predict it. The average error was 23%, but the weighted average was only 5%. Thus, SEER was more likely to over estimate the costs of low cost missions, and under estimate the cost of high cost missions. Additionally, when SEER under estimated the cost of a mission it was typically a small error, whereas over estimates tended to be larger errors. For example, the second smallest mission, CONTOUR, was over estimated by 99%, whereas the second largest mission, SMAP, was underestimated by -42%.



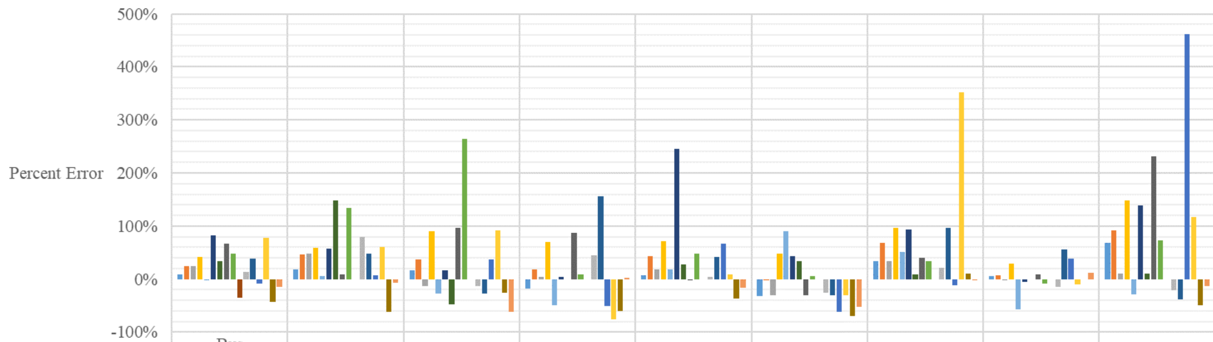
	Mission Total	WBS 1, 2, & 3 - PM, SE, SMA	WBS 5 - Payload Total	WBS 6 - Spacecraft Bus Total	Spacecraft PM, SE, IAT	WBS 10 IAT
Weighted Average	5%	22%	-1%	-3%	-28%	33%
Average	23%	62%	34%	16%	4%	64%
Median	-0.3%	12%	31%	2%	-33%	42%
Standard Dev	43%	99%	54%	39%	67%	70%
IBEX	-16%	-50%	1%	-22%	-62%	128%
CONTOUR	99%	264%	64%	92%	166%	165%
WISE	64%	95%	84%	-3%	-58%	127%
New Horizons	52%	18%		54%	15%	153%
MESSENGER	87%	244%	99%			66%
GRAIL	-5%	81%	103%			
Deep Impact	3%	-37%	32%	6%	-21%	18%
MAVEN	25%	77%	30%	20%	-32%	-16%
DAWN	-3%	6%		-3%	46%	-19%
Kepler	-11%	-19%	-45%	35%	-34%	10%
SMAP	-42%	-16%	-47%	-47%	-63%	-1%
Juno	-19%	1%	-21%	-23%	-41%	

Fig. 1 SEER Systems Comparison

Galorath claims that SEER estimates have an average error of -1%, with a standard deviation of 19%. It is not known whether the advertised -1% average error is a weighted average or not. Regardless, when compared to the 5% and 23% weighted average seen in the present study, SEER was not as accurate as advertised. Additionally, the present study found that the standard deviation of the error in SEER's estimates was 43%, over twice as much as the advertised 19%, meaning SEER's estimates were not as precise as advertised either. However, these large errors may not necessarily be the fault of SEER as described in Sec. IV.

The second column of Fig. 1 shows SEER's error in estimating WBS 1, 2, & 3 costs which represent NASAs Project Management (PM), Systems Engineering (SE), and Safety and Mission Assurance (S&MA). The last column shows the error in estimating WBS 10, the final Integration Assembly and Test (IAT) of the spacecraft with its payload and launch vehicle for each mission. All system level costs were estimated using the specific settings recommended by Galorath in their SEER Space Guidance document[3]. However, the weighted average error of WBS 1, 2, & 3 was 22%, and the weighted average error of WBS 10 was 33%. In contrast, the weighted averages for the Payload and Spacecraft Bus were -1% and -3%, respectively. The standard deviation of the WBS 1, 2, & 3 and WBS 10 estimates were 99% and 70%, respectively. In contrast, the standard deviation of the Payload and Spacecraft Bus estimates were 54% and 39%, respectively. Thus, the systems cost estimates at the mission level contributed more error and uncertainty than the payload and spacecraft cost estimates. The fifth column shows the systems level cost for the spacecraft bus. In general, the spacecraft bus's systems costs were underestimated.

Figure 2 shows SEER's errors in estimating each subsystem of the spacecraft bus for each mission. The weighted average error of all the subsystems was 8%. This is in contrast to the spacecraft bus total in Fig. 1 of -3%. Thus, SEER underestimated PM, SE, and IAT costs associated with a contractor building a spacecraft bus, but overestimated the costs of designing and producing the subsystem components. SEER overestimated the cost of the structures and mechanisms for all missions except SMAP and Juno, the two largest missions. SEER underestimated the cost of half of the thermal subsystems, but when it overestimated the costs, the magnitude of the error was larger, giving the thermal subsystems an overall weighted average error of 17%. The propulsion subsystems were typically overestimated, but when they were underestimated the error was less than -50%. This brought the weighted average of the propulsion



	Bus Subsystems Total	Structures	Thermal	Propulsion	GN&C	Comm	Electrical Power	Harness	C&DH
Weighted Average	8%	18%	17%	-19%	8%	-31%	35%	6%	68%
Average	24%	46%	38%	18%	43%	-2%	69%	7%	91%
Median	24%	48%	-14%	4%	19%	-30%	34%	-2%	10%
Standard Dev	41%	59%	90%	69%	72%	47%	97%	29%	148%
IBEX	-1%	6%	-27%	-50%	19%	90%	52%	-57%	-29%
CONTOUR	82%	57%	17%	4%	246%	44%	93%	-5%	139%
WISE	34%	148%	-48%		28%	34%	9%		10%
New Horizons	66%	8%	97%	88%	-1%	-31%	40%	9%	232%
MESSENGER	48%	134%	264%	8%	48%	5%	34%	-8%	73%
GRAIL	-36%								
Deep Impact	14%	80%	-14%	45%	4%	-26%	21%	-14%	-22%
MAVEN	39%	48%	-27%	157%	41%	-30%	97%	55%	-38%
DAWN	-9%	7%	37%	-51%	66%	-62%	-12%	38%	462%
Kepler	77%	60%	93%	-75%	9%	-30%	352%	-10%	118%
SMAP	-42%	-62%	-26%	-60%	-37%	-70%	10%	0%	-48%
Juno	-15%	-6%	-61%	3%	-16%	-52%	0%	12%	-13%

Fig. 2 SEER Subsystems Comparison

subsystem error down to -19%. GN&C errors were relatively small with the exception of CONTOUR. However, the GN&C subsystem of CONTOUR was only 4% of the subsystems' cost, so while the percent error was very large the absolute error was only moderate. SEER underestimated the majority of the communication subsystems costs. The communication subsystem is the subsystem most underestimated by SEER with a weighted average error of -31%. Electrical power was consistently over estimated except in the case of DAWN, the only mission in this study which utilized solar electric propulsion. The electrical power subsystem was the second most overestimated subsystem by SEER, with a weighted average error of 35%. The harness subsystem cost was most accurately estimated by SEER. The harness subsystem's weighted average error, median error and standard deviation are all smaller in magnitude than all other subsystems. The C&DH subsystem, in contrast, had the largest error and standard deviation of all the subsystems. The large error of the C&DH subsystem is likely due to the large uncertainties in estimating custom integrated circuits, such as Field Programmable Gate Arrays (FPGAs).

Next, SEER's uncertainty analysis capabilities were investigated. For every input in SEER the user inputs a "least," "likely," and "most" value corresponding to an optimistic, most likely, and pessimistic assumption about the input. SEER models uncertainty by assigning each "work element" a distribution of possible costs in addition to a mean cost. The least/likely/most inputs for each work element correspond to the lower bound, mode, and upper bound of the Probability Density Function (PDF) of a beta distribution given by

$$f(x) = \begin{cases} \frac{x^{(\alpha-1)}(1-x)^{(\beta-1)}}{B(\alpha,\beta)} & i_{least} \leq x \leq i_{most} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $B(\alpha, \beta)$ is a function of gamma functions given by

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{3}$$

α and β are shape parameters given by

$$\alpha = \frac{(\mu - i_{least})(2i_{likley} - i_{least} - i_{\text{most}})}{(i_{likely} - \mu)(i_{most} - i_{least})} \tag{4}$$

and

$$\beta = \alpha \frac{(i_{most} - \mu)}{(\mu - i_{least})}. \tag{5}$$

Where i_{least}, i_{likely}, and i_{most} are the least/likely/most inputs into SEER, and μ is the mean of the distribution given by

$$\mu = \frac{i_{least} + 4i_{likely} + i_{most}}{6} \tag{6}$$

When SEER runs a Monte Carlo simulation it selects the value of the inputs to its CERs in a random way determined by Eq. 2. Over all the trials of the Monte Carlo simulation SEER was able to build up a distribution of possible costs for the spacecraft. This distribution was then used to create a cumulative distribution function (CDF).

The frequency that the actual cost of a mission was within SEER's 80% confidence interval was used to evaluate the accuracy of SEER's uncertainty quantification capabilities. A confidence interval is the interval between two points, or confidence levels, on a CDF. When a model claims an estimate has an 80% confidence interval, that means 80% of the time the actual value should be within that interval. Thus, if SEER's uncertainty quantification capabilities are accurate it is expected that nine or ten of the twelve missions' (approximately 80%) actual costs will be within SEER's 80% confidence interval. 90% and 95% confidence intervals are much more common in literature, however, SEER only outputs the values of the CDF, known as confidence levels, at intervals of 10%. As a result, the user must be content with only knowing the 80% confidence interval instead of the more common 90% or 95% confidence intervals. The 80% confidence intervals generated by SEER for the twelve missions in this study can be seen in Fig. 3. Figure 3 shows that the only missions with actual costs outside SEER's 80% confidence interval were CONTOUR, WISE, and MESSENGER. Since twelve missions were included it would be expected that two or three would fall outside of the confidence interval which is exactly what was observed. It can be concluded that while the point estimates given by SEER had a wide variance, the uncertainty quantification function usually captured the actual cost.

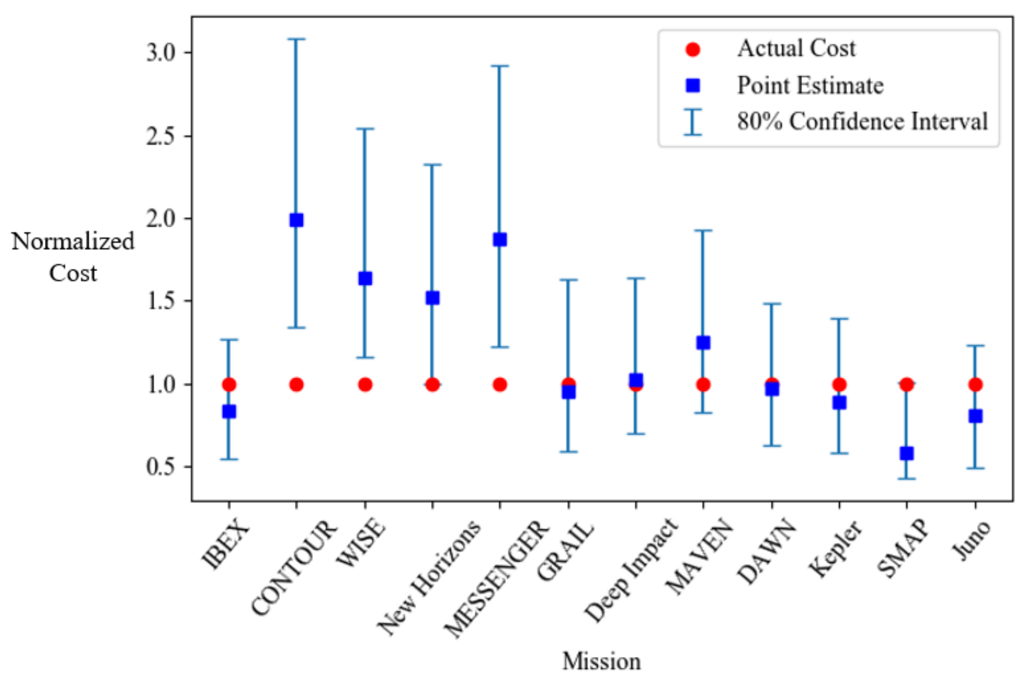


Fig. 3 SEER Uncertainty Analysis

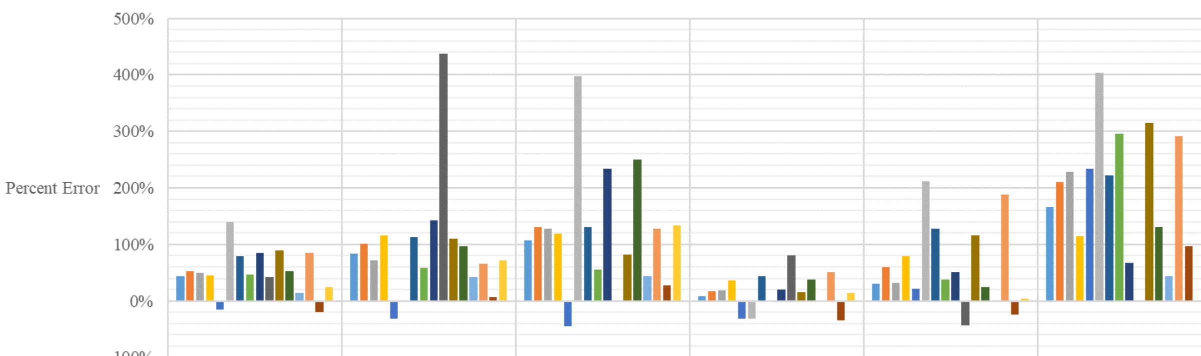
It should be noted that the confidence intervals seen in Fig. 3 are extremely large. This was primarily due to the uncertainty in modeling the spacecraft from the data available. In particular, it was extremely difficult to judge the level of heritage of a majority of the spacecraft components. In some cases, there was no information on whether a component was a brand new design, or a copy of a previously flown component. In these cases, the estimators would adjust the new design least, likely, and most inputs to reflect the uncertainty.

B. PRICE TruePlanning

The results for PRICE's performance in estimating mission total costs and systems costs can be seen in Fig. 4. As with the SEER results, the missions are ordered from least to most expensive. Of the twelve missions in the study, PRICE overestimated the cost of ten and underestimated the cost of two, and the median error was 50.0%. Additionally, the average error was 52%, and the weighted average error was only slightly better at 43%. Therefore, PRICE was much more likely to overestimate mission cost. Interestingly, the two missions that PRICE underestimated were the two missions that SEER underestimated by the largest margin. In fact, if you order the results by percent error the resulting rank order is very similar between the two tools, with the one significant exception being Kepler.

PRICE Systems claims that PRICE estimates have an average error of +1% with a standard deviation of 13%. With the major caveat that this study employed technical data from CDR, while PRICE used as built data, both average and standard deviation were found to be significantly higher than these values. This study found the average and weighted average error to be 52% and 43% respectively, and the standard deviation to be 47%. PRICE also over estimated IAT for every mission. The smallest error was an overestimation of 67% and the highest was overestimated by 403%. PRICE overestimated the WBS 1, 2, & 3 costs of all missions except for IBEX, with a weighted average error of 106%.

The most accurate and precise estimates from PRICE came from estimating the payload subsystems cost. PRICE's weighted average error was only 9% and the standard deviation of the estimates was only 37%. SEER's estimate of payload total cost was more accurate with a weighted average error of -1% but was less precise with a standard deviation of 54%.



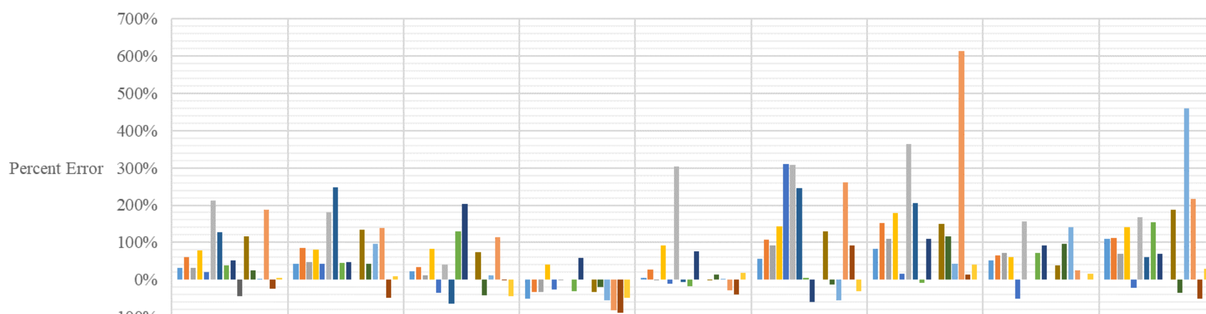
	Mission Total	All Systems Costs (WBS 1, 2, 3, & 10)	WBS 1 ,2, & 3 - PM, SE, SMA	Payload Subsystems Total	Spacecraft Bus Subsystem Total	WBS 10 IAT
Weighted Average	43%	83%	106%	9%	31%	166%
Average	52%	101%	131%	17%	60%	210%
Median	50%	72%	128%	18%	31%	228%
Standard Dev	45%	116%	118%	37%	79%	114%
IBEX	-15%	-31%	-45%	-31%	21%	234%
CONTOUR	140%		398%	-31%	213%	403%
WISE	80%	113%	130%	44%	127%	223%
New Horizons	47%	59%	56%		38%	296%
MESSENGER	85%	143%	235%	20%	51%	67%
GRAIL	42%	437%		81%	-43%	
Deep Impact	89%	110%	83%	16%	116%	315%
MAVEN	53%	97%	250%	38%	24%	130%
DAWN	14%	43%	43%		1%	44%
Kepler	86%	66%	128%	51%	188%	292%
SMAP	-20%	7%	27%	-35%	-24%	96%
Juno	24%	72%	134%	14%	4%	

Fig. 4 PRICE Systems Comparison

Figure 5 shows PRICE's errors for the estimation of each subsystem of the spacecraft bus for each mission in this study. The average and weighted average errors for all the spacecraft subsystems are 60% and 31% respectively. These errors were driven by the fact that PRICE overestimated the cost of every subsystem except propulsion, where the weighted average error was -51%. The only two missions PRICE underestimated the total spacecraft bus were SMAP and GRAIL. Interestingly SMAP was the only mission PRICE underestimated the structures cost, and GRAIL did not have subsystem costs broken down. PRICE was most accurate in estimating the GN&C subsystem where the weighted average error was only 5%, however, the average error and standard deviation were driven higher by the significant error for CONTOUR. On the other end of the spectrum, the three subsystems with the largest error were C&DH, Electrical Power, and Communications, with weighted average errors of 111%, 82%, and 57% respectively. The potential explanation for this degree of error is further examined in the discussion section.

Similar to SEER, PRICE too has the ability to perform uncertainty analysis. For every input in PRICE the user has the option to input an optimistic, most likely, and pessimistic value. Unlike SEER, in PRICE the user selects what inputs to apply uncertainty to, and what the inputs should be. Whereas SEER has default least/likely/most assumptions for nearly every input which can be adjusted by the user as needed. Within PRICE, the user must select what inputs to apply uncertainty to. In this study, the cost estimators applied uncertainty to weight and new design. PRICE then quantifies uncertainty in a similar way to SEER, except it uses a triangular distribution instead of a beta distribution. Details of their methods can be found in "FRISK-Formal Risk Assessment of System Cost Estimates"[5]. PRICE outputs the values of the CDF from the 5% to 95% confidence levels. It is therefore possible to generate a 90% confidence interval, however, for consistency with SEER, the 80% confidence interval is presented instead. The 80% confidence intervals generated from PRICE for the twelve missions in this study can be seen in Fig. 6. It is immediately clear that none of the PRICE estimates fall withing their confidence intervals. With an 80% confidence interval it would be expected that nine or ten of the twelve missions would fall within the bounds of the estimate.

Another significant difference between PRICE and SEER is the confidence level at which the point estimate falls. In SEER, the point estimate is the median value of the distribution, meaning the uncertainty inputs drive the point estimate. However, in PRICE the point estimate is generated using only the most likely input. Since the input distributions are



	Spacecraft Bus Subsystem Total	Structures	Thermal	Propulsion	GN&C	Comm	Electrical Power	Harness	C&DH
Weighted Average	31%	42%	23%	-51%	5%	57%	82%	52%	111%
Average	60%	85%	35%	-33%	28%	109%	151%	65%	113%
Median	31%	46%	12%	-32%	-2%	92%	109%	73%	71%
Standard Dev	79%	80%	82%	40%	92%	143%	178%	62%	142%
IBEX	21%	42%	-36%	-26%	-10%	310%	17%	-52%	-22%
CONTOUR	213%	180%	40%	-0.2%	305%	309%	363%	156%	168%
WISE	127%	247%	-65%		-7%	245%	205%		62%
New Horizons	38%	46%	129%	-31%	-16%	4%	-8%	73%	154%
MESSENGER	51%	46%	203%	58%	77%	-61%	109%	91%	71%
GRAIL	-43%								
Deep Impact	116%	135%	74%	-34%	-2%	130%	149%	37%	188%
MAVEN	24%	43%	-43%	-20%	14%	-13%	115%	96%	-34%
DAWN	1%	97%	12%	-56%	0.1%	-55%	43%	142%	461%
Kepler	188%	138%	113%	-82%	-28%	262%	613%	26%	217%
SMAP	-24%	-50%	-0.1%	-90%	-40%	92%	15%		-51%
Juno	4%	10%	-44%	-49%	19%	-31%	41%	16%	29%

Fig. 5 PRICE Subsystems Comparison

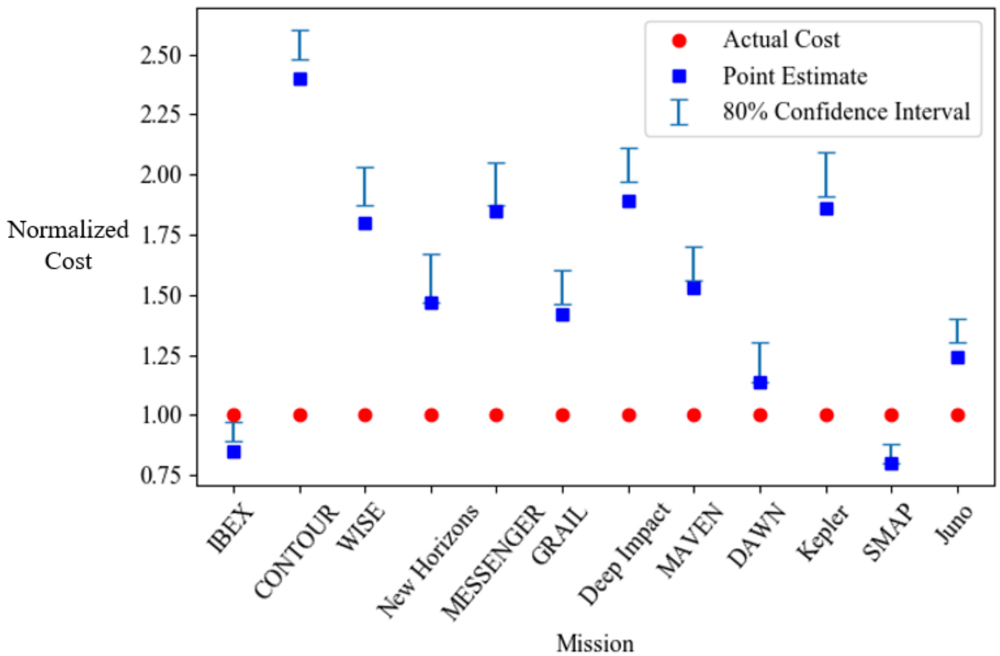


Fig. 6 PRICE Uncertainty Analysis

typically right skewed the end result is that the confidence level of the point estimate is substantially lower in PRICE compared to SEER. The distributions produced by PRICE are so right skewed that the majority of the point estimates fall below the 10% confidence level and outside of the 80% confidence interval.

IV. Discussion

In this section, a number of factors which may have adversely affected the outputs of PRICE and SEER in this blind study are discussed. Much of what is discussed in this section is speculation by the authors based off their experience as cost estimators and has not been rigorously tested.

A. Mass Estimation and Uncertainty

In PRICE, mass is the primary scaling mechanism for all components in estimating the cost. In SEER this is true for the majority of components in the structures, thermal, and propulsion subsystems, as well as a few components in the other subsystems. For the present study, it was assumed that the low mass input was the current best estimate, the likely estimate was the current best estimate plus the mass contingency, and the high estimate was 30% above the likely estimate. These assumed mass inputs are common in the cost estimating community and are recommended by the SEER Space Guidance document[3]. Despite being a common assumption in the cost estimating community, it is not clear where the assumption that the high mass estimate should be equal to the likely assumption plus 30% originates. It is possible that it comes from the 2006 paper "Evaluating the Impacts of Mass Uncertainty on Future Exploration Architectures," which showed that the average spacecraft increases in mass by 28.5% from its preliminary design to launch[6]. The 30% assumption does not take into account the design maturity of the components, or how close the mission is to completion. If the costing community switched to using the Jet Propulsion Laboratory mass margins approach defined in their handbook of design principles for flight systems[7], cost estimate accuracy may improve. The handbook recommends mass margins of 30% at PMSR, 20% at PDR, and 10% at CDR. The high 30% mass margin used for the present study is likely a contributing factor, for the high average error seen in the PRICE and SEER estimates.

B. Prototyping Assumptions

The number of prototypes affects the cost output from PRICE and SEER in different ways. In SEER, each additional prototype adds to the development cost. In PRICE, each component has an optimal number of prototypes where the cost is minimized. A standard assumption in the cost estimating community is a prototype input value of 1.3. A fractional prototype represents a prototype which does not have the full functionality of the working component, such as a breadboard, or boiler plate unit. For many spacecraft subsystems, there were details on how many prototypes were made for each component. However, when there was no data on the number of prototypes, the assumed value was 1.3. It is unclear where the assumption that the average spacecraft component for a NASA science mission will have 1.3 prototypes originated. The authors plan to investigate the effect of the number of prototypes on the accuracy of cost estimates in the future. The standard assumption of 1.3 prototypes most likely had a minimal effect on the PRICE estimates, but it may have inflated the SEER estimates as it is possible that in some cases the reason there was no prototype information was because there were no prototypes.

C. Subject Matter Experts

The role of subject matter experts is critical in any estimation. Unfortunately, since the missions in the present study have taken place over the past 20 years it was not possible to contact any of the subject matter experts working on these missions to clarify details missing in the CADRes or supporting documents. As a result, there was a wide range of uncertainty in the inputs and results. For example, CONTOUR, New Horizons, and MESSENGER were each significantly overestimated by both tools. One of the subsystems driving this overestimation was the C&DH subsystem. Each of these spacecraft shared a common C&DH electronics module with significant heritage. The documents from CDR described the electronics thoroughly, including circuit diagrams, and technical specifications. However, they did not mention that the majority of the circuit boards were copied from previous missions. It was assumed during the blind estimation that instead of being copied that these electronics would be modified to fit the requirements of each mission. A five minute conversation with an electrical engineer from any of these missions could have cleared up the confusion and significantly improved the accuracy of the estimate.

D. Experience of Estimators

The experience and skill of the estimator also plays a significant role in the accuracy of a cost estimate. Estimators B and C received formal training in using PRICE TruePlanning for space missions from PRICE Systems LLC. They did not receive any formal training using SEER but completed a number of training exercises over a few weeks before beginning the study. In an attempt to minimize this effect the estimators were advised by more experienced estimators who were on hand to answer any questions. However, we would be remiss not to mention that a cost estimator with several years of experience would likely have produced more accurate and precise results with both tools.

E. PRICE Error Analysis

The estimates developed using the Space Missions Catalog within PRICE TruePlanning in this study were relatively inaccurate. The most likely explanation for this is related to the dependence of the tool's CERs on mass, particularly for electronic components. The three subsystems where PRICE most severely overestimated costs were those with the highest number of electronic components: C&DH, Power, and Communications. The overestimation of these subsystems then effects the estimates of the systems level cost. PRICE's estimate of PM, SE, and S&MA (project and payload/spacecraft level) and IAT costs are strongly correlated with the subsystem level hardware costs. Further, many of the missions had large electronic boxes without sufficient information to break down the mass into individual boards. In SEER, where electronics cost depend on other factors such as clock speed, number of integrated circuits per circuit board, and number of pins per circuit board, the information was available on specification sheets. However, this information did not include mass and the estimators had to make educated guesses on not only the breakdown of mass between the individual circuit boards, but also the breakdown between structures and electronics mass (a significant cost driver). The reliance on mass is further exacerbated in the uncertainty analysis, where, as discussed above, the assumed input distribution is heavily right-skewed, potentially more than necessary for this point in the project's lifecycle. Another, and likely related, potential source of error was leveraging the SEER Space Guidance document for input assumptions when data in CADRe was lacking. This may not be appropriate in certain areas, particularly in the heritage assumptions of the spacecraft bus. A similar guidance document for PRICE would be a useful addition to the PRICE documentation. Lastly, there is simply more work required to determine, with certainty, the sources of error in the

PRICE estimates, particularly considering the Space Missions Catalog's leveraging of CADRe data (or related data) to develop the CERs.

V. Conclusions

The present work evaluated the parametric costing tools PRICE TruePlanning-Space Missions Catalog and SEER-H through a blind study. The study was blind in that the estimators had no knowledge of the actual costs of the missions being estimated. This was done to prevent cognitive biases, such as anchoring, from influencing the way the tools were used to estimate the costs.

The present study found that SEER had an average error of 23%, median error of -0.3%, and a standard deviation of 43%. Weighing the errors by the actual cost of the mission the average error was only 5%. The present study also determined that SEER tended to overestimate smaller missions and underestimate larger missions. SEER's uncertainty quantification capabilities worked well; of the twelve missions in the present study, nine of the missions' actual costs were within the 80% confidence interval given by SEER.

The present study determined that Price had an average error of 52%, median error of 50%, and standard deviation of 45%. Weighing the errors by the actual cost for each the mission, the weighted average error is 43%. PRICE's uncertainty quantification tool performed poorly, which may be due to the fact that uncertainty could be applied to far fewer input parameters than in SEER. None of the twelve missions fell within PRICE's 80% confidence interval.

There are several factors which may have affected the results of this study. These factors include the assumptions about mass margins and numbers of prototypes, as well as the experience of the estimators. In addition, the estimators were not able to ask clarifying questions from subject matter experts involved in the missions. The estimators had to rely on documents presented at major reviews only, which often omitted important information; such as, the heritage of a particular component. Lastly, and perhaps most significantly, the present study leveraged technical data from each mission's CDR while the tool providers modeled the missions using as built data. Future work will focus on the detailed investigation and identification of factors that may have contributed to the accuracy of the cost estimations reported in this study.

Acknowledgments

The authors would like to thank Robert Fairbairn and Hernani Tosoc for their advice and support while formulating this study.

References

[1] Trivailo, O., Sippel, M., and Şekercioğlu, Y., "Review of hardware cost estimation methods, models and tools applied to early phases of space mission planning," *Progress in Aerospace Sciences*, Vol. 53, 2012, pp. 1–17.

[2] Sanchez, S., and Kha, K., "SEER Validation Study Results for NASA Space Science Missions," , 2015.

[3] Galorath, "SEER-H Space Guidance: Revision 2.2," , 2016.

[4] Swaren, J., "TruePlanning®Validation Studies for Estimating Unmanned Space Projects," , 2017.

[5] Young, P., "FRISK-Formal risk assessment of system cost estimates," *Aerospace Design Conference*, 1992, p. 1054.

[6] Wilhite, A., Reeves, D., Stanley, D., and Wagner, J., "Evaluating the impacts of mass uncertainty on future exploration architectures," *Space 2006*, 2006, p. 7250.

[7] Yarnell, N., "Design, verification/validation and operations principles for flight systems," *Jet Propulsion Laboratory, D-17868, Rev*, Vol. 2, 2003.