Exploring the Limits of "Faster, Better, Cheaper" With Mission Cost Risk Assessment

Christian Smart, Ph.D.

Northrop Grumman IT

Cindy Fryer, Chief

Harry Born

Resource Analysis Office

Goddard Space Flight Center

Introduction

In May 2000, David Bearden of the Aerospace Corporation Presented "A Complexity-Based Risk Assessment of Low-Cost Planetary Missions: When is a Mission Too Fast and Too Cheap?" At the fourth IAA International Conference on Low-cost Planetary Missions, JHU/APL.

- Pioneering effort to establish risk, cost, and schedule tradeoffs
- Highly critical of "faster, better, cheaper" policy
- Featured in Aviation Week and Space Technology article

The May 2000 Aerospace Study and GSFC's Experience

- The Aerospace study's conclusions were different from Goddard space flight center's experience with "faster, better, cheaper"
  - e.g., the small explorer (SMEX) program

- In July 2000, we examined the aerospace study and tried to reproduce the results using Goddard data

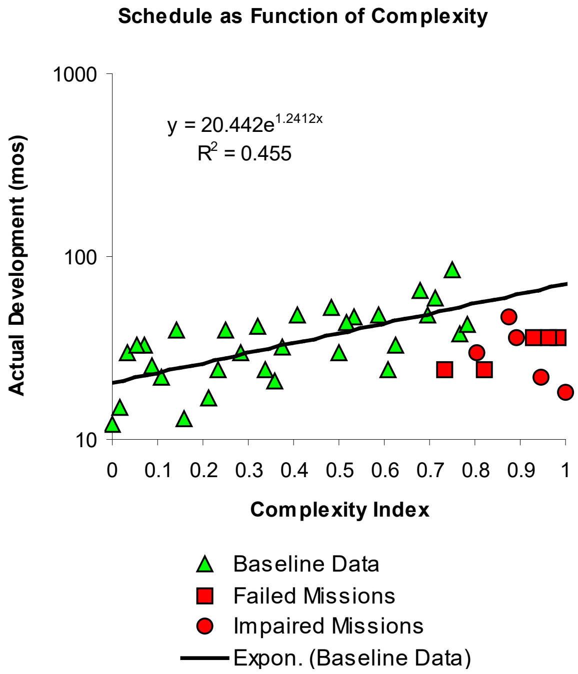
May 2000 Aerospace Corp. Study

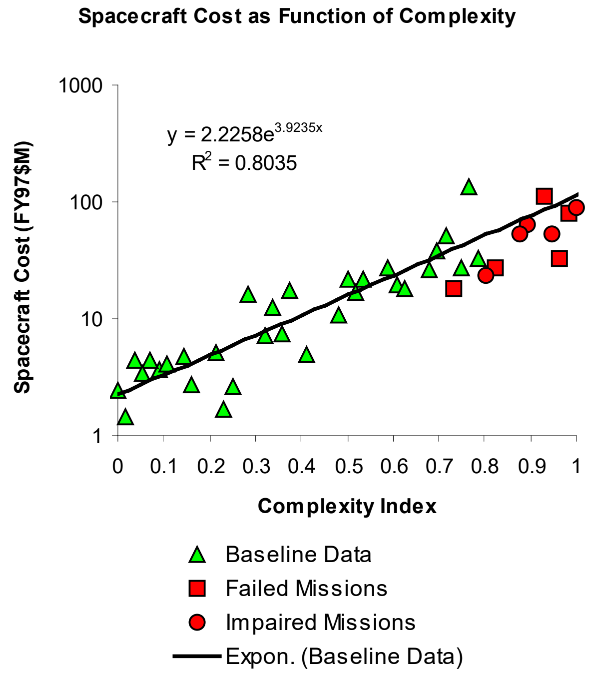
- Over 40 missions launched between 1990-1999 were studied
  - Includes earth-orbiting and planetary
  - Does not include non-U.S. missions
  - Includes both NASA and non-NASA missions
- Introduces "spacecraft complexity" metric
  - Arithmetic mean of the "complexity" levels of 21 technical and programmatic parameters for the spacecraft (e.g., mass, power, data rate)
  - Complexity levels for each parameter computed as a percentile rank
  - Spacecraft complexity ranges from 0 (least complex) to 1 (most complex)

May 2000 Aerospace Corp. Study (Cont'd.)

- Mission classification
  - Successful
  - Failed
  - Impaired
- Relates Spacecraft Complexity to Cost and Schedule
  - Cost strongly correlated
  - Schedule correlation is weaker but significant

May 2000 Aerospace Corp. Study (Cont'd.)





Aerospace Study Conclusions:

- More complex missions more likely to fail
- Regression curves can be used as parametric indicators

Task: Determine whether or not Goddard experience is consistent with the May 2000 Aerospace study

- In July 2000, we attempted to reproduce Aerospace Study using Goddard data

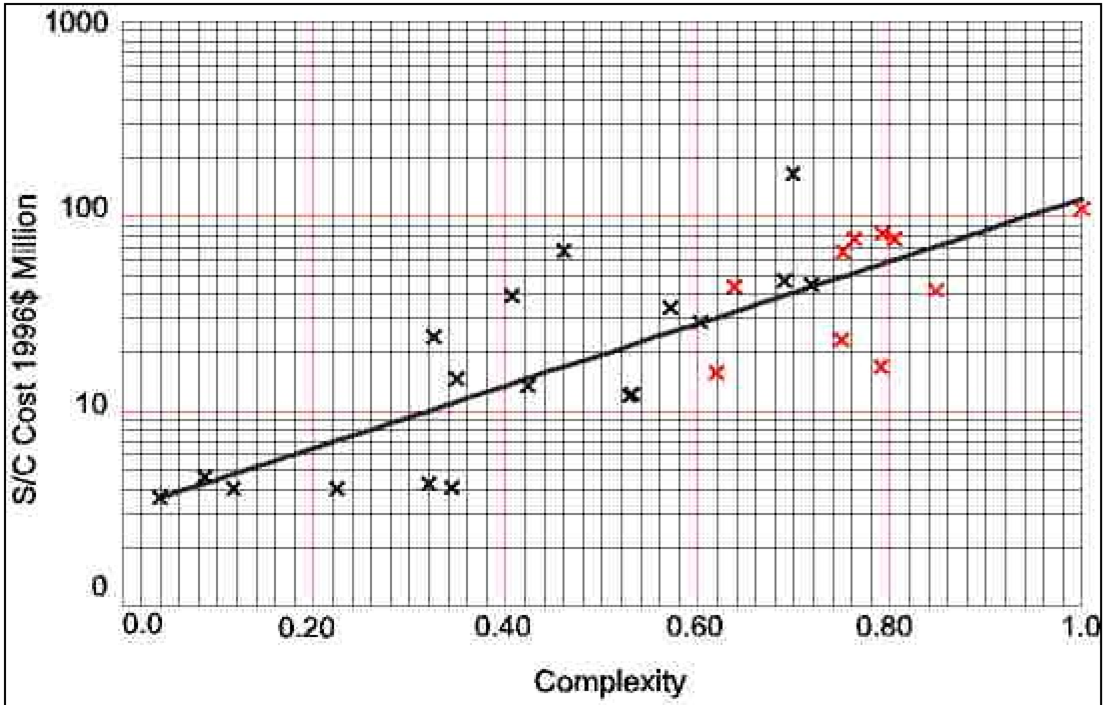
Task: Determine Whether or Not Goddard Experience Is Consistent With the May 2000 Aerospace Study

- Different results
  - Weaker, but still strong relationship between cost and complexity ($R^2 = 0.658$)
  - No relationship between schedule and complexity ($R^2 = 0.070$)
- Reasons for differences
  - Slightly different database (we did not have data for two non-NASA missions and thus could not include them in our study)
  - Used 14 parameters to calculate spacecraft complexity (several parameters used in Aerospace study are highly correlated – e.g., BOL Max power and EOL max power, which can give too much weight to certain subsystems, and skew the results)

Task: Determine Whether or Not Goddard Experience Is Consistent With the May 2000 Aerospace Study (Cont'd.)

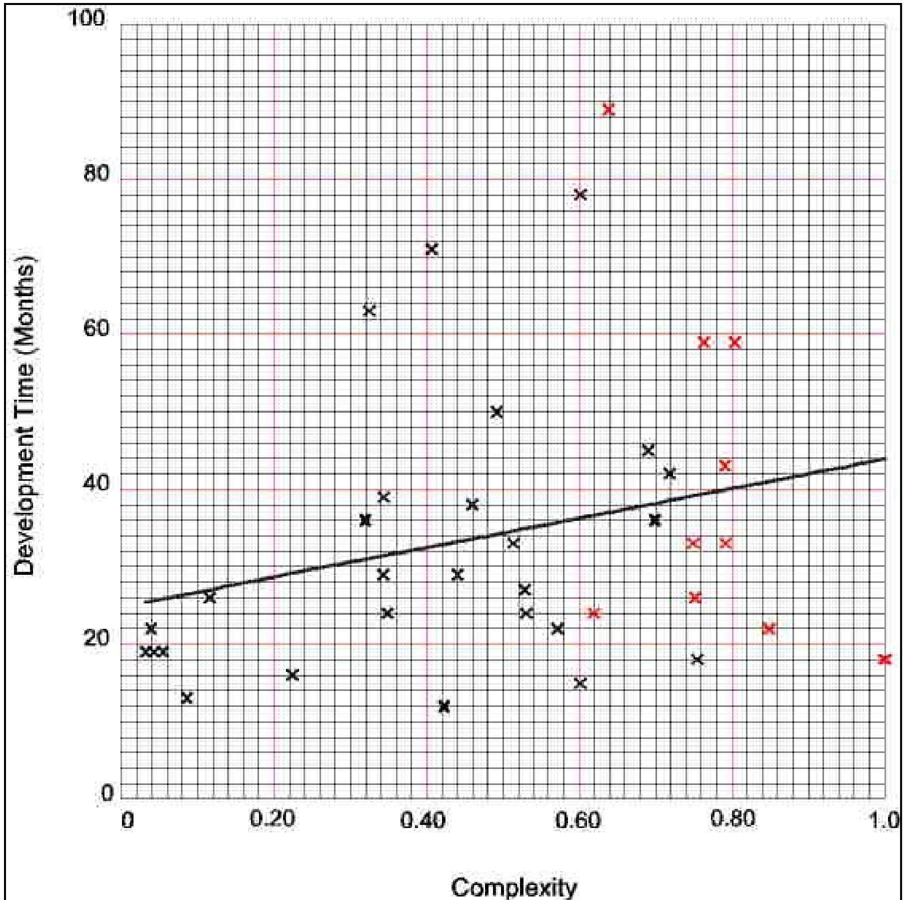
- Small changes in assumptions and data result in large changes in results (see charts that follow)
  - Only half the failed missions have cost or schedule below average for a given level of complexity
  - Model results seem sensitive to small changes in data

RAO Analysis Results
Cost Vs. Spacecraft Complexity



N1: S/C Cost = 3.094 * 39.841^ Complexity
N = 28
CC = 0.811
SEA = 0.677
RSQ = 0.658
CV = 0.216

RAO Analysis Results
Schedule Vs. Spacecraft Complexity



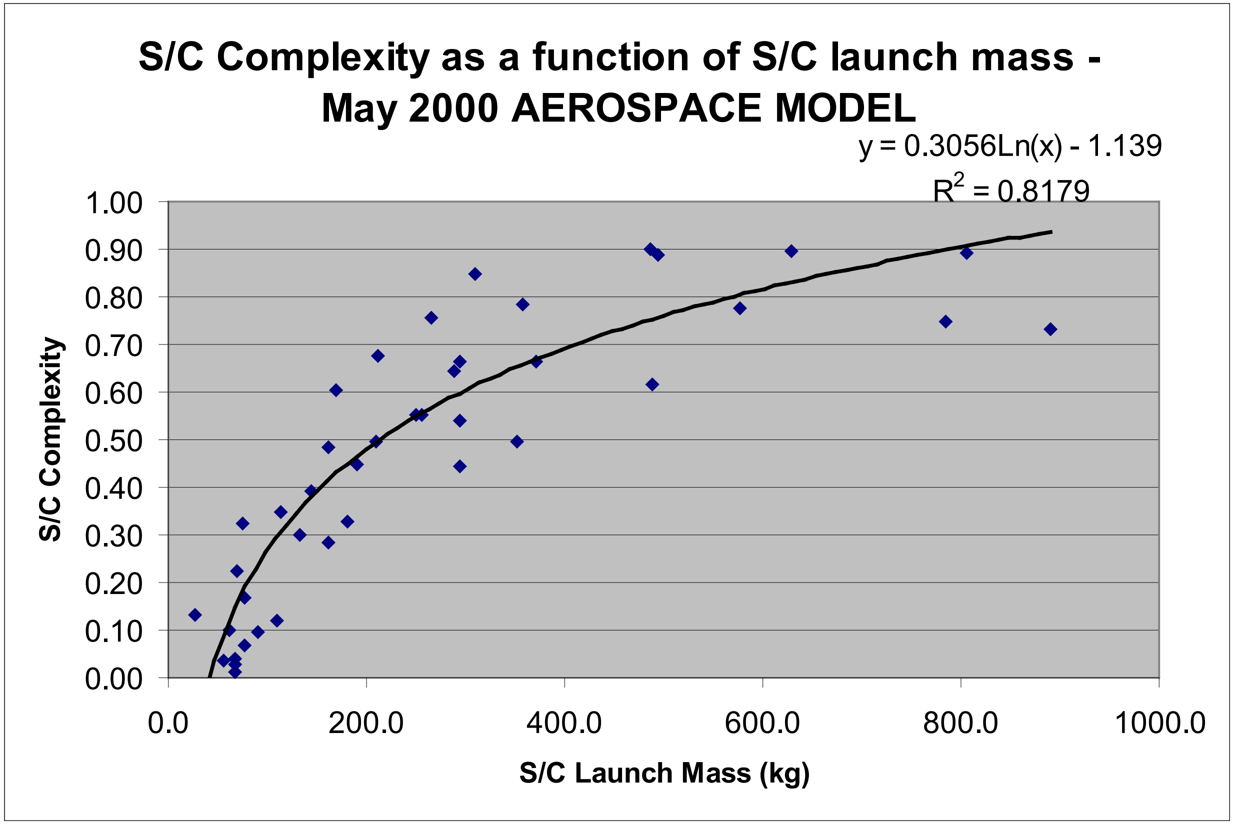
N1: Development Time = 24.854 +19.031 * Complexity
N = 37
CC = 0.265
SEA = 18.403
RSQ = 0.070
CV = 0.536

Preliminary Observations on the May 2000 Aerospace Model

- Data discrepancies, as discussed above
  - Half of missions in dataset are non-NASA
  - Database includes several small satellites with design lives of only a few hours or days
- Risk NOT measured directly
- Treats cost and schedule as independent from each other
- Complexity seems to be driven by spacecraft mass

Complexity and Mass

There is extremely high correlation between launch mass and Aerospace spacecraft complexity parameter.



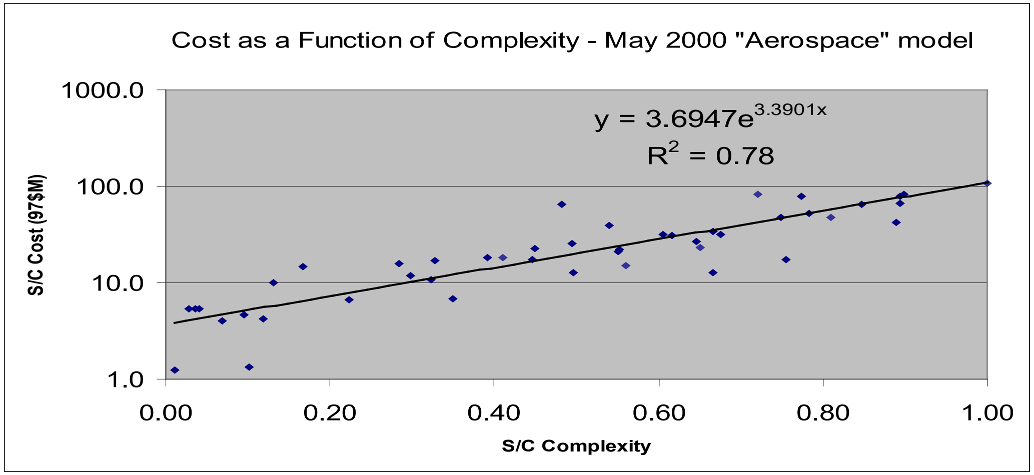
Complexity and Mass (Cont'd.)

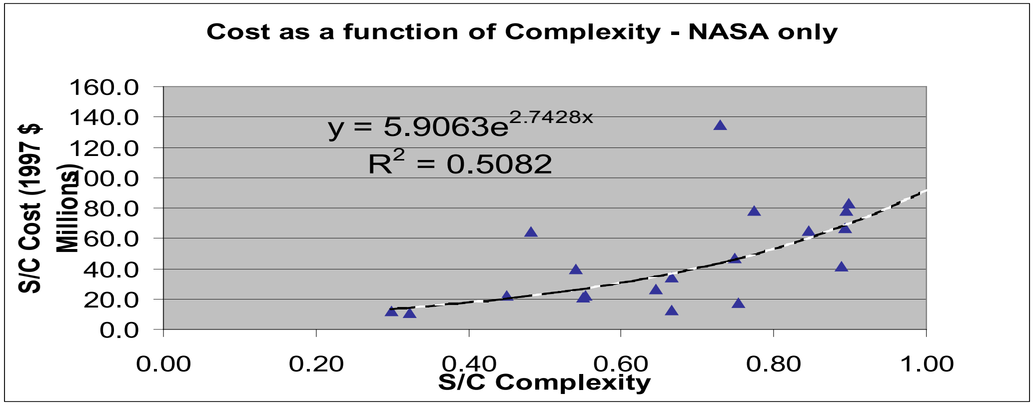
- High correlation between weight and complexity caused in part by considering non-NASA microsatellites
  - Weight seems to drive the complexity

Inclusion of Non-NASA Missions

- Including non-NASA missions seems to affect the results
  - Possibly overstates the goodness-of-fit of the model
- The impact of removing these missions
  - Approximated the May 2000 Aerospace model by using the graphs from the Aviation Week article ($R^2 = 0.78$ vs. $R^2 = 0.80$ for weight/complexity equation), then removed the non-NASA missions
  - R^2 dropped to 0.51

Inclusion of Non-NASA Missions



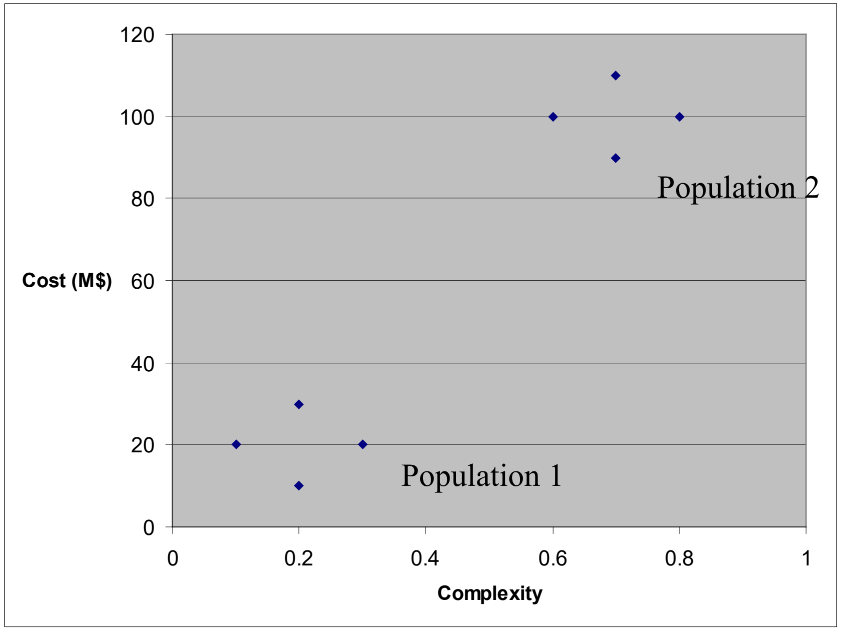


Remarks on Non-NASA Missions

- Non-NASA missions included in study:
  - represent a different culture
  - smaller than the average NASA mission
  - less complex than the average NASA mission
  - shorter in duration than the average NASA mission
- Consequence: including the non-NASA missions seems to overstate goodness of fit due to pseudo-correlation
- Conclusion: the non-NASA missions represent a different population and should not be included in the study of science-related missions

Pseudo-correlation: An Example

Comparing two distinct populations can introduce pseudo-correlation



Correlation within each population is zero, but considered as one population, correlation is 95%!

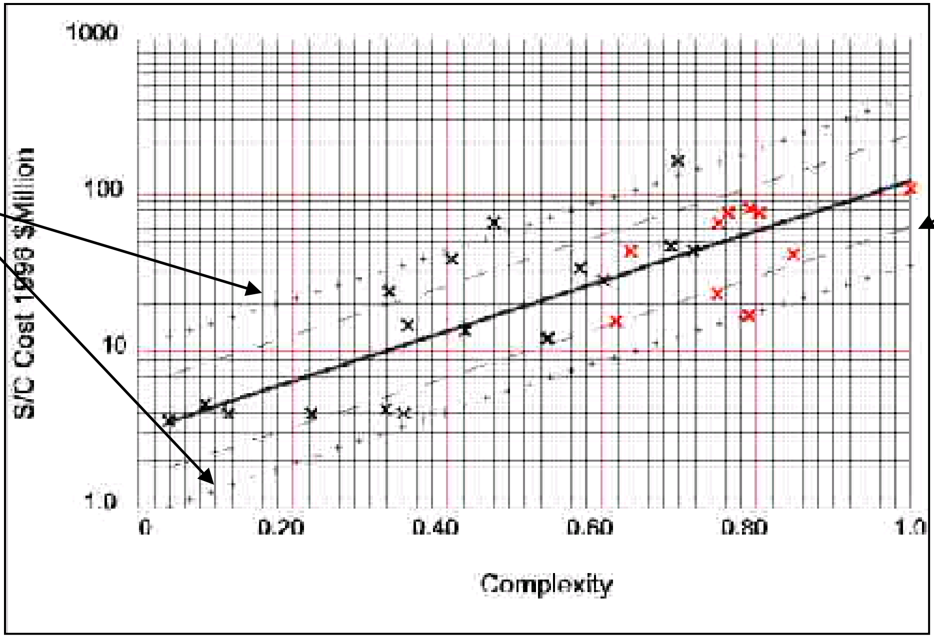
Regression Models

- Regression lines represent an average. An average is not sufficient information for parametric indicators.
- Must consider uncertainty – e.g., standard deviations, confidence limits
- Missions below the *average minus two standard deviations* could be considered to be too risky (for example)
- Another possibility: treat failure as a dummy variable – minimum distance between lines needed for significant difference

Regression Models (Cont'd.)

In the graph below, we have added error bands (average +/- 2 standard deviations), and 95% confidence limits

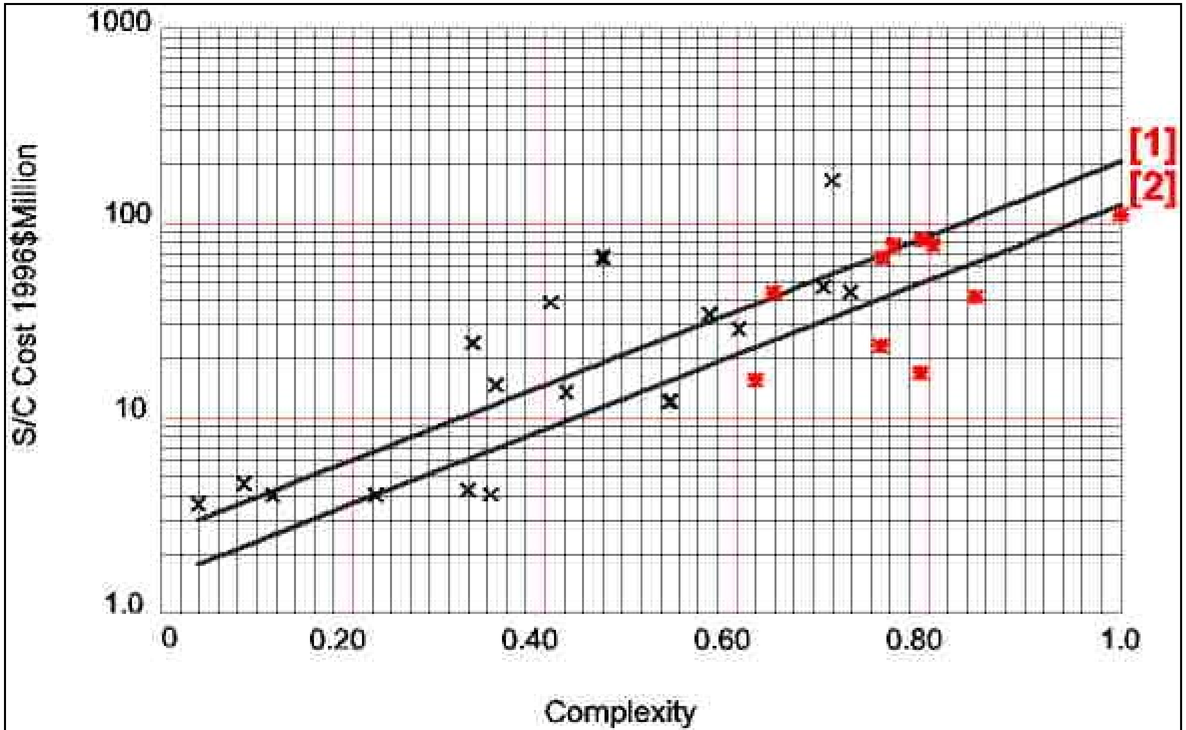
95% confidence limits



Error bands

Regression Models (Cont'd.)

Failure treated as a dummy variable



[1] REGULAR (x)

$S = 2.523 * (81.645)^{C}$

[2] FAILED (*)

$S = 1.502 * (81.645)^{C}$

CC = 0.826
SEA = 0.665
RSQ = 0.682
N = 28

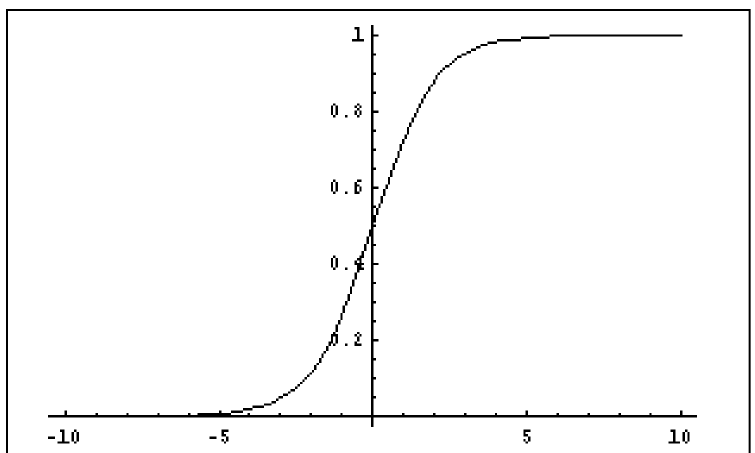
Lessons Learned

- May 2000 Aerospace model is a pioneering effort and a terrific first step, but there is room for improvement
- Consideration of non-NASA missions questionable
- Discrepancies in data
- Results sensitive to small changes in data
- Complexity metric needs more study – driven by weight?
- Model not directly related to risk
- Simple regression line not sufficient for parametric indicators
- Interdependency between cost and schedule not considered

Goddard Improvements

- Risk should be modeled directly
  - We have completed some **preliminary** work in this direction
  - Used over 50 NASA missions to directly predict risk, using logistic regression

For one independent variable, the model has the form:



Goddard Improvements (Cont'd.)

- Logistic regression
  - Algebraically, the logistic regression model has the form:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

where $g(x) = \beta_0 + \beta_1 * x$

- Logistic regression arose in epidemiological research, and is now commonly employed in business and finance, ecology, engineering, health policy, and linguistics

Goddard Improvements (Cont'd.)

- Directly measured the impact of cost, schedule, and spacecraft and payload complexity, reliability, and new technology advances
  - Based on our findings, these predictors better predict risk than spacecraft complexity alone
- Our model takes into account the relationship between cost and schedule (treated as independent in the May 2000 Aerospace model)

Goddard Improvements (Cont'd.)

- Predicts the probability of the return of scientific data - "mission success"
- A NASA metric, the level of technology maturation, is a better predictor of success/failure than "spacecraft complexity"
  - *Findings indicate that "spacecraft complexity" has little correlation with risk of mission failure*
- Model was tested and validated with Authority to Proceed (ATP) data

Goddard Improvements (Cont'd.)

Model Performance

Mission Result	Mission Success Level
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	99%
Successful	98%
Successful	98%
Successful	98%
Successful	98%
Successful	97%
Successful	97%
Successful	97%
Successful	97%
Successful	97%

Mission Result	Mission Success Level
Successful	96%
Successful	95%
Successful	95%
Successful	93%
Successful	93%
Successful	92%
Successful	92%
Successful	92%
Successful	91%
Successful	91%
Successful	91%
Successful	91%
Failure	**88%**
Successful	84%
Successful	79%
Successful	70%
Successful	67%
Failure	**65%**
Successful	61%
Successful	58%
Failure	**55%**
Successful	50%
Failure	**45%**
Failure	**41%**
Successful	34%
Successful	28%
Failure	**16%**

80% and above	"Fly" zone
51 -79%	"Caution" zone
50% and below	"No-fly" zone

Goddard Improvements (Cont'd.)

Model Performance

- The statistic $G = -2\ln[\frac{\text{likelihood without the variables}}{\text{likelihood with the variable}}]$ plays a central role in assessing goodness-of-fit for logistic regression. For our model, G = 15.615. G is X^2-distributed with six degrees of freedom. $Pr(X^2(6)>15.615)=0.015977$, which means the model is statistically significant up to the 97.5% confidence level.
- R^2 is not a very meaningful statistic for a logistic regression model, which is not exactly a regression model. Significant logistic regression models with good fits often have very low R^2 values, in the range of 0.1 - 0.2. For our model, $R^2 = 0.294$, which is quite good for a logistic regression model.
- The area under the ROC(Receiver Operating Characteristic) curve is 93%, which means the model has outstanding discrimination ability. Also, according to a ROC analysis, the optimum mission success level to distinguish between success and failure is 80%, the threshold for our "fly" zone.
- The Hosmer-Lemeshow statistic, C, has a p-value equal to 99.4%, which indicates an excellent fit.

Goddard Improvements (Cont'd.) Model Validation

- Mission success levels calculated at both ATP and at Completion:

ATP	Completion
99%	97%
99%	99%
99%	99%
96%	99%
96%	97%
91%	99%
90%	95%
90%	91%
87%	91%
82%	97%
42%	61%
41%	58%
17%	N/A

Summary and Conclusions

- Replicated May 2000 Aerospace Study with Goddard data
  - Obtained different results
    - Schedule has little relationship with spacecraft complexity
    - Cost has a positive correlation with spacecraft complexity, but only about half of the failed missions have a lower than average cost for a given complexity level
  - Our results are consistent with Goddard's experience, and do not invalidate "faster, better, cheaper"

Summary and Conclusions (Cont'd.)

- Developed a tool with Goddard data that is used for decision-making within Goddard
  - Model is used by the RAO to independently assess mission risk with respect to cost,and schedule
  - RAO employs the model in developing independent assessments as part of the mission confirmation process
    - Mission confirmation is the most important review during a mission's development
      - Point at which Goddard commits to cost and schedule