Spring 5-2016

Examining Cost Functionality and Optimization: A Case Study on Testing the Reasonableness of New Aircraft using Historical Aircraft Data

Katherine Jozefiak
Washington University in St. Louis

Recommended Citation

WASHINGTON UNIVERSITY IN ST. LOUIS

University College

Statistics

Examining Cost Functionality and Optimization: A Case Study on Testing the Reasonableness of

New Aircraft using Historical Aircraft Data

by

Katherine Jozefiak

A thesis presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

<u>Acknowledgments</u>

ABSTRACT OF THE THESIS

Examining Cost Functionality and Optimization: A Case Study on Testing the Reasonableness of

New Aircraft using Historical Aircraft Data

by

Katherine Jozefiak

Master of Arts in Statistics

Washington University in St. Louis, 2016

Professor Siddhartha Chib, Chair

When pursuing business by competing for government contracts, proving the submitted price is reasonable is often required. This proof is called a test of reasonableness. This study analyzes data from historical aircraft programs in relation of a new aircraft program in order to demonstrate the estimated cost of the new program is reasonable. The purpose of this study is to investigate three questions. Is the new program cost reasonable using current industry and government parameters? Is it better to look at programs from a total cost perspective or break the total cost into subcategory levels? Finally, this study applies a log linear model, Cobb Douglas model, and fitted linear model to the data. Does a proposed method provide greater statistical significance than the others?

Chapter 1: Introduction

When the government wants to purchase a new product, for example an aircraft, they will often have a competition. In a government competition companies that want to produce the solicited product put together a proposal. The proposal includes a best price offer and technical framework. This study will focus on justifying the best price offer to the customer. Justifying the best price offer requires a test of reasonableness. A test of reasonableness (TOR) uses historical program data to support the proposed price for a new program. The TOR must be understandable and clearly demonstrate the proposed price is achievable by the company.

The definition of a TOR has changed with the economy. The government answers to the tax payers for overspending. This is exemplified by past programs that spent significantly more than forecasted costs. In an economy where the government customer wants technically advanced products for the lowest price, competitions have become a crucial part of the proposal process. Companies strive to optimize production and reduce cost. This results in a unique TOR. It is no longer acceptable to find average historical cost or assume the same cost as a historical program economically adjusted for years. Instead the focus is on proving a new program can be achieved for less cost than historical programs. It is important to note the challenge of proving a new program can be operated at less cost, essentially targeting the lower range of historical cost.

Three models were selected for the purposes of this study to analyze new program cost and historical program cost. A log linear regression, Cobb-Douglas, and fitted linear regression model. Using a log linear regression model is one suggested method to determine proportional cost relationships (Corro and Shim 2000). The use of logarithmic scale helps normalize the data and creates symmetry. Cobb-Douglas is considered because it is a popular method for representing the relationship between multiple inputs (Henningsen and Henningsen 2011). The

fitted linear regression model fits the observed variables to the explanatory or predictor variables (Gelman and Hill 2009). This method gives us the benefit of applying least squares which will minimize the error term. We will further explore these models in Chapter 3.

This case study explores three questions. First, is the new aircraft program cost reasonable? This is key when supporting the TOR. Second, is there more meaning in examining program data at a total cost level or at subcategory levels. For the purpose of this project the subcategories are as follows: Structure and Subsystem Design, Flight Test, Mission Systems, Support and Training, Systems Engineering, Test and Validation, Manufacturing, and Program Management. These are common aircraft categories. The final question is which of the fore mentioned models is most significant for a TOR.

Due to its sensitive nature the data used in this case study was randomly generated. The actual cost data is competition sensitive and is proprietary. The data included within this study is modeled after the type of data companies have available to them. The application of the models and analysis of the results is not affected by using example data. The purpose of this case study is to identify the most significant means of developing a TOR. Further explanation of the data will be discussed in Chapter 2.

This paper is structured as follows: First in Chapter 2 the data and challenges in creating a data set are discussed. We then present the modeling frameworks in Chapter 3.

Chapter 2: Data and Parameters

In Chapter 1, it was mentioned the data used in this study were randomly generated. There are four variables, Pool(y), weight, material dollars (MTLDols), and months to first flight (M2FF). These are common aircraft industry variables.

- Pool(y): This variable represents the total labor cost in dollars where y can represent total program labor or total cost of a designated labor subcategory. Values range from $100,000-$20,000,000.

- Weight: The weight variable represents the entire total weight in pounds where weight follows the specified y from pool(y). Values range from 100lbs-52,000lbs.

- Material dollars (MTLDols): The MTLDols variable represents the material cost in dollars for the tangible hardware and physical material for an aircraft. Values range from $2,000,000-$4,000,000,000.

- Months to first flight (M2FF): The M2FF variable represents the time in months it takes an aircraft to start being built until it flies for the first time. Values range from 20 months – 105 months.

Pool(y) is the response variable. The cost of labor is dependent on the weight of the aircraft, material dollars, and the planned months to first flight. All three variables have a strong influence on pool(y). Weight and material dollars are strongly correlated. Typically, the heavier the aircraft the more material dollars are required. In general, the heavier the aircraft, the larger the aircraft, and larger aircrafts require more labor to construct, design, and test. In addition, the longer the months to first flight the more labor cost is incurred.

The first challenge the data set had was the number of aircraft programs. While this specific data set was developed for the purpose of the case study it represents a realistic number

of potential data points. Aircraft competitions do not occur often. There is not a steady market for new aircraft. Thus the data set available is small. There are different types of aircraft so key component in gathering data becomes what aircraft can be included in the data set and what aircraft should be removed. For example, a prototype aircraft may have been constructed once and not flown. In this example we included the month to first flight variable. This indicates we intend our new program aircraft to fly. Using an aircraft that never flew is not ideal when providing the customer a TOR. For this case study the sample size $N = 25$. This sample size was selected because it is a reasonable number of programs a company could have data on.

A second challenge was the time frame each data point was collected. Market conditions, processes, and technology are a few examples of parameters that have drastically changed with time. We discussed market conditions earlier. The customer wants more for less. Process and technology improvements have been developed to decrease cost and decrease months to first flight. When gathering a data set, the customer wants recent and relevant programs. It's up to the company providing the proposal to justify what is recent and relevant. For the purpose of this case study a parameter was aircraft built after 1999 or before 1999. In order to have a sample size of 25 all aircraft regardless of age, were included.

A third challenge was the type of aircraft available. Aircraft types include fighter, attack, unmanned, bomber, or carrier. Each aircraft has a different job and thus different design and complexity. For the purpose of this study we assumed a manned vehicle thus unmanned options were not included. This is because the cost and technological focus are very different. For the purposes of this study an unmanned vehicle was not a reasonable comparison to a manned vehicle.

There are other parameters that were investigated with the data set but were ultimately deemed noise factors and thus insignificant. These parameters included if the aircraft had fixed wings or if the aircraft was designed for long range missions. Both are considered technological complexities of the aircraft.

In summary, the data set is generated to reflect realistic aircraft data. The variables selected are industry accepted and the parameters reviewed to identify the sample size represent one method of industry data selection. For real world application understanding the aircraft data available, company goals and customer direction will have a large impact on sample size.

Chapter 3: Modeling Framework

3.1 Log- Linear Model

A Log-Linear model is fitted to the data. The log-linear is the transformation of random

variables, z with z ~ N(0,1). The probability density function (pdf) is $f_z(z,x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. The

data is mapped g: $\mathbb{R} \rightarrow \mathbb{R}$ given by g: t $\rightarrow e^t$. The result of this mapping is all negative values

will transform to be between zero and one. All positive values will be above one with a

maximum at one (Baio 2013). The cumulative density function (CDF) is $F_{g(z)} =$

$P''(g(Z) < t) = P'(z \le \ln(t)) = \Phi(\ln(t)), (-\infty, t]$. This is the composition of the previous

functions. Finally, $F_z(t) = \Phi(t)$. This model addresses a skewed distribution that results

from large incurred cost from a small number of programs (Baio 2013).

3.2 Cobb-Douglas

The Cobb-Douglas cost function was utilized to find levels of output for

combinations of input. The empirical relationship between variables suggested a

production function can be applied. A generalization of Cobb-Douglas is the Constant

Substitution of Elasticity (CES). The standard equation is $y = \gamma \exp(\lambda t) (\sigma x_1^{-p} + (1 -$

$\sigma) x_2^{-p})^{\wedge}(\frac{-v}{p})$ where n is the number of programs (Henningsen and Henningsen 2011).

Here, let x_1 represent labor cost in dollars of the subset category and x_2 represent the

weight of subset categories.

3.3 Linear Regression

A fit linear model was used to regress the data. The response variable was Pool(y) and the explanatory variables were weight, material dollars (MTLDol), and months to first flight (M2FF). In this approach multiple regression and analysis of covariance was examined to determine if there is no significant difference between Pool(y) on weight, MTLDol, and M2FF. Gelman and Hill 2009, states a fitted linear model is intuitive when predicting an outcome because it minimizes the predicted error.

Chapter 4: Model Specification

For this study the three modeling frameworks described were applied to the data sets. There is a two part question: Does one model have greater statistical significance over the other two? Does analysis at the subset category level or total program level have greater statistical significance? The goal of this analysis is to analyze each data subset and each program in aggregate. The same data sets will be processed using the Log-Normal Model, Cobb-Douglas, and a Bayesian Linear Regression model. Finally, we will identify if there is a method that is reasonable as a TOR.

There were two statistical computer programs utilized for this analysis, R and Co$tat. The Log-Normal Model utilized Co$tat in an effort to recreate what is available in industry. Co$tat is Cost Analysis Statistical Package of ACEIT that has been designed for the cost/price analyst. Hosted in Excel, CO$TAT has most features contained in major statistics programs plus many tailored functions for the cost estimator. R was used to simulate the Cobb-Douglas and Linear Regression models as seen in academia.

4.1   Case Study

The model specifications used in this study are described in this section. We have a set of twenty-five program data points. Each have a unique set of variables and parameters. Let program be equal to Program(i), we assign a set of m logical variables denoted by $m=(m_1,m_2,\ldots,m_{i)}$ where $m_i = 0, 1$. These are the parameters used to characterize each program. For example, m_1 is "Is program(1) prior to 1999", m_2 is "Is program(2) fixed wing". If yes to the first parameter, then $m_1 = 1$, if the response to the second parameter is no then $m_2 = 0$. With the parameter conditions answered we have identified the data set for this case study.

For the first model, log linear using Co$tat: We model the labor cost for each program using its weight and

$$Pool(y) = aw_i^{n+1} b_1^{m_1} b_2^{m_2} * \ldots * b_k^{m_k} \qquad (1.1)$$

This equation is distributed N(1,σ), where i denotes the ith historical program and k denotes the kth dummy variable with mean equal to one and the standard deviation is common to all data points meaning that they are all equal. The variables in the equation are as follows: a is a coefficient determined by regression, w_i is the weight coefficient determined by regression, and b=(b₁, b₂, …., bₙ) are stratification coefficients corresponding to the dummy variables m_i . The regression to determine a, w, and b_i coefficients will minimize the least squares regression error.

$$(log(N_1))^2 + (log(N_2))^2 + \cdots + (log(N_l))^2 = \sum_{i=1}^{l}(log(N_l))^2 \qquad (1.2)$$

Next we will investigate normalizing the first model. Normalize w_i and m_i to the new program, denoted by j. We define the normalization of the historical programs

$$Pool(y)_{i \to j} = aw_j^{n+1} * \left(\prod_{p=1}^{q} b_k^{m_j}\right) \qquad (1.3)$$

Adjusting for the new program we have the following

$$Pool(y)_{i \to j} = Pool(y)_i * \left(\frac{w_j}{w_i}\right)^{n+1} * \left(\prod_{p=1}^{q} b_k^{m_j - m_i}\right) \qquad (1.4)$$

In order to show that equation (1.3) and equation (1.4) are equivalent start with Pool(y).

$$Pool(y)_i = aw_i^{n+1} \left(\prod_{p=1}^{q} b_k^{m_i}\right) * N_i(1, \sigma)$$

$$N_i(1, \sigma) = \frac{Pool(y)_i}{w_i^{n+1}\left(\prod_{k=1}^{i} b_k^{mi}\right)}$$

$$Pool(y)_{i \to j} = aw_j^{n+1}\left(\prod_{p=1}^{q} b_k^{m_j}\right) * \frac{Pool(y)_i}{w_i^{n+1}\left(\prod_{p=1}^{q} b_k^{mi}\right)} = Pool(y)_i \left(\frac{w_j}{w_i}\right)^{n+1} \left(\prod_{p=1}^{q} b_k^{m_j - m_i}\right)$$

thus we have the original equation.

9

For the second model, the Cobb-Douglas function was applied to the data. Under this model we investigated economic modeling with a Constant Elasticity of Scale (CES) function. We derived the output quantity as Pool(y) divided by weight. The quantity of intermediate inputs was derived as Pool(y) plus MTLDols divided by weight. Using the derived cost estimating relationship of Pool(y) divided by weight we have the dependent variable. Pool(y) and weight are the independent variables. The equation is as follows:

$$f(Pool(y), weight) = (\tau_1 Pool(y)^p + \tau_2 weight^p)^{\frac{1}{p}} \qquad (1.5)$$

Now we consider the third model, fitted linear regression. The equation is as follows:

$$Pool(y) = \mu + \beta_1 * w_i + \beta_2 * MTLDols_i + \beta_3 * M2FF_i \qquad (1.6)$$

We applied this method to the total program and each of the eight subcategories.

4.2 Results

The focus of this study is to demonstrate the best method to support a TOR. Is the cost of a new program reasonable compared to historical programs? Secondly, will the results show significance at the total program level or subcategory levels?

The first method is to analyze the data using the log-normal model. This approach expanded the analyses by normalizing Boolean parameters and weight. The data was run through Co$tat to generate a log linear regression model. The log linear model was selected because the dependent and independent variables are related proportionately. Taking the log also addressed the non-linear relationship between the dependent and independent variables.

The Pool(y) variable in the model represents one of the following labor dollars: Structure and Subsystem Design, Flight Test, Mission Systems, Support and Training, Systems Engineering, Test and Validation, Manufacturing, Program Management, and Total Program. These are the variables we want to analyze. Once the Log Linear Regression Model was created

the data was adjusted to create a log model for each individual program. Each aircraft's historical program Pool(y), weight, material dollars, and M2FF are applied to the logarithmic scale. Using the log values applied in the regression equation normalized historical program values are computed.   The results are shown below in Table 4.1 and Table 4.2. Figure 4.1 shows the residuals in log space. The data points on the graph fall relatively in a straight line indicating the data is approximately normally distributed.

Category	DF	Sum of Square (SS)	Mean SQ = SS/DF	F-Stat	P-Value
Structure/Subsystem Design	12	40.8121	3.4010	7.1622	0.0047
Flight Test	10	13.6659	1.3666	2.1689	0.2844
Mission Systems	12	13.9492	1.1624	0.6305	0.7590
Support and Training	12	32.6759	2.7230	1.8019	0.1912
Systems Engineering	12	38.2576	3.1881	2.8019	0.0492
Test	12	26.3275	2.1940	1.9354	0.1420
Manufacturing	11	23.3857	2.1260	1.6141	0.5522
Program Management	12	42.6816	3.5568	6.8156	0.0086
Total Program	12	33.0973	2.7581	10.9992	0.0002

Table 4.1 Model 1 Analysis of Variance

| Category	Standard Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
Structure/Subsystem Design	0.6891	91.48%	78.71%	0.9565
Flight Test	0.7938	87.85%	47.35%	0.9373
Mission Systems	1.3578	65.42%	-38.33%	0.8088
Support and Training	1.2293	70.61%	31.42%	0.8403
Systems Engineering	1.0667	75.35%	48.46%	0.8680
Test	1.0647	67.86%	32.80%	0.8238
Manufacturing	1.1477	94.67%	36.02%	0.9730
Program Management	0.7224	92.12%	78.60%	0.9598
Total Program	0.5008	92.31%	83.92%	0.9608

Figure 4.2 Model 1 Goodness-of-Fit
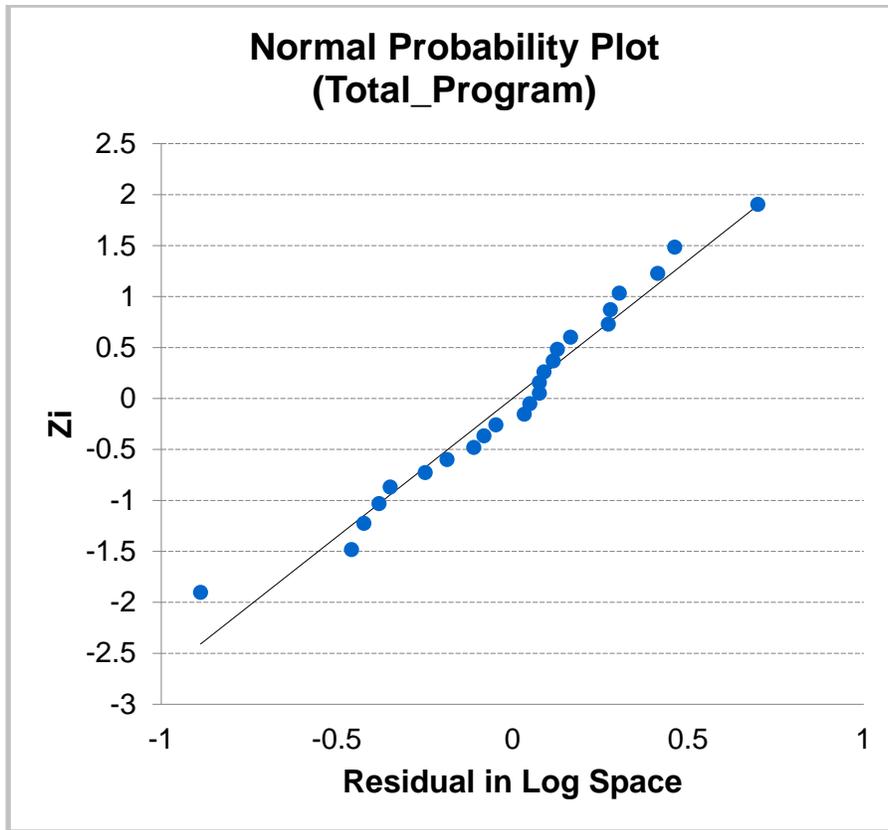
Figure 4.1 Model 1 Normal Probability Plot for Total Program

We assume $\alpha = 0.05$. This is an academically accepted value and addresses concerns over making a Type I error. The p-values for Flight Test, Mission Systems, Support and Training, Test and Manufacturing are greater than α, this indicates we should not reject the null for these categories. This indicates the aforementioned categories are significant. Further, analysis of the adjusted R-square to determine how well the model fits the data indicates that Structure/Subsystems Design, Program Management, and Total Program are good fits.

The Cobb-Douglas Cost function is one way to investigate the question of cost. This is a widely used function to represent the relationship between two or more inputs. It is also highly effective when evaluating technological relationships, such as the aircrafts used as an example of programs in this exercise. The Constant Elasticity of Substitution (CES) package in R was used to compute the relationship between the independent variables. This package allows for practical

analysis utilizing the Douglas-Cobb function. The first step was to compute the output value. In this case Pool(y) divided by weight. Next, use the "cesEST" function to compute the CES with Pool(y) and weight with fixed returns of scale. Then, calculate the fitted values. The summary table from the "cesTotalCWeight" variable which holds the CES function calculation shows statistical significance. The standard error (SE) of the CES and the t-value are shown in the table below, Table 4.3

| Category	Standard Error (SE)	P-value
Structure/Subsystem Design	0.3055	0.754
Flight Test	0.8273	<2e-16
Mission Systems	0.8669	0.0205
Support and Training	0.9370	0.596
Systems Engineering	1.1170	0.02
Test	0.7667	<2e-16
Manufacturing	0.6874	0.478
Program Management	0.2471	0.888
Total Program	0.8904	0.279

Table 4.3 Model 2 Elasticity of Substitution

We assume $\alpha = 0.05$. The unadjusted data shows we do not reject the null for Structure/Subsystem Design, support and Training, Manufacturing, Program Management and Total Program. This indicates the above categories are significant. The adjusted data shows we do not reject the null for Flight Test, Mission Systems, Support and Training, Systems Engineering, Test and Manufacturing. The adjusted values appear to be artificially inflated thus this method is not ideal for analyzing the data.

In this regression analysis we used the fitted linear model. Model 3 analyzes the model for each category as a whole and the significance between explanatory variables. This was achieved by using the LM function in R. Table 4.4 shows the results and Figure 4.2 shows the Normal Q-Q plot for each category.

| Category	Variable	P-value
Structure/Subsystem Design	Weight	0.0000
	MTLDol	0.0663
	M2FF	0.0000
Flight Test	Weight	0.02323
	MTLDol	0.8207
	M2FF	0.6395
Mission Systems	Weight	0.0284
	MTLDol	0.6228
	M2FF	0.9746
Support and Training	Weight	0.0000
	MTLDol	0.000
	M2FF	0.0021
Systems Engineering	Weight	0.0249
	MTLDol	0.3060
	M2FF	0.5046
Test	Weight	0.0120
	MTLDol	0.0000
	M2FF	0.0043
Manufacturing	Weight	0.0177
	MTLDol	0.5478
	M2FF	0.8334
Program Management	Weight	0.0000
	MTLDol	0.071
	M2FF	0.0000
Total Program	Weight	0.0000
	MTLDol	0.0019
	M2FF	0.0000

Table 4.4 Model 3 P-Values by Subcategory by Variable
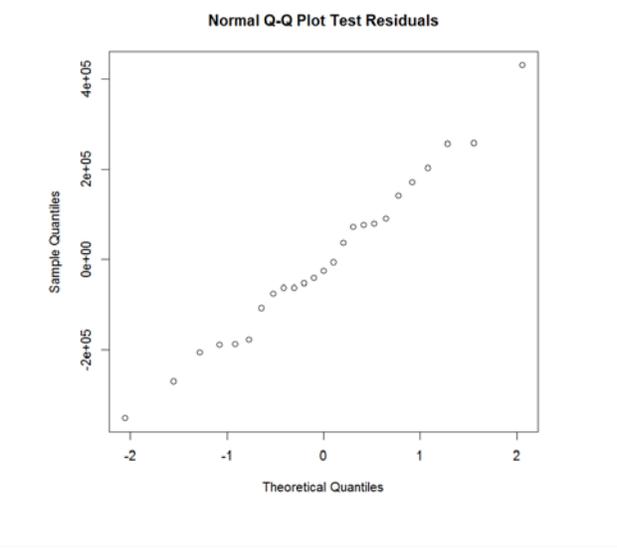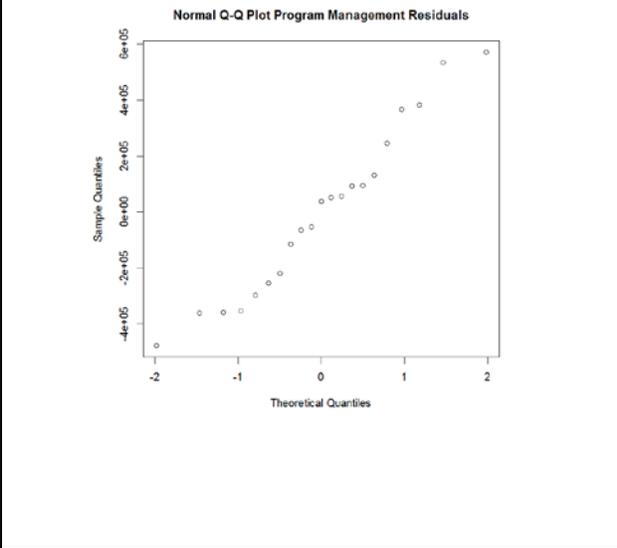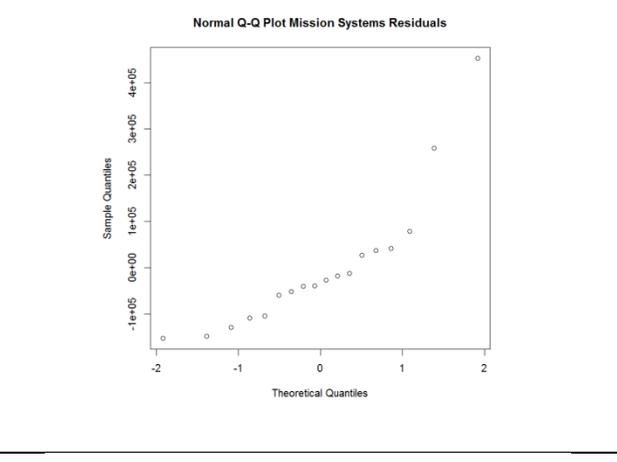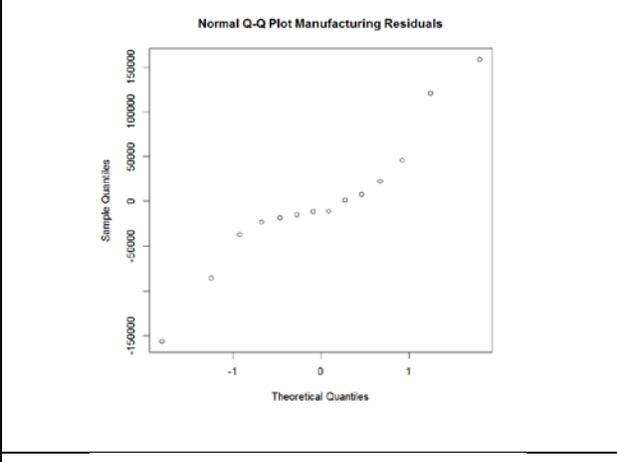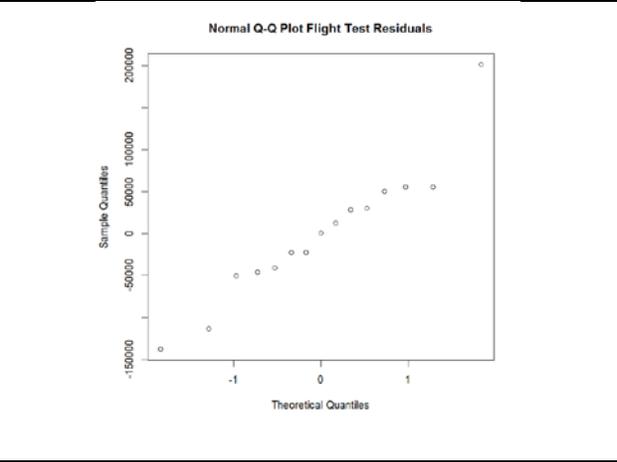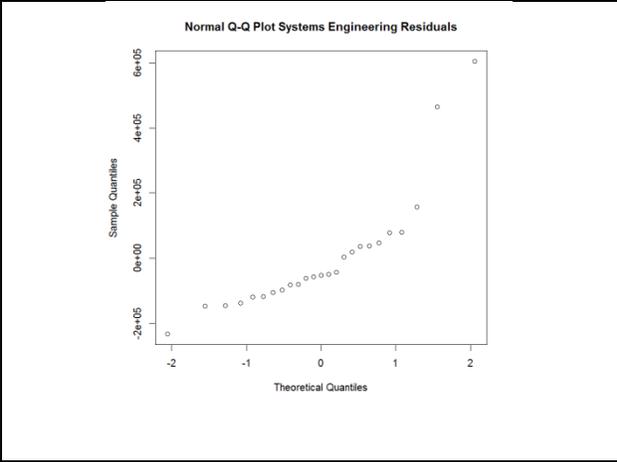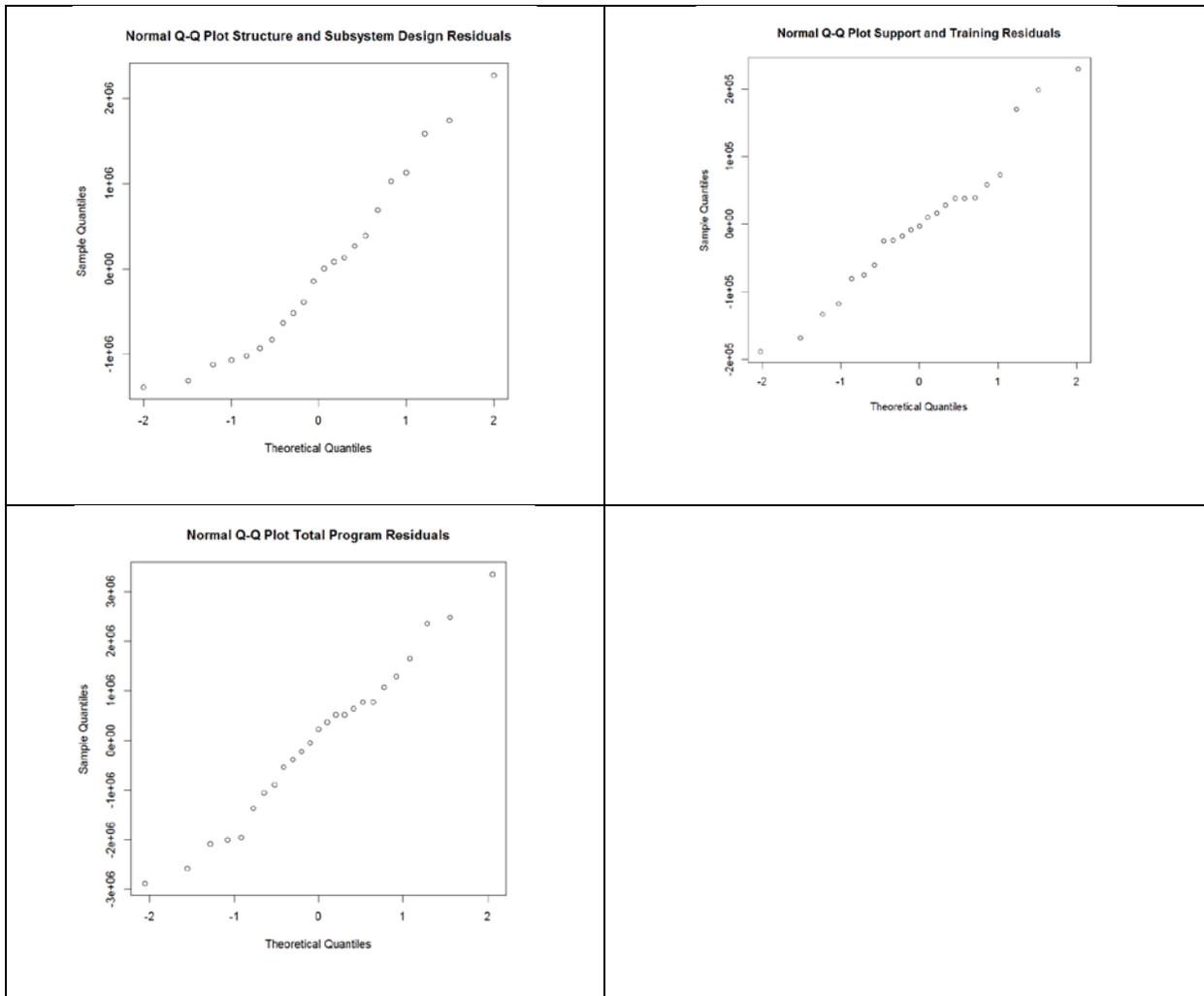
Table 4.2 Model 3 QQnorm Graph of Residuals

We assume α = 0.05. In the Structure/Subsystem Design and Program Management models MTLDol is significant thus we do not reject the null. For the Flight Test, Mission System, System Engineering, and Manufacturing models MTLDol and M2FF are significant. There is no statistical significance evident for the Support and Training, Test, and Total Program when applying this method. Reviewing the QQ-norm graphs indicate that generally each category is linear and thus the assumptions are valid.

Chapter 5: Discussion

There are several questions being examined. Is the price of a new program reasonable compared to historical programs? Does a given model provide a better analysis than another? Is it significant to analyze each program at a category level or total program level?

The cost of the new program is reasonable though variation from history causes concerns with comparative data. The three models are examined to determine if any one model has greater significance in analyzing the aerospace data. The first model, a log linear model utilize Co$tat. It shows that there is statistical significance at the total program level. However, there is variability at the category level. This variability could be a result of sample size. The second model, a Cobb-Douglas Model, showed the greatest significance at the category level. The P-values indicated a large majority of the categories should not reject the null, thus there is significance in this analysis. The final model is a fitted log linear. This model appears to provide the least significance. It demonstrated MTLDols appears most frequently as significant. This is expected since Pool(y) is the total price and MTLDols is strongly correlated. It also, shows that in some cases Months to First Flight (M2FF) was significant. Due to external influences and understandability by the audience the Log Linear Model is recommended.

Based on the Log Linear Model the new program is reasonably priced and aggressively priced compared to history. This supports the industry need to lower cost. A log linear model is the most advantageous for analytical purposes. There is significance in looking at cost by category. There is meaning in performing a TOR at the total program level and subcategory level. A TOR may not be required for every category thus performing a TOR on the strongest statistically significant categories is advisable.

<u>Works Cited</u>

[1] Corro, Dan and Shim, Kyumin, *Monitoring Cost Changes with Log-Linear Cost Models: Lessons from a Case Study*, National Council on Compensation Insurance, Inc. 2000

[2] Baio, Gianluca. *Bayesian Models for Cost-Effectiveness Analysis in the Presence of Structural Zero Cost*, 2013, Statistics in Medicine Volume 33 Issue 11

[3] Gelman, Andrew and Hill, Jennifer, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, 2007

[4] Hogg, Robert V., McKean, Joseph W., and Craig, Allen T., *Introduction to Mathematical Statics*, Pearson Education Limited, 2013

[5] Raluca Vernic, Sandra Teodorescu and Elena Pelican, *TWO LOGNORMAL MODELS FOR REAL DATA*, An. Șt. Univ. Ovidius Constanța, Vol. 17(3) 2009

[6] Henningsen, Arne and Henningsen, Geraldine. *Econometric Estimation of the "Constant Elasticity of Substitution: Function in R: Package micEconCES"*, FOI Working Paper 2011/9, Institute of Food and Resource Economics, University of Copenhagen. 2011

[7] Rumbaugh, Margaret G, *Understanding Government Contract Source Selection*, Management Concepts. Inc. 2010