Cost Growth Models for NASA'S Programs: A Summary

Tze-San Lee
Western Illinois University
Macomb, Illinois 61455
(Tze-San_Lee@ccmail.wiu.edu)

L. Dale Thomas
National Aeronautics and Space Administration
Marshall Space Flight Center
Huntsville, Alabama 35812
(dale.thomas@msfc.nasa.gov)

Abstract. Under two cost growth indices, annual absolute and relative cost growth, probability models were constructed for basis functions of NASA's different technology readiness levels by using Johnson's four parameter family of bounded, unbounded, or lognormal distributions. In addition, statistical prediction models were built for the programs. The result of this research showed that the program's annual absolute cost growth was estimated to grow at a rate of 5.9% of the program's initial cost estimate with a standard error of 0.3%, while the program's annual relative cost growth was estimated to grow at a rate of 1.6% of the weighted average of technology readiness level from the program's components with a standard error of 0.3%.

Key words: Cost growth indices, Johnson's four-parameter distributions, maximum likelihood estimation, multiple linear regression model, NASA's technology readiness levels.

1. INTRODUCTION

The goal of this paper is to build probability and statistical models for the cost growth of NASA's programs. To build probability models for the program, our approach which basically followed along the same line used in the COMPRE (Complex Organizational Metric for Programmatic Risk Environment) model was to construct basis functions for different levels of technology maturity of the program's subsystems/components. The idea of COMPRE model was to break down the whole program into subsystems/components and then rated the technology maturity of each subsystem/component using a scale of 1 to 9 of NASA's Technology Readiness Levels (TRL's), where 1 and 9 represents the lowest and highest degree of the technology maturity (see Table 1). COMPRE model was developed by L. D. Thomas and R. A. Mog with an intention to serve as a decision making tool for the NASA's program managers (Mog 1997, SAIC 1997). Mog (1977) adopted the schedule duration as an indirect measure of the cost growth and used the family of exponential functions to construct the basis functions for different TRLs. In this paper we took a totally different approach. We first proposed two new measures, annual absolute and relative cost growth, for measuring the cost growth and then constructed probability models for basis functions for TRL 3-9,whereas statistical prediction models for the entire programs.

2. DATA

The data were taken from Resource Data Storage and Retrieval data base (REDSTAR), established in 1971 as a NASA's major repository of cost, technical, and programmatic information pertaining to the space program, which was maintained by the Science Applications International Corporation (SAIC). In total, the data from 31 programs were used in this paper. For each component of the program, the data were collected on five variables: ICE (Initial Cost Estimate, $million), FTC (Final Total Cost, $million), IDE (Initial Duration Estimate, year), FTD (Final Total Duration, year), and TRL (Technology Readiness Level). We first grouped together those components having the same TRL from different programs. Then, for each TRL, two new variables, ACG (annual Absolute Cost Growth, $million per year) and RCG (annual Relative Cost Growth, fraction per year), are created, respectively, by setting ACG = (FTC – ICE)/FTD and RCG = ACG/ICE. To save the space, the data of ACG and RCG for different TRL are omitted, but can be found as Tables 2-3 in Lee-Thomas (2000). Among 31 programs, no component of all the programs was rated as TRL = 1. For TRL = 2 & 9, there were only two observations. Consequently, basis functions could not be constructed for these three levels. Again, the data for 28 entire programs are omitted here, but can be found as Table 6 in Lee-Thomas (2000). The data of WTRL (Weighted average of TRL for the entire program) were obtained by calculating the weighted average of TRL of each component multiplied by their corresponding percent of the allocation cost against the entire program's cost.

3. METHOD

3.1 Johnson's Family of Distributions for TRLs

After plotting the histograms for the ACG and RCG data for different TRLs, all histograms were not symmetric and had exhibited some skewness (Lee-Thomas (2000)). We decided to adopt Johnson's 4-parameter families of [bounded, unbounded, and lognormal] distributions as probability models for both ACG and RCG. As far as selection of the appropriate Johnson's distribution was concerned, we followed the selection criteria set up by Slifker-Shapiro (1980). However, we did not use their equal-distance quantile formula for parameter estimation since the results were unsatisfactory. Instead, the method of maximum likelihood estimation was used to estimate the model parameters.

Basically, Johnson's family of distributions is comprised of three subfamilies: (i) bounded (X_B), (ii) unbounded (X_U), and (iii) lognormal (X_L) and their probability density functions are given, respectively, by

(i) $$f_B(x) = \frac{\eta\lambda}{\sqrt{2\pi}[(x-\varepsilon)(\lambda+\varepsilon-x)]}\exp(-\frac{1}{2}[\gamma+\eta\ln(\frac{x-\varepsilon}{\lambda+\varepsilon-x})]^2),\ \varepsilon < x < \lambda+\varepsilon, \tag{1}$$

where $\lambda > 0$, $\eta > 0$, $-\infty < \varepsilon < \infty$, $-\infty < \gamma < \infty$.

(ii) $$f_U(x) = \frac{\eta}{\sqrt{2\pi[(x-\varepsilon)^2+\lambda^2]}}\exp(-\frac{1}{2}[\gamma+\eta\sinh^{-1}(\frac{x-\varepsilon}{\lambda})]^2),\ -\infty < x < \infty, \tag{2}$$

where $\lambda > 0$, $\eta > 0$, $-\infty < \varepsilon < \infty$, $-\infty < \gamma < \infty$.

(iii) $$f_L(x) = \frac{\eta}{\sqrt{2\pi}\cdot(x-\varepsilon)}\exp(-\tfrac{1}{2}[\gamma+\eta\ln(x-\varepsilon)]^2),\ x > \varepsilon, \tag{3}$$

where $\eta > 0$, $-\infty < \varepsilon < \infty$, and $-\infty < \gamma < \infty$.

Incidentally, the mean and variance of X_B can be obtained from the k^{th} moment of X_B which is given as follows:

$$\mu_{k;X_B} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[\lambda+\varepsilon-\frac{\lambda}{1+e^{(z-\gamma)/\eta}}]^k e^{-z^2/2}dz. \tag{4}$$

Although it is possible to obtain a closed form expression for (4) [also see Johnson (1949), equations (56) and (57)], they are very complicated and not useful for numerical calculations. Instead, we apply the 11-point Gauss-Hermite numerical quadrature in Stroud-Secrest (1966) to evaluate (4).

In contrast to X_B, the mean and variance of X_U have simple closed forms as follows (Johnson, et al 1994):

$$\mu_{X_U} = \varepsilon+\lambda\cdot\exp((2\eta^2)^{-1})\cdot\sinh(\gamma/\eta), \tag{5a}$$

and

$$\sigma^2_{X_U} = \frac{\lambda^2}{2}[\exp((2\eta^2)^{-1})-1][\exp((2\eta^2)^{-1})\cdot\cosh(2\gamma/\eta)+1]. \tag{5b}$$

Similarly, the mean and variance of X_L have simple closed forms as follows (Johnson, et al 1994):

$$\mu_{X_L} = \varepsilon+\exp(\frac{1-2\gamma\eta}{2\eta^2}), \tag{6a}$$

and

$$\sigma_{X_L}^2 = \exp(\frac{1-2\gamma\eta}{\eta^2})\cdot[\exp(\frac{1}{\eta^2})-1]\,. \quad (6b)$$

In addition, since both the distributions of X_U, X_B, and X_L are skewed, their medians, given, respectively, by (Johnson 1949)

Median of $X_U = \varepsilon - \lambda\sinh(\gamma/\eta)$, (7)

Median of $X_B = \varepsilon + \lambda/(1+\exp(\gamma/\eta))$, (8)

and

Median of $X_L = \varepsilon + \exp(-\gamma/\eta)$, (9)

are also calculated as a comparison against their means.

3.2 Selection Criteria for Johnson's Families

Let x_{3z}, x_z, x_{-z}, and x_{-3z} be the estimates of the 95th, 71st, 29th, and 5th percentiles of Johnson's random variable from the given sample data, where z was chosen as 0.5483 so that 3z (= 1.645) is the 95th percentile of standard normal distribution. Then, selecting one of the three Johnson's (unbounded, bounded, and lognormal) distributions to fit the data is based upon the following criterion:

define

$m = x_{3z} - x_z$, (10a)

$n = x_{-z} - x_{-3z}$, (10b)

$p = x_z - x_{-z}$, (10c)

$q = mn/p^2$. (10d)

(a) If $q > 1$, then select Johnson's unbounded distribution X_U;

(b) If $q < 1$, then select Johnson's bounded distribution X_B;

(c) If $q = 1$, then select Johnson's lognormal distribution X_L.

Note that since m, n, and p are random quantities, q is also a random quantity. Therefore, the probability of $q = 1$ is, in theory, nil. However, this does not rule out the possibility that X_L might be a more appropriate model for the data than either X_B or X_U if the calculated q value is close to 1. Indeed, this was the case in some of the models for the cost growth data as shown in Section 4.

3.3 Maximum Likelihood Estimation

Generally, it is difficult to estimate the unknown model parameters by the principle of the maximum likelihood if they are embedded in the domain of the probability density function. For Johnson's family, both its bounded and lognormal subfamilies have indeed the unknown model parameters occurring in the domain of the probability density function defined by Eqs. (1) and (3), respectively. A difficulty in employing the maximum likelihood estimation for the unknown Johnson's model parameters was long recognized by Johnson (1949). This is the reason why he instead used the method of moment and quantile estimation for estimating his unknown model parameters. Although some authors did try to apply the principle of maximum likelihood to estimate the unknown parameters of Johnson's bounded distribution (Olsson (1979), Siekierski (1992)), either they dealt with only the grouped data or their method had an error. Here, based upon the method of nonlinear programming (Luenberg (1984)), a complete solution for applying the principle of maximum likelihood to estimate the unknown model parameters of all of Johnson's families was obtained as follows:

(i) Given that the sample data $\{x_1, x_2, ..., x_n\}$ are fixed, the likelihood function of $\vec{\theta} = (\gamma, \eta, \lambda, \varepsilon)$ for X_B is given as follows:

$$L_B(\vec{\theta}) = \prod_{i=1}^{n}\{\frac{1}{\sqrt{2\pi}}\cdot\frac{\eta\lambda}{(x_i-\varepsilon)(\lambda+\varepsilon-x_i)}\exp(-\frac{1}{2}[\gamma+\eta\ln(\frac{x_i-\varepsilon}{\lambda+\varepsilon-x_i})]^2)\}\,. \quad (11)$$

Let $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$, and $\hat{\varepsilon}$ denote, respectively, the maximum likelihood estimators for γ, η, λ, and ε. Then, $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$, and $\hat{\varepsilon}$ must satisfy the following constrained maximization problem:

Maximize $L_B(\vec{\theta})$ (12a)

Subject to

$$\varepsilon < x_{(1)}\,, \tag{12b}$$
$$\lambda > 0\,, \tag{12c}$$
$$\eta > 0\,, \tag{12d}$$

where $L_B(\vec{\theta})$ is given by (11).

(ii) Given that the sample data $\{x_1, x_2, ..., x_n\}$ are fixed, the likelihood function of $\vec{\theta} = (\gamma, \eta, \lambda, \varepsilon)$ for X_U is given as follows:

$$L_U(\vec{\theta}) = \prod_{i=1}^{n} \{ \frac{\eta}{\sqrt{2\pi[(x_i-\varepsilon)^2+\lambda^2]}} \exp(-\frac{1}{2}[\gamma + \eta \sinh^{-1}(\frac{x_i-\varepsilon}{\lambda})]^2) \}\,. \tag{13}$$

Let $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$, and $\hat{\varepsilon}$ denote, respectively, the maximum likelihood estimators for γ, η, λ, and ε. Then, $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$, and $\hat{\varepsilon}$ must satisfy the following constrained maximization problem:

$$\text{Maximize } L_U(\vec{\theta}) \tag{14a}$$

Subject to

$$\lambda > 0\,, \tag{14b}$$
$$\eta > 0\,, \tag{14c}$$

where $L_U(\vec{\theta})$ is given by (13), and $x_{(i)}$ denote the i[th] order statistics of $\{x_i\}$, i.e., $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$.

(iii) Given that the sample data $\{x_1, x_2, ..., x_n\}$are fixed, the likelihood function of $\vec{\theta} = (\gamma, \eta, \varepsilon)$ for X_L is given by using (13) as follows:

$$L_L(\vec{\theta}) = \prod_{i=1}^{n} \{ \frac{1}{\sqrt{2\pi}} \cdot \frac{\eta}{x_i-\varepsilon} \exp(-\tfrac{1}{2}[\gamma + \eta \ln(x_i-\varepsilon)]^2) \}\,. \tag{15}$$

Let $\hat{\gamma}$, $\hat{\eta}$, and $\hat{\varepsilon}$ denote, respectively, the maximum likelihood estimators for γ, η, and ε. Then, $\hat{\gamma}$, $\hat{\eta}$, and $\hat{\varepsilon}$ must satisfy the following constrained maximization problem:

$$\text{Maximize } L_L(\vec{\theta}) \tag{16a}$$

Subject to

$$\varepsilon < x_{(1)}\,, \tag{16b}$$
$$\eta > 0\,, \tag{16c}$$

where $L_L(\vec{\theta})$ is given by (15).

3.4 Statistical Prediction Models

Two multiple linear regression models without intercept were built for the entire program's data as follows:

$$ACG = \alpha_1 \cdot ICE + \alpha_2 \cdot IDE + \alpha_3 \cdot WTRL + e_A\,, \tag{17}$$

and

$$RCG = \beta_1 \cdot IDE + \beta_2 \cdot WTRL + e_R\,, \tag{18}$$

where e_A and e_R denoted unobservable errors for ACG and RCG, respectively. Assume that $e_A \sim N(0, \sigma_A^2)$ and $e_R \sim N(0, \sigma_R^2)$.

4. RESULTS and DISCUSSIONS

Results of modeling TRL3 – TRL8 under ACG and RCG are given, respectively, in Tables 2 and 3. Those numbers listed in the first eight rows of Tables 2 and 3 were obtained, respectively, from applying Eqs. (10a-10d) to fit the ACG and RCG data. Then, based on the q values in the eighth row of Tables 2 and 3, appropriate Johnson's distrinution were chosen as probability models for TRLs. However, there were two exceptions for TRL6 under ACG and TRL5 under RCG. Since their q values, 0.743 and 0.941, were close to 1, we decided to fit X_L to the data too. After fitting both X_B and X_L to the data, it was observed that X_L was more reasonable than X_B by examining their corresponding means and standard deviations. The values of $\hat{\gamma}$, $\hat{\eta}$, $\hat{\lambda}$, and $\hat{\varepsilon}$ listed in the tenth row to the thirteenth row of Tables 2 and 3 were obtained by using Eqs. (124a-12d), or (14a-14c), or (16a-16c) depending on whether Johnson's model was X_U, or X_B, or X_L. Note that the parameter λ was not used in the probability distribution of X_L. Hence, the values of $\hat{\lambda}$ were not listed for TRL6 under ACG and TRL5 under RCG. The values of mean and standard deviation listed in the fourteenth and sixteenth row of Tables 2 and 3 were obtained by substituting the maximum likelihood estimators for the unknown parameters of Eqs. (4), or (5a-5b), or (6a-6b) into the corresponding Johnson's models. The value of median listed in the fifteenth row of Tables 2 and 3 were obtained by using Eqs. (7-9) depending on what the corresponding Johnson's model was. The values of $\bar{x}$ and s listed in the seventeenth and eighteenth row of Tables 2 and 3 are the sample mean and sample standard deviation.

Under ACG, the mean, median, and standard deviation of all TRL, except TRL6, apparently had a decreasing trend as the level of TRL increased (Table 2). For TRL6, the anomaly was attributed to a single large observation (192.329) in the data set. If the observation of 192.329 was not used in the model fitting, then the q value became 0.825, and the mean, median, and standard deviation for TRL6 after re-calculation became 1.379, 1.259, and 2.450, respectively. Both the mean and standard deviation were changed from 3.773 and 8.037 to smaller values 1.379 and 2.450, respectively. However, since the data of 192.329 was a genuine observation, we preferred to retain it in the model fitting. Similarly, a general decreasing trend seemed to exist under RCG as the level of TRL increased (Table 3). The larger standard deviation for TRL6 and TRL8 under RCG indicated that the judgmental ranking for these two levels was comparatively less accurate.

Additionally, if we compared the values of $\bar{x}$ and s listed in the seventeenth and eighteenth row of Tables 2 and 3 against their corresponding mean and standard deviation in the fourteenth and sixteenth row of the same table, we noticed that they agreed very closely with each other. In particular, they agreed more closely under RCG than under ACG. This indicated that although selections of Johnson’s probability models for different TRLs were not the same under both ACG and RCG, probability models obtained in Tables 2 and 3 were accurately reflecting the characteristic of their data.

After fitting, respectively, Eqs. (17) and (18) to the data, results of numerical calculation were given in Tables 4(a) and 5(a). Judging from their p-values, ICE and WTRL were the only significant predictor variables for ACG and RCG, respectively. Hence, prediction equations for ACG and RCG were fitted again to the data by using ICE and WTRL only. From Tables 4(b) and 5(b), predicted regression models are thus given, respectively, by

$$A\hat{C}G = 0.059 \cdot ICE \,, \qquad (19)$$

and

$$R\hat{C}G = 0.016 \cdot WTRL \,, \qquad (20)$$

where $A\hat{C}G$ and $R\hat{C}G$ denote, respectively, the predicted value of ACG and RCG.

From Table 4(b), the value of the test statistics was

$$t = (\bar{e}_A - E(e_A)) / (\hat{\sigma}_A / \sqrt{n}) = (3.167 - 0) / (26.021 / \sqrt{28}) = 0.64$$

which was not significant at the level of $\alpha = 0.05$. Hence, the model assumption on e_A in Eq. (17) was satisfied. Similarly, the model assumption on e_R in Eq. (18) was proved to be satisfied because the value of the test statistics from Table 5(b) was - 0.27 which was not significant either at the level of $\alpha = 0.05$.

Also, we noticed that the values of R^2 were 0.933 and 0.604 for Eqs. (19) and (20), respectively (Tables 4(b) and 5(b)). Hence, 93.3% and 60.4% of variation in the observed data of ACG and RCG were accounted for variation in ICE and WTRL, respectively. Still, we wanted to see how well the prediction models for ACE and RCG performed. Observed and predicted values of ACG and RCG were, respectively, listed in Table 6, where predicted values of ACG and RCG were calculated according to Eqs. (19) and (20). Under ACG, Programs 2, 3, 11, 19, 28 were identified as the outliers in fitting Eq. (19) to data. Program 11 under-predicted the actual annual absolute cost growth by an amount of \$88.011 (= 130.611 - 42.6) million per year. This was the largest under-predicted amount of ACG among 28 programs. Programs 2 and 3 also under-predicted the actual annual absolute cost growth by \$42.484 and \$46.304 million per year, respectively, while Programs 19 and 28 over-predicted the actual annual absolute cost growth by \$38.636 and \$56.383 million per year, respectively. Under RCG, no programs were identified as the outliers in the process of fitting Eq. (20) to data. Note that the largest under-predicted fraction of RCG among 28 programs was Program 14 which under-predicted the actual annual relative cost growth by 17.3% (= 30.1% - 12.8%).

5. CONCLUSIONS

The idea of COMPRE was employed to build cost growth models for NASA's programs. Two cost growth indices, annual absolute and relative cost growth, were used as measures for the cost growth of NASA's programs. NASA's technology readiness levels were used to break the entire program into subsystems/components. Then, Johnson's four-parameter families of bounded, unbounded, or lognormal distributions were applied to build probability models for basis functions of different technology readiness levels. The method of maximum likelihood estimation was employed to estimate the unknown parameters of Johnson's models. The obtained probability models were considered to be very good because they accurately reflected the characteristic of the observed data. In addition, statistical prediction models were obtained, respectively, under the two cost growth indices based on the data of the twenty-eight programs. It was shown that the program's annual absolute cost growth was estimated to grow at a rate of 5.9% of the program's initial cost estimate with a standard error of 0.3%, while the program's annual relative cost growth was estimated to grow at a rate of 1.6% of the weighted average of technology readiness level from the program's components with a standard error of 0.3%.

Table 1 Definitions of Technology Readiness Level used by NASA.

Level	Definition
TRL 1	Basic principles observed and reported
TRL 2	Conceptual design formulated
TRL 3	Conceptual design tested analytically or experimentally
TRL 4	Critical functions and characteristics demonstrated
TRL 5	Component/Brassboard tested in relevant environment
TRL 6	Prototype/engineering model tested in relevant environment
TRL 7	Engineering model tested in space
TRL 8	"Flight-qualified" system
TRL 9	"Flight-proven" system

Table 2 Parameter estimates of probability models for different levels of TRL under ACG.

	ACG					
	TRL3	TRL4	TRL5	TRL6	TRL7	TRL8
x_{3z}	37.665	29.672	24.043	6.596	6.849	0.873
x_z	9.572	6.455	4.789	2.568	2.663	0.411
x_{-z}	1.074	0.766	0.169	0.047	-0.166	-0.988
x_{-3z}	-0.206	-2.312	-6.363	-1.126	-0.807	-7.169
m	28.093	23.217	19.254	4.028	4.186	0.462
n	1.280	3.078	6.532	1.173	0.641	6.181
p	8.498	5.689	4.620	2.521	2.829	1.399
q	0.498	2.208	5.892	0.743	0.335	1.460
Johnson's model	X_B	X_U	X_U	X_L	X_B	X_U
$\hat{\eta}$	0.565	1.866	2.422	1.364	0.661	408.781
$\hat{\gamma}$	1.001	-4.355	-4.115	-2.708	0.713	-2148.632
$\hat{\lambda}$	46.102	2.5	5.7	-	8.738	10.5
$\hat{\varepsilon}$	-0.744	-8.911	-12.118	-5.754	-1.234	-1007.669
Mean	18.248	5.838	4.286	3.773	2.598	-1.008
Median	5.956	3.865	2.946	1.528	0.983	-1.005
Standard deviation	9.370	8.648	7.524	8.037	1.799	2.463
$\bar{x}$	9.81	5.974	4.370	7.206	1.342	-1.004
s	13.025	9.531	8.435	33.326	2.461	2.624

Table 3 Parameter estimates of probability models for different levels of TRL under RCG.

	RCG					
	TRL3	TRL4	TRL5	TRL6	TRL7	TRL8
x_{3z}	0.637	0.566	0.399	1.004	0.397	1.091
x_z	0.139	0.093	0.146	0.134	0.225	0.369
x_{-z}	0.031	0.022	0.004	0.004	-0.033	-0.059
x_{-3z}	-0.003	-0.033	-0.071	-0.093	-0.070	-0.144
m	0.498	0.473	0.253	0.870	0.172	0.722
n	0.034	0.055	0.075	0.097	0.037	0.085
p	0.108	0.071	0.142	0.130	0.258	0.428
q	1.452	5.161	0.941	4.993	0.096	0.335
Johnson's model	X_U	X_U	X_L	X_U	X_B	X_B
$\hat{\eta}$	0.769	1.837	1.835	6.306	0.898	0.940
$\hat{\gamma}$	-3.790	-2.890	2.663	-4.963	1.599	0.711
$\hat{\lambda}$	0.001	0.09	-	1.5	1.096	1.515
$\hat{\varepsilon}$	-0.010	-0.140	-0.172	-1.122	-0.124	-0.294
Mean	0.151	0.101	0.100	0.201	0.120	0.265
Median	0.059	0.068	0.062	0.184	0.034	0.190
Standard deviation	0.338	0.153	0.152	0.320	0.171	0.298
$\bar{x}$	0.152	0.106	0.100	0.203	0.087	0.228
s	0.321	0.185	0.166	0.364	0.163	0.465

Table 4 Parameter estimates of a prediction model of Eq. (17) for 28 programs.

(a)

ACG = ICE + IDE + WTRL + e_A					
Predictor variable	$\hat{\alpha}_i$ (s.e.($\hat{\alpha}_i$))	p-value	R^2	$\bar{e}_A$	$\hat{\sigma}_A$
ICE	0.059 (0.004)	0.000	0.934	0.064	25.991
IDE	0.204 (3.042)	0.947			
WTRL	0.488 (2.305)	0.834			

(b)

ACG = ICE + e_A					
Predictor variable	$\hat{\alpha}_i$ (s.e.($\hat{\alpha}_i$))	p-value	R^2	$\bar{e}_A$	$\hat{\sigma}_A$
ICE	0.059 (0.003)	0.000	0.933	3.167	26.021

Table 5 Parameter estimates of a prediction model of Eq. (18) for 28 programs.

(a)

RCG = IDE + WTRL + e_R					
Predictor variable	$\hat{\beta}_i$ (s.e. ($\hat{\beta}_i$))	p-value	R^2	$\bar{e}_R$	$\hat{\sigma}_R$
IDE	-0.007 (0.008)	0.345	0.617	-0.002	0.075
WTRL	0.022 (0.006)	0.002			

(b)

RCG = WTRL + e_R					
Predictor variable	$\hat{\beta}_i$ (s.e. ($\hat{\beta}_i$))	p-value	R^2	$\bar{e}_R$	$\hat{\sigma}_R$
WTRL	0.016 (0.003)	0.000	0.604	-0.004	0.077

Table 6 Observed and predicted values from Eqs. (19) and (20) of ACG and RCG for the 28 programs.

	ACG		RCG	
Program	Observed	Predicted	Observed	Predicted
1	476.595	484.767	0.058	0.081
2	140.332	97.848	0.085	0.065
3	70.212	23.908	0.174	0.081
4	45.416	27.916	0.097	0.085
5	22.551	32.314	0.041	0.072
6	13.23	28.939	0.027	0.074
7	31.52	10.289	0.182	0.075
8	0	10.089	0	0.081
9	0	0.89	0	0.078
10	0.332	0.23	0.086	0.092
11	130.661	42.6	0.182	0.077
12	12.445	17.021	0.043	0.076
13	4.994	7.658	0.039	0.111
14	18.395	3.632	0.301	0.128
15	3.12	1.389	0.133	0.115
16	11.412	2.614	0.259	0.115
17	3.296	5.419	0.036	0.121
18	6.564	1.651	0.236	0.110
19	25.323	63.959	0.023	0.076
20	0.764	11.122	0.004	0.097
21	2.61	3.162	0.049	0.087
22	7.223	5.059	0.085	0.125
23	11.426	6.104	0.111	0.106
24	32.491	21.476	0.09	0.116
25	15.308	22.837	0.04	0.069
26	24.067	17.699	0.081	0.09
27	1.973	16.557	0.007	0.063
28	-16.4	39.983	-0.024	0.082

ACKNOWLEDGEMENTS

This research was partially supported by an NASA's grant NAG8-1611 to T-S. Lee for the period of May 13, 1999 – May 13, 2000. Also, the first author would like to thank Ms. T. Clayborn who kindly provided the data to be used in the statistical analysis. The computer software, MAPLE and SYSTAT, were used to facilitate numerical calculations and data analysis.

REFERENCES

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. Biometrika, **36**, 149-176.

Johnson, N.I., Kotz, S., and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1, second edition, John Wiley & Sons, Inc., New York.

Lee, T-S. and Thomas, L. D. (2000). Cost Growth Models for NASA's Programs, Final Report, Marshall Space Flight Center, Huntsville, AL.

Luenberger, D. G. (1984), Linear and Nonlinear Programming, Second Edition, Addison Wesley Publishing Company, Reading, Massachusetts.

Montgomery, D.C. and Peck, E.A. (1982). Introduction to Linear Regression Analysis, John Wiley & Sons, New York.

Mog, R.A. (1997). Systems Engineering Metrics: Organizational Complexity and Product Quality Modeling, Final Report, OR Applications, Huntsville, AL.

Olsson, D. M. (1979). Fitting Johnson's S_B and S_U systems of curves using the method of maximum likelihood, Journal of Quality Technology, **11**, 211-217.

SAIC (1997). Complex Organizational Metric for Programmatic Risk Environments Model (COMPRE'), Final Report, Huntsville, AL.

Slifker, J. F. and Shapiro, S. S. (1980). The Johnson System: selection and parameter selection, Technometrics, **22**, 239-246.

Siekierski, K. (1992). Comparison and evaluation of three methods of estimation of the Johnson S_b distribution, Biometrical Journal, **34**, 879-895.

Stroud, A. H. and Secrest, D. (1966). Gaussian Quadrature Formulas, Prentice-Hall, Inc., Englewood Cliffs, N.J.

Thomas, L. D. and Mog, R. A. (1998). A quantitative metric of system development complexity, unpublished manuscript.