A Comparative Study of Target Costing Methods

Timothy Muia

A Thesis

in

The Department

of

Mechanical and Industrial Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Masters of Applied Science (Industrial Engineering) at

Concordia, University

Montreal, Quebec, Canada

December 2012

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Timothy Muia
Entitled: A Comparative Study of Target Costing Methods

and submitted in partial fulfillment of the requirements for the degree of

Degree of Masters of Applied Science (Industrial Engineering)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

______________________________ Chair
Dr. L. Kadem

______________________________ Examiner
Dr. G. Gouw

______________________________ Examiner
Dr. I. Dostaler

______________________________ Thesis Supervisor
Dr. N. Bhuiyan

Approved by ____
Graduate Program director : Dr. S. Narayanswamy

_______ 2012

____
Dr. Robin A.L. Drew, Dean
Faculty of Engineering & Computer Science

ABSTRACT

A Comparative Study of Target Costing Methods

Timothy Muia

The aerospace industry is highly competitive and is constantly seeking to develop new innovative aircraft programs in order to remain competitive and to gain new market segments. Due to the present economic situation, the necessity of cost initiatives is essential to comprehend, especially at an early conceptual stage. Cost understanding has become a crucial role in the supply chain to accurately predict target costs for all major structures and commodities in order to launch a new aircraft program within mandate and assist in commercial negotiations. To continuously improve turnaround time, costing models can be developed to improve the design to cost effort at the early stages of a new program. In this thesis, parametric models for estimating the cost of a component will be developed and compared. The study uses a comparative analysis between linear and non linear parametric models in order to determine which estimation method enhances the credibility of the cost estimate. The reliability of the costing models is evaluated by two different methodologies to determine which parameter(s) become significant and will therefore be used to determine the cost. These methodologies are the analysis of variance and path analysis. It is concluded that the non linear regression analysis achieves a lower level of error when comparing it to the linear regression. Moreover, possible future studies to develop topics pertaining to target costing are presented in this thesis.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Dr. Nadia Bhuiyan who has helped and guided me throughout my thesis. Starting from when she was my INDU 211 professor 9 years ago, Dr. Nadia has encouraged me along with hundreds of other students to exceed expectation while being a leader in the Industrial Engineering field. I am truly privileged to have been under your direction.

I wish to thank my parents, Christine and Dominic, for their unceasing support and encouragement in concluding my thesis.

I wish to thank my beautiful fiancée Sophia who has motivated and supported me in the accomplishment of my masters. She is truly an inspiration.

I wish to thank my sister, Kimberly, for being incredible role model in her career and education. Kim guided me in making the right decisions as an Industrial Engineer and in my personal life.

Lastly, I would like to thank Bombardier Aerospace for allowing me to achieve my thesis in the strategic costing team and for supporting continuous education.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

$\Delta C'$	Cost differential between trend-line and baseline
$\Delta C_{Complexity}$	Cost differential due to cumulative complexity factor
ANN	Artificial neural networks
ANOVA	Analysis of variance
BA	Bombardier Aerospace
β_0	Intercept
β_1	Regression coefficients or slope
B_i	Regression coefficient
β_k	Regression coefficients
β_m	Regression coefficients
CE	Concurrent engineering
CER	Cost estimating relationship
CERs	Cost estimating relationships
C_{Pred}	Predicted manufacturing cost
Con	Concurrency
CV	Coefficient of variation
DA	Dimensional analysis
DC	Degree of change
DE	Difficulty to expertise
DE	Experience of departmental personnel
D_{fan}	Fan diameter
D_m,	Effort driver (m)
DP	Design parameters
e	Residual
$\hat{E}$	Estimated design effort in hours
ε	Error
ε_i	Residual value of observation i
$\bar{\varepsilon}$	Mean of residuals
GE	General Electric
ISPA	International Society of Parametric Analysts
L	Volume
lbs	Pounds
ln	Natural log
LSE	Least squares estimation
LRM	Linear regression model
m	Number of fundamental dimensions in a relationship
M	Dimension
M_i	Profit margin of the ith stage
m_{data}	Original data
MDO	Multidisciplinary design optimization
MLG	Main landing gear
MLRM	Multiple linear regression model
MTOW	Maximum takeoff weight
n	Number of dimensions in a relationship

n	Number of stages in the distribution channel
N	Number of data points
NVT	Number variance technique
OLS	Ordinary least squares
P	Price paid by the end user
PA	Path analysis
PC	Product complexity
P_{0i}	Path coefficient
P_{ou}	Path coefficient for the uncorrelated residual
r	Coefficient of correlation
R^2	Coefficient of determination
SEE	Standard error of estimate
SLR	Simple linear regression
SLRM	Simple linear regression model
SS_R	Residual sum of squares
SS_T	Total sum of squares
T	Time
TC	Target cost
TD	Type of design
μ	Mean of data set
VIF	Variance inflation factor
x	Observed data point
X_k	Specified cost driver
X_m	Specified cost driver
X1	Weight of the MLG
X2	MTOW of the program
X3	Height of the MLG
y	Observed value
$\hat{y}$	Predicted value
Y	Actual result
Y_i	Masked cost
$\hat{Y}$	Target cost
$\hat{Y}i$	Predicted cost
z_{data}	Vertical shift from the original data
Z_{ε_i}	Standardized residual of observation i
σ^2	Variance
σ_ε	Standard deviation of residuals
σ_i	Standard deviation of the CER
σ_y	Standard deviation of the MLG cost

1.0 INTRODUCTION

Successful organizations depend on the ability to continuously develop new products while meeting customer demand for improved cost, delivery, quality and flexibility. Many industries such as the aerospace industry, are subjected to extremely competitive markets in which companies require accurate business cases and strategies for their products. Being in a competitive environment, cost has increasingly become one of the main parameters for clients. In response to improved cost, many manufacturers have begun to adopt tools and techniques: one such technique is target costing (TC). Target costing is used to understand the actual cost of a system, which will in turn find opportunities in reducing or improving the cost. The objective of this thesis is to accurately determine a fundamental question known to businesses: how much will the product cost?

1.1 Background

Target costing draws important links to concurrent engineering (CE), which aims at optimizing engineering design cycles. Bhuiyan et al. (2004) state that CE reduces the overall lead time to design components. As can be seen in the below picture, the CE process normally entails an iterative approach of four main activities in product development: product idea, evaluation, analysis and synthesis. Comprehending the target cost of the product idea phase will eliminate multiple time-consuming revisions of the process, hence optimizing engineering design cycles. CE refers to an integrated product development team consisting of engineering, finance, supply chain, marketing etc. in order to minimize the effort required downstream in the product development cycle. This team of experts will use a systematic approach described in the following figure to create a lean process in a products development in order to satisfy the customer's needs.

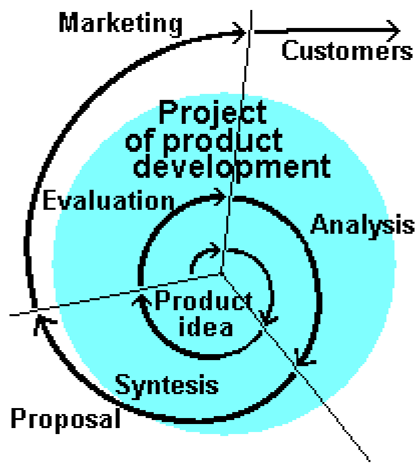

Figure 1: Concurrent Engineering Product Development Cycle (Routio, 2007)

Target cost models are utilized as critical decision tools to approximate the cost of products. According to Blanchard & Fabrycky (1998), ten to fifteen percent of the total cost spent during the design phase commits eighty percent of the total cost in the life cycle. Moreover, Davila (2000) argues that 70-80% of a product's cost is set during product development and cannot be changed when the product reaches production.

Hence, cost models used in early design phases are extremely important and the level of accuracy must be significantly high in order to reduce cost early on, when effective cost models will be most beneficial due to the lack of product knowledge. In other words, the technical specifications for the products are unknown or unclear most of the time. These technical specifications can be referred to as cost estimation relationships (CER). CERs correspond to a positively correlated relationship between a dependant variable and the corresponding independent variable. For example, in the aerospace industry, CERs are based on full data sets consisting of all available costs and technical data associated with a particular product (Book *et al.*, 2011). Book *et al.* (2011), developed an extension to the traditional CER named the adaptive CER. The adaptive CERs goal is to have smaller estimating errors and narrower prediction bounds. CER's are used as fundamental knowledge required in building target cost models and will be explained in more detail in Section 3.4.

Increasing competition in global markets drives companies to deliver high quality goods at lower price. Monden and Hamada (1991) discuss the necessity of target costing in new product development of an industrial assembly based manufacturers. Traditionally, businesses used to calculate the cost of a product, add a profit margin and then sell the product to the public where the firm can meet an acceptable rate of return (Sudhir, 2009). Moreover, traditional cost accounting was developed from mass production where profitability is maximized when labour and machine utilisation is maximized. Traditional costing practices are no longer effective since the competitive environment, where multiple suppliers can provide the same product, makes cost reduction initiatives, optimization and continuous improvement a necessity to the organization. In the end, a higher selling price for a similar product does not adhere to the end customer (Cooper and Chew, 1996). Traditional accounting structures do not support organizations and rarely enable effective decision making by managers (Maskell, 2006). Traditional accounting systems are no longer effective due to the following inefficiencies in the reporting (Maskell and Baggaley, 2006):

1. Large, complex and wasteful processes
2. Measurements and reports that motivate large batch production
3. No concept to measure lean improvements

4. Use of standard product costs for decision criteria

Organizations are moving towards target costing which is a component of lean management. Target costing, a subset of lean accounting, seeks to replace traditional costing practices by providing more timely and relevant management information. This method creates value to the end customer by establishing the right market price and then works backwards to design and manufacture a product in a lean fashion.

Ansari and Bell (1997) explained in their work a very simplistic way in which target costing can be derived. One must commence by assessing the selling price of a product determined by its market forces. Thereafter, the profit margin of a specific product must be established. From the above statements, the technique known as target costing can be formally stated as follows:

Target price - Target profit = Target Cost

1.2 Thesis Objectives and Scope

In the target costing approach, the cost of a new product is no longer an outcome from the product design process; the cost plays an important role and becomes an input into the process (Ansari and Bell, 1996). It is understood that target costing is suited to meet the needs of the customer in today's competitive environment (Sani and Allahverdizadeh, 2012). Products and their corresponding prices are determined by the customer's needs. In order to satisfy the customer needs in terms of acceptable pricing, target costing models need to come into account in the early stages of product development. The objective of this thesis is to present the concept of target costing and how it can be utilized in the early stages of product development. The methodology and models discussed can be applied to a wide variety of commodities and industries. The target costing models are then applied to a case study in the aerospace industry.

The outline of the thesis is as follows. Chapter II presents a literature review of the topic under study. Chapter III introduces the parametric models used to estimate the cost, its suitability along with its linearity assumptions. Moreover, this section depicts the methodologies used to determine which type of regression is the most suitable. Subsequently, Chapter IV describes a case study where the models were applied. Furthermore, given the fact that the data used in this thesis is confidential and cannot be disclosed, data masking was necessary; thus multiple data masking techniques are presented. Finally, the conclusions, future research and recommendations are identified in Chapter V.

2.0 LITERATURE REVIEW

In this chapter, four main aspects will be discussed. The chapter includes the origins of target costing, the importance of costing accuracy, cost models and introduction to parametric estimating principles with relevant examples.

According to the car manufacturer Toyota, TC is a process for ensuring that a product launched with specified functionality, quality and sales price can be produced at a life-cycle cost that generates a satisfactory level of profitability (Cooper and Slagmulder, 1997). In addition, TC can be perceived as a market-driven strategy that can help a company assess the ideal value of a product. This is essential in identifying a baseline price for the purpose of negotiation with suppliers. The TC approach concentrates on determining costs for a product during the planning and design stage. This approach makes use of cross-functional teams within an organization made up of cost experts, supply chain agents and engineering. The cost expert's role in TC is to determine the ideal value of a product based on cost drivers.

2.1 Target costing

The research of Leahy (1998) explains that Japanese companies tried to maximize desirable product attributes while at the same time minimizing product costs. The above process became known as "value engineering" and was rapidly adopted in Japan. Value engineering was then combined with the idea of reducing product costs as early as possible during the planning and development stages of a product (Buggert and Wielpütz, 1995). From that point on, target costing originated and was quickly implemented across

Japan in the 1960s. Many years later, upon realizing its successful results, extensive efforts were made to convey target costing to Western companies.

According to Rösler (1996), Toyota used the expression “genka kikaku” in their daily vocabulary, which translates into “target costing”. Figure 2 describes the derivation of each individual term from the Japanese language.

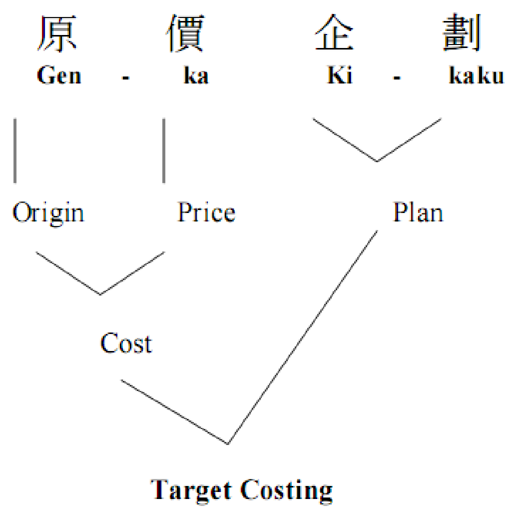

Figure 2: Translation of expression "genka kikaku" (Rösler, 1996)

In the 21st century, the aerospace industries globally are facing highly competitive markets. Manufacturing companies observe that about 80% of a product’s total cost is determined during the planning and development process (Ansari and Bell, 1997). This above statement can also be shown in the work of Blanchard and Fabrycky (1998). Target costing is an attempt at the planning and development phase, of a product life cycle, to attain a specified cost that is decided on by management. Such importance is placed in the early design phase of an emerging product because any alterations will result in high costs once it had been launched to the market (Cooper and Chew, 1996). It was also demonstrated that assessing the cost and building a strategic alliance between a company and its suppliers is extremely important to sustain competitiveness (Kaplan and Atkinson, 1998).

According to Ulrich and Eppinger (2011), target costing is the manufacturing cost at which the company and its distribution partners can make adequate profits while still offering the product to the end customer at a competitive price. It is a process of setting the manufacturing cost based on the price the company hopes the end user will pay for the product, thus it is the reverse of the cost-plus approach of pricing. Similarly, according to Sani and Allahverdizadeh (2012), target costing is a method that takes financial, manufacturing and customer needs into consideration during the early conceptual phase and helps firms in making product design decisions to increase the value of the company. On the other hand, the cost-plus approach begins with what the firm expects its manufacturing costs to be and then sets its prices by adding its expected profit margin to the cost. The cost plus approach does not take into account that the prices are driven by market demand and no lean initiatives are implemented in the product development phase in order to reduce cost. The target cost formula is given by the following expression:

$$C = P \prod_{i=1}^{n} (1 - M_i) \tag{1}$$

where,

C, Target cost

P, Price paid by the end user

n, Number of stages in the distribution channel

M_i, Profit margin of the ith stage

As presented in the work of Cooper and Chew (1996), traditional costing techniques are not profitable in a competitive environment and it would be necessary to start investing in new management accounting tools, such as target costing.

Tanaka (1993) utilized the concept of target costing in his work and relates it to cost reduction. His research allows companies to sets goals for cost reduction at the design stage and then tries to achieve these new targets through design changes that will accomplish the cost reduction goals. Toyota performs target costing by taking the original product and summing up all the modifications with extreme detail. This approach is known as a detailed bottom up type of estimate to assess the target price of the new product. Bottom up estimation entails the summation of all the specific tasks to complete the component in terms of manufacturing, labour, wrap rates, overhead, etc. Afterwards, the product must go through a vigorous testing phase to judge the costs of the new design in comparison with the old one, in order to guarantee a cost reduction after implementation of the new product. The main idea that Takana uses to achieve companywide goals can be seen as a lean enterprise strategy. As can be demonstrated in today's market, many companies have greatly distinguished themselves from its competition and this is in part due to the implementation of target costing. These companies are:

1. Rockwell Collins and Goodyear in the aerospace industry
2. Mercedes, Toyota and Nissan in the automotive industry
3. Texas Instrument and Canon in the electronics industry

A synonymous costing practice to target costing with a Japanese heritage is named kaizen costing (Sani and Allahverdizadeh, 2012). Kaizen refers to the process of seeking continuous improvement where some Japanese companies must link there planning process with a kaizen process once the products are in production. Kaizen costing indicates that regular kaizen events are scheduled throughout the product life cycle in order to continuously reduce its cost. It makes all employees conscious that the company must continually reconsiders how the task is undertaken and whether there is a better way of doing it.

Incorporating kaizen and target costing practices within an organization can be beneficial in many ways, such as:

1. Improves customer satisfaction (Cooper and Chew, 1996)
2. Increases knowledge of supply chain cost structure (Swenson *et al*, 2003)
3. Emphasis on cost reduction in the early stages of product development (Cooper and Chew, 1996)
4. Considers the whole product life cycle (Swenson *et al*, 2003)

Hart *et al*. (2012) made a comparative study between multiple costing techniques in order to improve the fidelity of the cost estimate for an engineering system. They compare three types of costing practices: regression analysis method, neural networks analysis method and Kriging regression method. The Kriging model is based on treating the data as a stochastic process. A seven-step approach is summarized in their research to perform a Kriging regression model. The main objective of their work is to create a cost estimation model with a high level of predictive capability and fidelity. A case study addressing the fabrication of a submarine pressure hull is developed in order to compare the three types of analysis. Their research concludes that the Kriging approach demonstrates better cost predictive capabilities compared to the neural network and regression analysis. The Kriging method is limited, as all other cost modeling methods, to the available historical cost data. Thus, Kriging will not be useful in situations where completely new technologies or designs are considered.

Dal-Ri *et al.* (2005) propose a methodology based on the fuzzy logic concepts to take into consideration the uncertainty and subjectivity in the target costing process. Fuzzy logic assumes a degree of pertinence within the 0 to 1 range, which deals with approximate reasoning rather than exact numbers. Dal-Ri *et al.* (2005) use the production of a tennis racquet as example in order to illustrate these fuzzy concepts. There research shows evidences that fuzzy logic enables a decision-maker to gain additional insights in the relationship between costs components and products. Fuzzy logic can be considered a helpful tool to handle the subjectivity and the uncertainty inherent in the complex process of organization's decision making.

Roy et al. (2001) describes the development of a target cost model for predicting the cost of engineering design effort during the conceptual stages of product development. With the use of linear regression and factory analysis to identify the final CERs, the authors demonstrate that using regression analysis can predict future design effort required, based on limited product definition at the conceptual design stage.

Womack and Jones (2003) discussed five lean principles to eliminate waste. The final element of value in Lean Thinking is related to value definition: target costing. According to Womack and Jones (2003), target costing is the most important task in specifying value, once the product is defined. A target cost should be determined based on the amount of resources and effort required to make a product if all the waste were removed from the process. In contrast, many companies set target selling prices based on what they believe the market will bear. They then work backwards, to determine acceptable costs to ensure an adequate profit margin for any new product.

Sani and Allahverdizadeh (2012) recommend as guide a six step process in order to properly implement target costing in any organization. The six step process is the following:

1. Establish the target market price
2. Establish the target profit margin and cost to achieve
3. Calculate the probable cost of current and new products and processes
4. Establish the target cost
5. Attain the target cost
6. Pursue cost reductions once production has started

In the end, target costing is a cost management tool that planners use during product design to drive improvement efforts aimed at reducing the product's future manufacturing costs (Kaplan and Atkinson, 1998). The six step process related to target costing is suited to meet the needs of all organizations in today's competitive environment. Target costing needs to be incorporated early in the developmental phase of product development where members from operations, marketing, accounting and procurement departments conduct a concurrent engineering approach rather than sequential process (Zengin and Ada, 2009). This will help optimize decisions based on functionality, quality and market price with a view on design for manufacturing and assembly.

On the other hand, a simplistic method to estimate the cost of a product with no research and little knowledge about the subject can be described by this quote:

> When I build something for somebody, I always add $50 million or $60 million onto the price. My guys come in, they say it's going to cost $75 million. I say it's going to cost $125 million, and I build it for $100 million. Basically, I did a lousy job. But they think I did a great job. (Trump 2010)

The above rough order of magnitude type of estimate is not the right solution which leads us to develop an accurate cost estimation model.

2.2 Accuracy

Accuracy is crucial in assessing the target price for a complex product. Overestimating the target price in the conceptual phase may cause the product to be oriented towards a less expensive design with worse performance. Underestimating the target price in the early design phases for a product along with its material, labor, tooling and engineering effort will definitely lead to a product redesign which will consume additional resources at a large cost in a later phase. This will ultimately increase the originally anticipated budgeted cost of a new product. The accuracy of cost estimates is essential for cost experts to control.

Accuracy of an estimate improves with the data availability. Generally, with the evolution of product design, more and more information can be provided to make estimation more accurate. Creese and Moore (1990) provided a discussion on the degree of accuracy in the different design stages as shown in Figure 3.

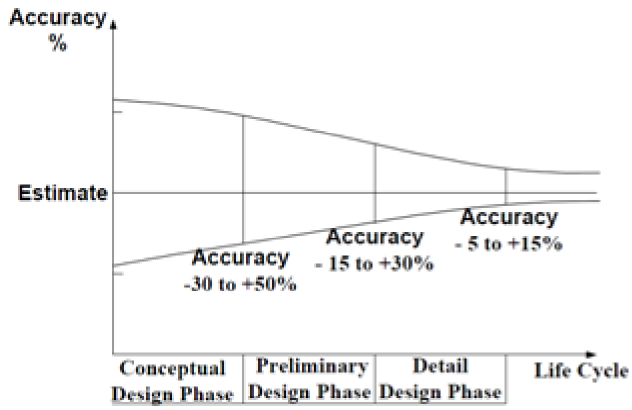

Figure 3: The degree of cost estimation accuracy (Creese and Moore, 1990)

During the conceptual design phase, information is scarce and cost estimation must rely primarily on the use of preliminary cost drivers. The accuracy ranges from -30% to +50% of the real cost. As the design evolves and potential trade studies are eliminated, the available data pertaining to the product becomes accessible and cost estimates can be made based on historical cost data. Frequently, regression type of estimation can be produced with the use of historical data. Further elaboration on regression analysis will be mentioned at a later stage of the thesis. The accuracy of cost estimates in the preliminary design phase ranges from -15 to +30%. During the detail design phase, the information about the product is known and the degree of cost estimating accuracy should be within - 5 to +15%. As previously mentioned, the preliminary design phase is the critical phase where companies can benefit from accurate cost models based on the available information on hand.

2.3 Cost Modeling

Understanding cost in the early stages of product development is essential to comprehend. Having the possibility to rely on a cost model to forecast the price of a product, within a certain acceptable degree of error, is beneficial to any enterprise. These cost models would alleviate many stresses incurred by management when faced with important cost related decision at such an early stage of product development. The accuracy of the cost models will be essential in the credibility of an estimate. Cost modeling is vital for industries due to the following reasons:

1. Deals with uncertainty of new products
2. Copes with product requirements (specification, design)
3. Assures better decision making (accurate lifecycle costing)
4. Assures better understanding of what drives the cost of a product
5. Ability to manage and predict cost throughout the design process at every level of design

Curran *et al* (2006) developed a cost model for manufacturing at an early design phase for the nose cowl section of a wide range of nacelles. Their research entails defining prominent manufacturing cost elements in order to associate the cost estimating relationships into cost models. The cost modeling approach is readily adapted to deal with the early conceptual design phases. The cost model is composed of different cost elements as a cumulative sum that can be estimated using higher level complexity ratings.

$$C_{Pred} = m_{data} D_{fan} + z_{data} - \Delta C' + \Delta C_{Complexity} \quad (1)$$

where

C_{Pred}, Predicted manufacturing cost

m_{data}, Original data

D_{fan}, Fan diameter

z_{data}, Vertical shift from the original data

$\Delta C'$, Cost differential between trend-line and baseline

$\Delta C_{Complexity}$, Cost differential due to cumulative complexity factor

In the end, the early cost model rendered an average error of less than ten percent when correlated with actual findings based on eleven different nose cowls.

2.4 Parametric estimation

Parametric estimating is a method used for predicting cost based on historical relationships between cost and one or more predictor variables (Roy *et al.* 2000). Parametric models can be classified as simple or complex. For this research, simple models are cost estimating relationships (CERs) consisting of one cost driver. Complex models, on the other hand, are models consisting of multiple CERs. According to the International Society of Parametric Analysts (ISPA), it is estimated that one can save from 40 to 80 percent of an employee's time by performing parametric estimation over a bottoms-up approach. Parametric tools and techniques have much more versatility than traditional estimating techniques. There are numerous benefits of using parametric estimation, such as:

1. Accurate estimates in conceptual design
2. Efficiency in model creation
3. Statistical link between the technical and cost proposals

According to Fabrycky and Blanchard (1991), parametric cost estimation is used in the early phases of conceptual design. Thereafter, once more understanding is available on the product under study, different costing techniques can be utilized to yield a higher degree of accuracy, such as: analogous and detailed estimates. The below figure describes the relation between the model types and their use in the product development cycle.

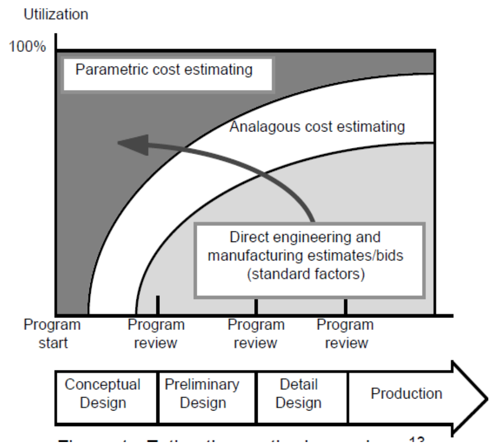

Figure 4: Estimating methods vs. product development phases (Fabrycky and Blanchard, 1991)

Marx *et al.* (1995) utilized parametric estimation to identify and quantify the economic benefits associated with the wing cost of a high speed civil aircraft concept. The goal of their research is to provide an estimate of manufacturing technology earlier on in the design process.

Bashir and Thomson (2004) developed a parametric model to estimate the time required to design hydroelectric generators for General Electric (GE). They analyzed fifteen designs made between the years 1985-1989. They performed parametric estimation where specific cost drivers are as follows;

1. Product complexity (PC)
2. Difficulty to expertise (DE)
3. Type of drawing to be submitted to customer (TD)
4. Involvement of design parameters (DP)

The resulting equation with the aforementioned specific cost drivers is,

$$\hat{E} = 0.12PC^{2}DE^{0.41}TD^{0.35}DP^{0.69} \quad (2)$$

Having applied their equation to 15 jobs, they found that their model was a better estimation tool for the company reducing the mean magnitude relative error of the prediction from 27% to 13%.

Kahyani and Basiri (2011) explore parametric estimation in order to determine the relationship between tree cross section cover and basal area in forest ecology studies. The simple linear parametric equation utilized can be validated with the work of Helsel and Hirsch (2002). The case study will determine if there is a relationship between the basal area and the cross section cover area in the forestry industry. Three groups were studied and the results are shown below.

Table 1: A summary of statistics derived from models obtained, Kahyani and Basiri (2011)

Models	R^2(%)	R_P^2(%)	*MSE*	*PRESS*	*t* test (value*P*)	*F* test (value*P*)
Group I	62	54.2	4.5	149.2	0.99	0.27
Group II	74	70.7	5.5	189.1	0.99	0.31
Group III	60	54.6	2.2	67.8	0.99	0.13

Multiple criterion were used to determine the validity of the simple linear regression model, they are: regression model, they are: the coefficient of determination, the mean sum of squares error, the t test error, the t test and the F test. All criterion previously mentioned play a role in determining the best determining the best test group. In order to select the best parametric equation between the the independent variable, basal area, and the dependant variable, tree cross section cover, a a summary of the rankings are presented in

Table 2.

Table 2: Ranked values of estimation error statistics, Kahyani and Basiri (2011)

Models	*MAE*	*SSE*	*RMSE*	*MAPE*	Ranks Mean	Ranks Mean
Group I	1	1	1	2	5	1.25
Group II	2	3	3	1	9	2.25
Group III	3	2	2	3	10	2.5

Kahyani and Basiri (2011) ranked the parametric models as the highest percent error was given the lowest ranking. As such, group III is deemed the best model since the model has the highest rank at 2.25.

Results of correlation in this study indicate that there was significant linear relationship between the two variables. The final model is show below.

$$y = 8.4 + 0.03x \tag{3}$$

where

y, Basal area

x, Cross section of the tree canopies

Salehi and Hematfar (2011) compared the linear and non-linear relationships between accounting variables and dividends listed on the stock exchange from 2005 to 2010 using ten different models. The parametric equations under study are described in the following table:

Table 3: Linear and non linear parametric models, Salehi and Hematfar (2011)

Linear	$Y = \alpha + \beta x$
Semi logarithmic	$Y = \alpha + \beta Lnx$
Inverse	$Y = \alpha + \beta / x$
Quadratic	$Y = \alpha + \beta_1 x + \beta_2 x^2$
Cubic	$Y = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$
Compound	$Y = \alpha \beta^x$
Power	$Y = \alpha x^{\beta}$
S	$Y = \exp(\alpha + \beta / x)$
Growth	$Y = \exp(\alpha + \beta x)$
Exponential	$Y = \alpha e^{\beta x}$

The different models can be represented by the graph plotted against all ten regression models, where the vertical axis represents the stock returns and the horizontal axis represents the net income.

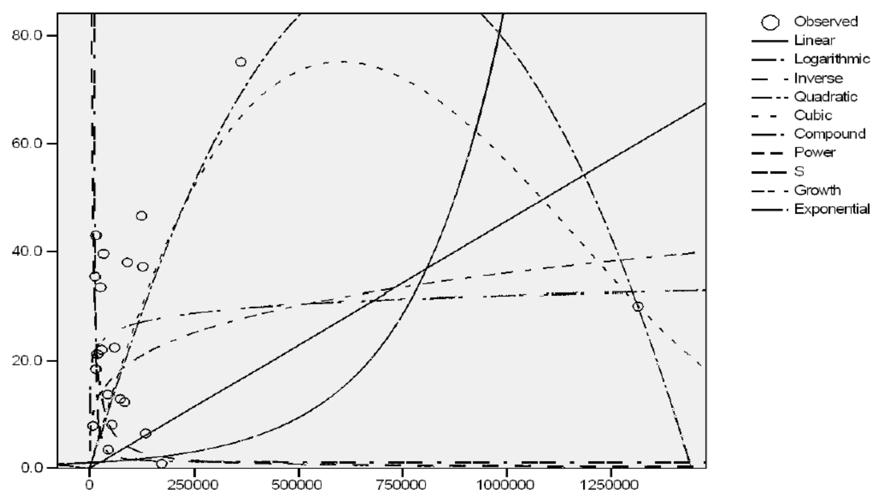

Figure 5: Regression graph between net income and stock returns, Salehi and Hematfar (2011)

It can be demonstrated that the non linear relationship between accounting variables and stock returns is stronger than the linear relationship. In some cases, 95% of the variables were able to be explained by the stock returns in a non linear model.

Salam (2009) presented a parametric model for estimating design effort of a compressor fan involving the engineering departments at Pratt & Whitney Canada. The non linear regression model utilized can be validated with the work of Walston and Felix (1977) and is shown below.

$$\hat{E} = aPC^{b} D_1^{c_1} D_2^{c_2} \cdots D_m^{c_n} \tag{4}$$

where

$\hat{E}$, Estimated design effort in hours

PC, Product complexity

D_m, Effort driver (factor m)

a, b, c_n, Constants (weights) that are estimated from historical data

Four potential parameters were identified for parametric modeling in his research. These factors are:

1. Type of design (TD)
2. Degree of change (DC)
3. Concurrency (Con)
4. Experience of departmental personnel (DE)

Upon validating with actual data, it was found that the proposed parametric model provides a maximum relative error of less than 10%. The jackknife technique was used to calculate the values of the constants. The parametric equations generated from this analysis can be summarized below.

Table 4: Summary of findings of all departments

Department	Design effort $(\hat{E})$	Maximum relative error (%)
D1	$4.078\text{x}10^4\ TD^{-3.328}\ DC^{2.108}\ Con^{-0.6107}\ DE^{-4.762}$	2.32
$D1_{\text{No Con}}$	$5.026\text{x}10^4\ TD^{-3.209}\ DC^{2.081}\ DE^{-4.792}$	4.68

D2	$2.63 \times 10^2\ TD^{1.75}\ DC^{0.221}\ Con^{-0.131}\ DE^{-0.345}$	2.99
D2 $_{\text{No Con}}$	$2.69 \times 10^2\ TD^{1.76}\ DC^{0.225}\ DE^{-0.325}$	3.30
D3	$6.09 \times 10^2\ TD^{1.18}\ DC^{-0.184}\ Con^{0.229}\ DE^{-0.906}$	2.29
D4	$5.26 \times 10^2\ TD^{1.34}\ DC^{-0.253}\ Con^{-0.446}\ DE^{-0.411}$	6.06
D4 $_{\text{No Con}}$	$7.28 \times 10^2\ TD^{1.14}\ DC^{-0.068}\ DE^{-0.562}$	8.46
D4 $_{\text{No Con, No DC}}$	$7.36 \times 10^2\ TD^{1.05}\ DE^{-0.570}$	9.88

When the data set shows irregular tendencies, it is recommended to use a model that has different levels of natural hierarchy to best suit the data set. The hierarchical regression model can adjust to the unpredictability of a data set.

The hierarchical model, otherwise known as a polynomial model, is a form of non linear regression in which the relationship between the independent variable and the dependant variable is modeled as an nth order polynomial (Faraway, 2000). The equation pertaining to the polynomial regression model can be denoted as follows.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{5}$$

where

y, Target cost

β_0, Intercept

$\beta_{1,2}$ Regression coefficients or Slope

x, Specified cost driver

ε, Error

Fernandez-Juricic et al. (2003) examined the effect of human disturbance on the nesting of house sparrows. They counted breeding sparrows per hectare in 18 parks in the city of Madrid, and also counted the number of people per minute walking through each park.

Figure 6 describes the data points from their analysis on three different models: linear regression, quadratic regression and cubic regression. While using linear and quadratic regressions, it can be concluded that the results are not significant. However, the cubic regression graph is significantly superior with a coefficient of determination = 0.765 and a P-value = 0.0001. As this example depicts, a polynomial model is the most appropriate type of regression model to represent the human disturbance on the nesting of house sparrows. The cubic equation is as follows;

$$Y = 0.0443x^3 - 2.916x^2 + 50.601x - 87.765 \tag{6}$$

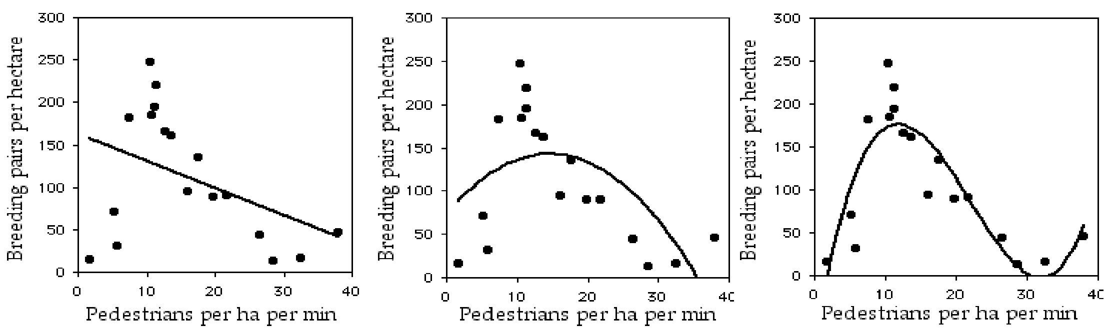

Figure 6: Graph of sparrow abundance vs. human disturbance, Fernandez-Juricic et al. (2003)

2.5 Summary

Target costing is primarily based on the products price, quality and functionality requirements as defined by the customers (Sani and Allahverdizadeh, 2012). Target costing is vital for any new products on the market in any type of industry. Being able to estimate cost at an early stage of product development is critical to understand in order to eliminate cost overrun.

As can be seen in this chapter, there has been previous work performed on target costing and parametric estimation that show promising results. This thesis will aim at comparing two types of target cost models and determining which method provides the most accurate result. These models will be applied to products in the aerospace industry. Since cost is a major a part of product development, target costing tools can add value to the product development process. A comparative parametric approach between linear and non linear regression is discussed in the following chapter utilizing the technique performed by Salehi and Hematfar (2011), Kahyani and Basiri (2011), Bashir and Thomson (2004). A great deal of attention pertaining to cost modeling has been given to one type of regression analysis in particular and will be shown in the results section.

3.0 METHODOLOGY

In order to determine a target cost model, one must first comprehend the different types of models under study. Regression models are classified into two broad categories namely linear and non-linear models (Rajarathinam and Parmar, 2011). These models are considered to be veritable tools for describing the functional form of the relationship between variables (Okereke, 2011). Many commercially available cost estimating packages use weight as the baseline cost driver and then generate measures of secondary cost drivers to refine the cost estimate (Curran *et al,* 2006). However, it is difficult to get an accurate weight estimate early in the early conceptual stages of design. To achieve the purpose of this study, two types of parametric equations will be presented and compared. They are the linear and non linear regression models.

3.1 Simple Linear Regression Model

The simple linear regression model (SLRM) assumes that the relationship between the dependent variable, denoted y, and the independent variable, denoted x, can be approximated by a straight line. Linear regression is the most common method of studying the linear relation between two or more variables (Kahyani and Basiri, 2011). The simple linear regression (SLR) is utilized when only one proven CER is capable of determining the cost of the product. The SLR model can be denoted as follows.

$$y = \beta_0 + \beta_1 X_1 \tag{7}$$

where

y, Target cost

β_0, Intercept

β_1, Regression coefficient

X_1, Specified cost driver

Ordinary least squares (OLS) is a method for estimating the unknown parameters in a linear regression model. OLS calculates the straight line through the data set which minimizes the error.

When there is more than one CER present in the regression model and the data set continues to portray linear tendencies, the multiple linear regression model will be utilized.

3.1.1 Multiple Linear Regression Model

In the event that a cost model depends on several parameters, the multiple linear regression model (MLRM) will be employed. According to Kutner *et al*. (2004), this type of parametric model is considered complex. The relevant formula for the MLRM is as follows:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \tag{8}$$

where

$\hat{y}$, Target cost

β_0, Intercept

β_k, Regression coefficients

X_k, Specified cost drivers

In order to calculate the regression coefficients, the method of least squares was used.

3.1.2 Least Squares Estimation

In least squares estimation (LSE), the unknown values of the parameters are called regression coefficients. The German scientist Karl Gauss (1777-1855) proposed estimating the regression coefficients to minimize the sum of the squared deviations between the observed data and the estimated data (Montgomery, 1984). This method can be observed in the below figure.

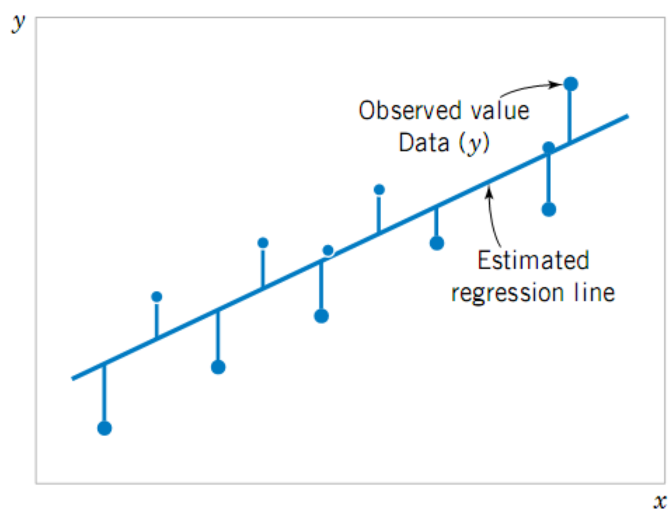

Figure 7: Deviations of the data from the estimated regression model (Montgomery *et al.*, 1984)

The necessary derivation of calculating LSE can be demonstrated in the work of Faraway (2000), Montgomery (1984) and Sajadifar and Allameh (2008). The equation to calculate regression coefficients is listed below.

$$\hat{\beta} = (X'X)^{-1} X'y \tag{9}$$

Sajadifar and Allameh (2008) have created different methods to compute multiple linear regression coefficients. They have modified the existing way to compute regression coefficients in order to make the computation more efficient. The least squared method of estimating the parameters under study is typically preferred to other methods as it yields unbiased estimators (El-Salam, 2011).

3.2 Multiple Non Linear Regression Model

In statistics, the NLM is a type of regression that utilizes data modeled in the form of non-linear combinations (Montgomery, 1984). As stated in the literature review, Salam (2009) introduced a parametric model based on estimating design effort of a compressor fan at Pratt & Whitney Canada. Below is the formula for the NLM.

$$\hat{y} = \beta_0 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \tag{10}$$

where,

$\hat{y}$, Target cost

X_m, Specified cost drivers

β_m, Regression coefficients

Upon examining the above nonlinear model, the natural log (ln) of the entire equation must be taken in order to be in the proper format for regression analysis. The assumptions of linearity hold when manipulating the data from non linear to linear. It is much easier to use simple linear regression to estimate the parameters of a nonlinear model. Following the data linearization, evaluating the model parameters concludes direct solutions using the least squares method. The linear equation generated is described as follows and takes into account the error (ε) that postulates to the logarithmic equation below:

$$\ln \hat{y} = \ln(\beta_o) + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \beta_3 \ln(X_3) + \varepsilon \tag{11}$$

In regression, it is known that the units of the dependant variables do not equate that of the independent variable. Therefore, the concept of dimensional analysis needs to be addressed. Dimensional analysis (DA) is a method employed to restructure the variables of a regression model into a set of independent dimensionless products (Vignaux and Scott, 1999). DA explicitly uses the constraint that all terms of the model must have the same dimension. Parametric equations utilize DA as it takes multiple cost estimating relationship of any dimension, transforms them into dimensioned constants, called regression coefficients, to form a correct dimensional relationship that is homogeneous. Dimensionally homogeneous implies that all the dimensions of all terms are the same. DA divides each factor by the reference value in order to generate dimensionless units.

For example, an equation can be composed of multiple dimensions including mass with dimension [M], time [T] and volume [L]. Moreover, problems in economics often add a dimension of cost to the function. The regression analysis process takes the multiple variables and creates dimension constants, also known as regression coefficients, to form a dimensional relationship. These dimensionless products of the variables can be explained by the Buckingham Π theorem (Buckingham, 1914). Buckingham (1914) stated that all relationships can be reformulated as a function of a set of dimensionless products of the variables. The Buckingham Π theorem is a process that satisfies physical dimensional homogeneity which involves n multiples variables and reduces them to Π dimensionless variables (Bender, 1978). The following will be explained by the two subsequent formulas.

The original relationship of multiple dimensioned variables is written as the following equation, where n represents the number of dimensions.

$$f(x_1, x_2, \ldots, x_n) = 0 \tag{12}$$

If the above formula is dimensionally homogeneous, the Π theorem states that the expression can be expressed as a new function of a set of dimensionless parameters written in terms of Π's. Here m represents the number of fundamental dimensions in the relationship.

$$\phi(\Pi_1, \Pi_2, \ldots, \Pi_{n-m}) = 0 \tag{13}$$

The above formula represents a new function that is equivalent to the old one with fewer variables. Further research about DA and Buckingham Π theorem can be found in the work of most applied mathematics and physical modelling textbooks such as; Langhaar (1951) and Huntley (1967).

The units are removed using the concept of DA furthermore they are still considered linear. Evaluating the suitability of the data is essential to postulate their linearity assumptions. Below are three techniques used to validate the linearity assumptions of a given data set.

3.3 Data Linearity Assumption

As stated by Okereke (2011), linear regression models are those that are linear in parameters. Data linearity is a critical parameter in regression. In order to determine if the function is linear, several key assumptions must be satisfied. These assumptions are listed below.

1. Linear relationship
2. Multicollinearity

3. Homoscedasticity

3.3.1 Linear Relationship

Firstly, linear regression requires that the relationship between the dependant and independent variables to be linear. It is important to validate all the data points and determine if there are any outliers. Scatter plots or residual plots can test the linearity of a data set. These plots allow for visual assessment of the relationship between the response and predictor variable (Weisberg, 2005).

In the case of the scattered plot, the standardized residual plotted against the non-standardized predicted value will determine if the data set is linear. The standardized residual is calculated with the following equation.

$$Z_{\varepsilon_i} = \frac{\varepsilon_{\mathrm{i}} - \bar{\varepsilon}}{\sigma_{\varepsilon}}, \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n} \tag{14}$$

where

Z_{ε_i}, Standardized residual of observation i

ε_{i}, Residual value of observation i

$\bar{\varepsilon}$, Mean of residuals

σ_{ε}, Standard deviation of residuals

The scatter plot should not have any visual curvilinear patterns as demonstrated in the below figure.

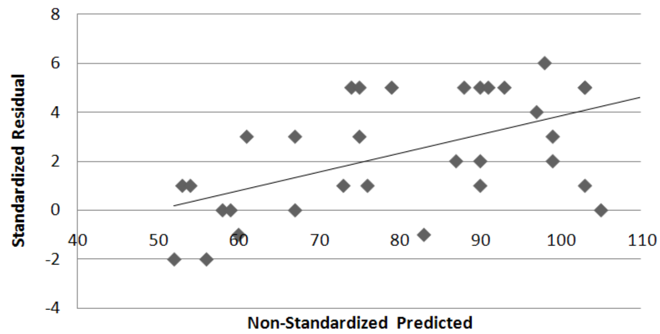

Figure 8: Linear scatter plot

In the case where the data set renders a curvilinear pattern, the linearity test fails. A curvilinear representation can be seen in the following figure.

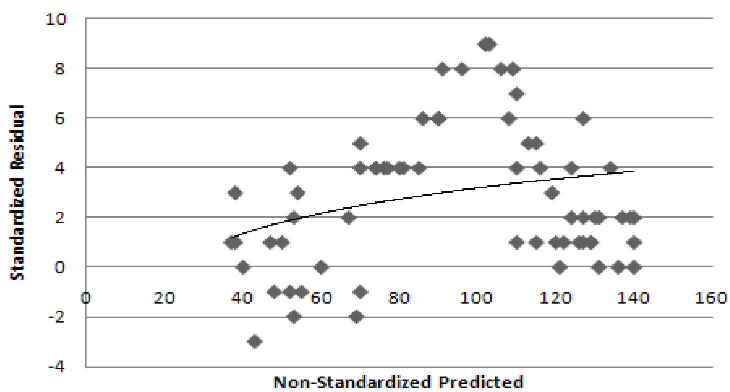

Figure 9: Non linear scatter plot

The residual plot is a graph which demonstrates the residuals on the vertical axis and the independent variable on the horizontal axis. The difference between the observed value and the predicted value is called the residual. Below is a formula to calculate the residual error.

Residual = Observed value – Predicted value

$$e = y - \hat{y} \tag{15}$$

If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data, indicating a good fit for a linear model. Otherwise, a non linear model is more appropriate. The residuals of a regression can be tabulated in the following example based on a sample size of ten units.

Table 5: Example of table of residuals

	Observed value	Prdicted value	Residual (Error)
Sample 1	98	103	-5
Sample 2	102	99	3
Sample 3	102	104	-2
Sample 4	102	97	5
Sample 5	105	105	0
Sample 6	103	101	2
Sample 7	98	97	1
Sample 8	103	101	2
Sample 9	95	96	-1
Sample 10	101	100	1

The linear regression model can be employed if the linearity assumptions are satisfactory. To test the linearity of a data set, the scatter plot of the standardized residual against the predicted values is required. Figure 10 demonstrates a random dispersion of data which represents a linear relationship. The residual plot of the aforementioned example is displayed below.

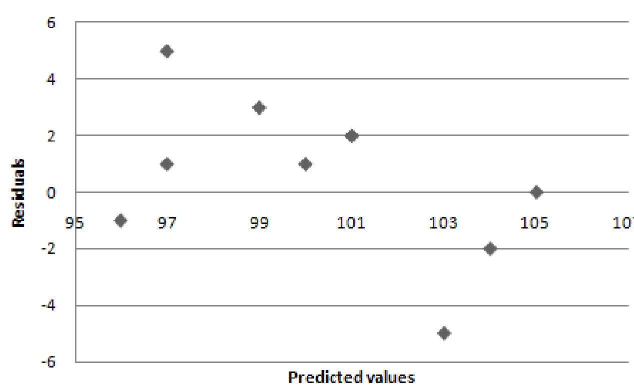

Figure 10: Example of residual plot

3.3.2 Multicollinearity

Secondly, linear regression assumes that there is little or no multicollinearity in the data. In regression analysis, it is expected to have dependencies between the response variable and the regressor (Montgomery and Runger, 2007). Multicollinearity occurs when two or more predictors in the model are correlated and provide redundant information about the response. Multicollinearity can have serious effects on the estimate of the regression coefficients and on the general applicability of the estimate model. Multicollinearity can be tested several different ways, such as: the correlation matrix and the variance inflation factor (VIF).

The correlation matrix must yield correlation coefficients smaller than 1 in order to assume that multicollinearity is not present. Correlation can be interpreted as a statistical measurement of the relationship between two variables. Possible correlations range from +1 to –1. A correlation of zero indicates that there is no relationship between the variables. A correlation of +1 indicates a perfect positive correlation, meaning that both variables move in the same direction together. The correlation matrix will be demonstrated in the data analysis section.

The VIF of the linear regression must render a value smaller than 10 which determines if multicollinearity is present (Kutner, 2004). If a VIF greater than 10 is yielded, there is certainly multicollinearity in the data set. The equation for the variance inflation factor is shown below.

$$VIF = 1/(1-R^2) \tag{16}$$

where,

VIF, Variance inflation factor

R^2, Coefficient of determination

As can be seen in the work of Faraway (2000), who worked on a multicollinearity example on employment within a population, rendered a VIF value of 42. This value can be interpreted that the standard error is 42 times larger than it would have been without the presence of multicollinearity. To remove the presence of multicollinearity, examine the correlation matrix and remove the variables that do not have a large pairwise correlation with the other variables (Faraway, 2000). Thereafter, the process of calculating the R^2 is repeated to determine the new VIF value with the omission of one less variable in the equation.

A stronger linear dependency of the independent variable(s) will enable a larger coefficient of determination. Hence, the VIF value will also yield a larger value. Montgomery and Runger (2007) discuss in detail about the presence of multicollinearity and different measures for solving this issue.

3.3.3 Homoscedasticity

The last data linearity assumption is homoscedasticity. Homoscedasticity can be observed when the data set exhibits similar amounts of variance across the range of values for an independent variable (Kim and Bentler, 2002). Equal variance is essential across the data for the linearity assumption to hold since the variance measures the dispersion of a set of data points around their mean value. The equation to calculate the variance is as follows;

$$\sigma^2 = \sum (x - \mu)^2 / N \tag{18}$$

where

σ^2, Variance

$\sum$, Sum

x, Observed data point

μ, Mean of data set

N, Number of data points

The scatter plot is an excellent validation tool in order to determine whether the data demonstrates homoscedasticity. As can be seen in the following example, the variance across the data set is very similar, represented by the grey bars, which deems that the data set displays linear tendencies.

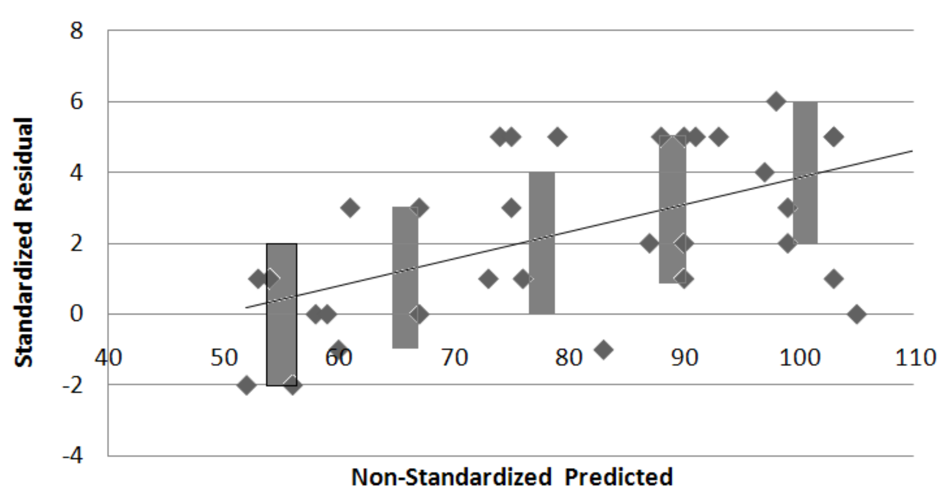

Figure 11: Variance across data set

In this study, the three linearity tests will be conducted in the case study in Chapter IV. In order to comprehend which parametric model will generate a more accurate target cost between both models aforementioned, the percent errors will be compared and two different methodologies will be employed to determine the significant factors.

The reliability of the models will be validated by two different methodologies to determine which parameter(s) become significant and will therefore be used to determine the cost. These methodologies are the analysis of variance and path analysis. It should be noted that these methodologies take into account the cost estimating relationships. The requirements to determine the most suitable regression model along with the ideal technical parameter(s) will be discussed in the following segment.

3.4 Cost Estimating Relationship

According to ISPA, cost estimating relationships can be defined as a mathematical expression of varying degrees of complexity expressing cost as a function of a cost driving variable. Cost drivers are any factors which cause a change in the cost of work performed in the lifecycle of a product. The identification and selection of cost drivers are fundamental to a cost estimating model. Without adequate data and CER, the cost models will have no added value in the early conception phase of estimation. The CER is an integral part of regression analysis and its validation is crucial.

3.4.1 CER Validation

Once the data collection and validation has occurred, it is imperative to pair the raw data with valid CERs. The ultimate test of the goodness will determine whether or not a particular CER can accurately predict the cost of a component. Several mathematical tests are available to comprehend the most significant CER. These tests are as follows:

1. Standard error of estimate (SEE)
2. Coefficient of correlation
3. Coefficient of determination

The standard error of estimate measures the accuracy of the prediction. It can also be termed as the standard deviation of the data set. Therefore, if the data demonstrates large dispersions, a higher SEE will be calculated. This represents that the data set in the study tends to be far from the regression line. SEE uses the regression line that minimizes the sum of squared deviations from the prediction. The standard error of estimate's equation is listed below.

$$SEE = \sqrt{\frac{\Sigma\,(Y - \hat{Y})^2}{N}} \tag{19}$$

where

SEE, Standard error of estimate

∑, Sum

Y, actual result

Ŷ, predicted result

N, number of data points

The coefficient of correlation (r) and the related coefficient of determination (R^2) are certainly the two most commonly used measure of goodness of fit. The value of r is calculated from the following equation.

$$r = \pm\sqrt{\mathrm{R}^2} \tag{20}$$

These statistical results measure the amount of correlation between the prediction and the corresponding data set. Moreover, it is an excellent indicator of the accuracy and denotes the strength of the equation. For example, if the coefficient of correlation is r = 0.875, then the coefficient of determination is R^2 = 0.766. This represents that 76.6% of the total variation can be explained by the linear relationship in the regression. The remaining 23.4% of the total variation in the equation remains unexplained. Hence, a high coefficient of determination yields a minimum amount of disparity in the equation. A metric for determining the proportion of the variation explained by the independent variables is the coefficient of determination (Montgomery *et al.*, 2001) given as:

$$R^2 = 1 - \frac{SS_r}{SS_T} \tag{21}$$

where

$$SS_r = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \tag{22}$$

$$SS_T = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{23}$$

where

$\bar{y}$, mean

SS_T, total sum of squares

SS_R, residual sum of squares

There are multiple methods to determine the validity of a cost driver. The following section describes quantitative metrics to establish the significance of a CER.

3.4.2 CER Significance

The significance of the CER generates confidence in the regression equation and the assurance of its forecasting capability. It is known that there are numerous ways in statistics to evaluate the significance of a cost estimating relationship. The below table demonstrates several key criterion when determining the ideal CER.

Table 6: Key attribute for regression analysis (ISPA, 2011)

Criteria	*Good*	*Marginal*
P-value of the f-test	≤ 0.10	≤ 0.15
P-value of the t-test	≤ 0.10	≤ 0.15
Coefficient of Variation (CV)	≤ 0.25	0.25 to 0.30
R-squared	≥ 0.70	0.35 to 0.70

A caution is warranted when performing statistical analysis of a relationship. There is no one statistic that disqualifies a CER or model, nor is there any one statistic that validates a CER or model. The math modeling effort must be examined from a complete perspective, starting with the data and logic of the relationship.

3.4.3 CER Strengths

CERs can be excellent predictors when implemented correctly, and they can be relied upon to produce quality estimates when used appropriately. Several CER strengths pertaining to estimation are as follows:

1. Capability to reduce the amount of time to evaluate cost estimates
2. Ability to develop and produce prompt estimates

3. Minimal information is required concerning the product estimated
4. Practicability when estimating in early conceptual phase of a program

3.4.4 CER Weaknesses

On the other hand, CERs must be used carefully and the process of selecting the ideal CER is essential. Below are certain examples of cost estimating relationship weaknesses.

1. Performing a detailed estimation can be more reliable than CER estimation
2. Employing incorrect cost or technical data may skew the chosen CER

3.5 Analysis of Variance

The analysis of variance (ANOVA) can be used to test the significance of regression analysis. According to Montgomery (1984), analysis of variance is a method of decomposing the total variability in a set of observations, as measured by the sum of the squares of these observations from their average, into component sums of squares that are associated with specific defied sources of variation.

In order to judge the adequacy of a regression model, the coefficient of determination is utilized. The coefficient of determination represented by R^2, measures the percentage of variation explained by the model between 0% and 100%. A marginal result can be found between 35% and 70%, however above 70% is considered good (ISPA, 2011). Additionally, the t-statistic and the related p-value are both important in this methodology since it estimates the probability level at which the statistical test would fail, suggesting the relationship is not valid. As per the ISPA, a p-value less than 0.10 is considered acceptable for inferring that the selected factor remains a significant cost driver. Moreover, a p-value less than 0.15 is deemed marginal. If the CERs are considered insignificant, the analysis is repeated by removing one CER at a time and the analysis is continuously repeated until only significant factors remain. Furthermore, the confidence interval utilized for significance in this study is set at 90%. To do so, the probability level at which the statistical test would fail, suggesting the relationship is not valid should have a p-values less than 0.10.

The analysis of variance will be employed to determine which CER(s) become significant. The ANOVA methodology will be compared to the path analysis which takes into account the effects of each parameter on the outcome.

3.6 Path Analysis

Wright (1934), known for his influential work on path analysis (PA), takes into account the approach used to study the direct and indirect effects of variables. This methodology analyses the effect that each of the parameters will have on the output of the equation. By examining the possible linkages between each potential cost driver, their respective path coefficients will determine a standardized method to conclude the significant factor(s). PA is designed to produce measures of relationship between variables (Smith and Murray, 1978).

The path analysis process begins by determining the correlation between the potential cost drivers. In statistics, the correlation indicates the strength and direction of a linear relationship between two random variables, which is essential to comprehend their associations (Kutner *et al*, 2004). The higher the correlation, the better the effect of the CER will pertain to the regression.

The regression coefficients, B_i, are generated in the summary output of the ANOVA. Thereafter, they are utilized in the below equation to determine their path coefficients. The path coefficients presented by (Li, 1975) are the result of the following equation:

$$P_{0i} = \frac{B_i \sigma_i}{\sigma_y} \tag{24}$$

where,

P_{0i}, Path coefficient

B_i, Regression coefficient

σ_i, Standard deviation of the CER

σ_y, Standard deviation of the MLG cost

It should be noted that path analysis is not merely an ordinary regression analysis. Path analysis employs regression analysis to compute path coefficients. In path analysis, an equation represents a causal link whereas in regression analysis, an equation represents the dependant variable as a function of the independent variable (Smith and Murray*,* 1978).

Lastly, the value of U in relationship with the final equation is calculated, which represents the uncorrelated residual of the function (Li, 1975). The path coefficient for the uncorrelated residual is calculated by;

$$P_{0u} = \frac{\sigma_e}{\sigma_y} \tag{25}$$

where,

P_{ou}, Path coefficient for the uncorrelated residual

σ_e Standard deviation of the error

σ_y, Standard deviation of cost

A path diagram is a graphic display of the order in which variables are assumed to affect on another. The path analysis diagram is shown below.

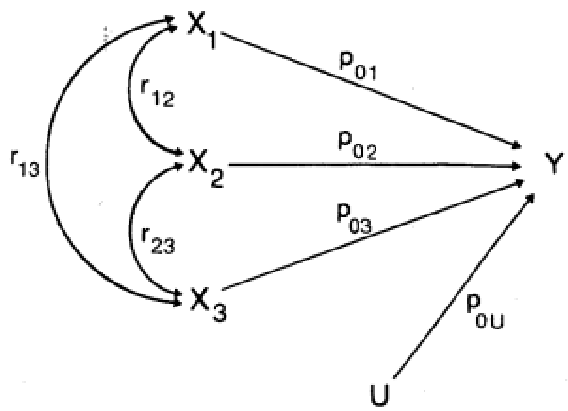

Figure 12: Path diagram (Li, 1975)

Land (1969) describes the convention for drawing path diagrams. It consists of arrows that can be drawn from variables acting as causes to variables acting as effects. The initial variables are linked to one another by curved lines with double arrows. Then the initial variables are linked to the output by a straight arrow and finally the output is linked to the uncorrelated value of the function by a straight arrow. The path analysis' strengths and weaknesses are as follows.

3.6.1 Strengths of Path Analysis

Path analysis provides the study with unique advantages not available with other methods, such as;

1. Provides means for modeling of complex problems: Path analysis provides means of incorporating unobservable variable into evaluative studies. The

resultant models are more representative of the dynamic reality (Smith and Murray, 1978).

2. Presents reflection of alternative models: Path analysis requires the explicit specification of presumed causal relationships and forces the researcher to consider several alternative models (Chamberlin, 1965).

3. Enables the study of both direct and indirect effects: Path analysis enables the study of both the direct and indirect effect on dependant variables by analyzing their correlations

3.6.2 Weaknesses of Path Analysis

Some of the technical problems associated with the use of path analysis procedures are presented below:

1. Multicollinearity: No possibility to determine what proportion of the variance is accounted for by the variable when the data under study are interrelated. Multicollinearity often arises when multiple indicators are used in the regression for the cost drivers under study (Pedhazur, 1975).

2. Measurement errors: One of the assumptions in path analysis is the measurements are free of error. Otherwise a greater amount of ambiguity in the path coefficients will be inherent.

3. Complexity of interpretation: Path analysis models with four or more variables become increasingly complex to interpret. Moreover, the

calculation of the coefficients and their interpretation can be tedious and inconclusive.

In the end, path analysis provides a unique capability to study the direct and indirect effects of the variable on the output.

3.7 Summary

When used appropriately, regression analysis can be a powerful tool (Constantine, 2012). A comparative analysis will be presented between two different types of regression cost models, in the following section. Thereafter, the regression models will be used in a case study in order to determine which model will yield greater results. The reliability of the models will be validated by two different methodologies to determine which parameter(s) become significant. These methodologies are the analysis of variance and path analysis. It is essential to determine the most predominant CER's that may have significant effects on the independent variable of the case study. Finally, as the data used in this thesis is confidential, data masking technique are described in chapter 4 along with the recommended data masking technique. The following chapter will present a case study along with the data analysis from the different types of estimation methods under study.

4.0 CASE STUDY AT BOMBARDIER AEROSPACE

This chapter provides a comparative analysis between linear and non-linear regression models in the aerospace industry. These models will be applied to a case study involving the main landing gear at Bombardier Aerospace. The reliability of the models will be validated by two different methodologies to determine which parameter(s) become significant and will therefore be used to determine the cost. These methodologies are the analysis of variance and path analysis. The ultimate goal of this research is to determine which target costing tool will increase the credibility of the cost estimate to predict the target cost of the main landing gear at Bombardier Aerospace.

4.1 Bombardier Aerospace

Bombardier is the world's only manufacturer of both planes and trains. The Canadian founded company focuses on evolving mobility and making it easier for people to connect with others. Bombardier Aerospace holds the most comprehensive aircraft portfolio totalling 13 programs, including:

1. Business aircraft: Learjet, Challenger and Global platforms
2. Commercial aircraft: Q-Series, CRJ Series and CSeries
3. Amphibious aircraft: Bombardier 415

Bombardier's core values include commitment to excellence, customer orientation and shareholder focus. Recently, commercial considerations have become an integral part of all key decisions. As the threat of new aircraft manufacturer are being established in low cost countries such as: Japan, China, Russia and Brazil, competitive pricing is playing a crucial role for aircraft operators.

According to Wall (2011) two main challenges can be faced when integrating target costing practices, they are: the re-design challenge and the development rhythm challenge. In order to continuously drive down the cost of the aircraft, supply chain has collaborated with the advanced design team at Bombardier aerospace in order to optimize the cost of new aircraft design in the early conceptual phases of product development. Jörgensen and Messner (2009) describe how the necessity of integrating multidisciplinary teams to the product development phases are necessary whereas Mouritsen et al (2009) illustrates how accounting practices are problematic for re-designs phases. By creating a financial awareness to the engineering community, target costing is seen as a central process for linking product development to customers, owners and suppliers (Östman 2009). Moreover, previous studies have demonstrated that value engineering plays an important role in target costing (Agndal and Nilsson, 2009). Value engineering can be seen as any activity deemed necessary to improve the function or reduce the cost of a product.

In order to counter the challenges of today economy, increased competition and increased costs, a multidisciplinary design optimization (MDO) tool is in the process of being created in order to make the appropriate trade offs between design and cost for future programs. Consequently, the cost of research and development will significantly decrease with the understanding of a costing capability upfront in the design process. This cost understanding will eliminate potential design concepts that are deemed too expensive and would eventually lead to cost overruns. As stated by Davila and Wouters (2004), reducing costs while simultaneously solving technical problems is a challenging task.

Regression models, when adapting them to the aerospace industry, take into account technical specifications that drive the cost of the product. In this thesis, regression models will be developed and integrated in the MDO tool in order to help the decision makers comprehend the cost aspect in the conceptual design phase of the aircraft and reduce the overall life cycle cost of a new program. This chapter will present the case study pertaining to the main landing gear at Bombardier Aerospace.

4.2 Landing Gear

The landing gear, also known as the undercarriage, is utilized as an interface between the aircraft and the ground (Pazmany, 1986). The landing gear can be divided into different sub-systems such as: main landing gear (MLG), nose landing gear, extension and retraction system, alternate release system, steering system and brake control system. The main functions of the landing gear are to absorb loads upon landing, taxiing and braking. The load during landing is absorbed by the gears and it is proportional to the maximum takeoff weight (MTOW). The MTOW, typically measured in pounds (lbs), is the heaviest weight at which the aircraft can fly and meets all the applicable airworthiness requirements. The following figure is a visual representation of a landing gear.

Figure 13: Superjet 100 Landing gear (Messier-Bugatti-Dowty, 2010)

After conducting multiple interviews with the engineering community and the landing gear supplier chain agents, the potential technical specifications that can drive the cost of the main landing gear and which will be analysed for this case study are as follows:

1. The maximum take off weight
2. The weight of the main landing gear
3. The height of the main landing gear

Interviews were conducted with a multitude of candidates from all levels of management and the discussions were triangulated using the questionnaire placed in the appendix section. In order to establish target cost models, credible data must be collected before it can be used effectively. The following section provides an overview of the processes required to collect data to be used in parametric applications.

4.3 Data Collection

The data collection phase is one of the most critical and time consuming steps in producing an estimate. All parametric estimating techniques require credible data before they can be used effectively. Without credible data and data sources, the outcome of a cost model, which is the target cost, is irrelevant. The parametric techniques utilized in this thesis require two types of data collection. Firstly, historical cost data is gathered and assembled specifically for the current research project. The data is collected with the use of the contractual prices. Secondly, technical data is gathered seeing as it describes the physical, performance, and engineering characteristics of a product. According to Bengtsson and Sjöblow (2006), expert interviews are necessary to create understanding and avoid misunderstanding of the system under study. In essence, cost interviews were conducted with numerous supply chain agents and managers in order to normalize the cost in a comparable fashion. The cost data can be subject to a lot of variability due to the fact that the negotiations were performed on different programs, commercial versus business, and in different fiscal years dating back in the early 80's. Therefore, negotiation techniques and business cases have drastically changes over the past 30 years, cost transparency increases as well as increases in the aircraft planning base which drives the cost of the system. Similarly, technical interviews with the engineering and advanced design teams were conducted in order to get multiples point of views of the different potential cost drivers to be studied. Interviewing multiples sources and triangulating the results is a key factor in order to build credibility in the technical and commercial data gathering phase. The interview grids can be seen in APPENDIX C. In order to make a successful study, once the technical parameters are assessed, the technical data must be collected and validated by numerous sources. It is known that the aerospace industry is subject to data scarcity. Bombardier Aerospace was launched in 1986 with the acquisition of Canadair Ltd. Hence, the data points collected are limited to the programs on hand. Knowing that a new aircraft program is launched every few years, data scarcity can be a predominant factor in this study. The cost data for this thesis was collected with the contractual agreement between the supplier and BA in a given year. The technical data are precise constraints that can be found in the technical requirements document. This document provides technical requirements of the commodity under study. All data points have been cross validated by many sources in the company. The assessment of data linearity is crucial in order to suggest a good fit for a linear regression model. The validity and reliability of the data are an important factor due to the fact that it is used as a basis of the research.

4.3.1 Data Validity and Reliability

To assure the credibility of a cost model, special attention must be placed on the validity and reliability of the data. Data validity can be defined as the ability of accurately measuring information, whereas data reliability is best described as the degree to which the data are free from error and yield consistent results (Eriksson and Wiedersheim-Paul, 2001)

The collected data must be adjusted for items such as production rate, improvement curve, and inflation. This is also referred to as the data normalization process. According to Bengtsson and Sjöblow (2006), it is important to reflect upon which errors could have occurred while gathering information.

In the aerospace industry, the main variability within the commercial data is due to the commodities negotiation strategy. Some commercial strategies involve paying the suppliers development costs up front in order to obtain a reduction on shipset price. Whereas, in some cases where funding is scarce, commodities tend to make the supplier absorb all development costs which yields a higher recurring price. The thesis takes into account as a baseline that all commercial agreements are treated equally and that the collected data is assumed to be fair. Due to confidentiality purpose, the data collected at Bombardier aerospace cannot be divulged to the public therefore data masking techniques were performed.

4.4 Data masking

Data masking is the process of protecting sensitive information for wider visibility (Edgar, 2000). Data masking techniques are often used to protect confidential and numerical data from unauthorized individuals while providing maximum information. Data masking, otherwise known as data sanitation or data perturbation, is the process of disguising sensitive information by overwriting it with data of a similar type. Data security and protection of sensitive data has received considerable attention in literature in recent years. Therefore, in order to protect proprietary information of the industrial collaborator, the cost data and any sensitive technical data must be masked.

There are many techniques which can be used to sanitize sensitive data; several of these techniques will be discussed below.

4.4.1 Number Variance Technique

The number variance technique (NVT) is useful on numerical data. The technique entails modifying each numerical value in the data set by some random percentage of its real value. The major advantage of this procedure is that it provides the data set with a reasonable disguise for the numerical data while still keeping the range and distribution of values in the column within viable limits (Edgar, 2000). For example, the original data set will be similar if the observed data are changed within a range of 10%, as displayed in the following table. NVT is advantageous in preventing others to correlate the known numerical data.

Table 7: Example of number variance technique

	Observed value	Number Variance Technique	Masked Data	Difference
Sample 1	96	+-10%	90	-6
Sample 2	97	+-10%	90	-7
Sample 3	102	+-10%	99	-3
Sample 4	100	+-10%	96	-4
Sample 5	96	+-10%	100	4
Sample 6	105	+-10%	100	-5
Sample 7	101	+-10%	96	-5
Sample 8	103	+-10%	95	-8
Sample 9	95	+-10%	94	-1
Sample 10	96	+-10%	104	8

4.4.2 Substitution Technique

The technique of substitution consists of randomly replacing the contents of a column of data with similar information but is completely unrelated to the real details (Edgar, 2000). This can be an effective technique to use if dealing with non-numerical data. For example, aircraft names can be modified with code names in such a way that the public would not be able to recognize the programs. However, if the need to substitution millions of data points is required, substitution can become a tedious task. Substitution is quite powerful, reasonably fast and preserves the look and feel of the data. Finding the required random data to substitute and developing the procedures to accomplish the substitution can be a time consuming effort.

4.4.3 Character technique

This data masking technique entails replacing certain fields with a mask character (such as an X). This effectively disguises the data content while preserving the same formatting on front end screens and reports. This minimizes the risk that the confidential information is disclosed to unauthorized personnel. This technique is useful when disguising credit card information, for example:

4346 6454 0020 5379

4493 9238 7315 5787

4297 8296 7496 8724

After the masking operation the information would appear as:

4346 XXXX XXXX 5379

4493 XXXX XXXX 5787

4297 XXXX XXXX 8724

The character data masking technique effectively remove the sensitive content from the record while still preserving the look and feel. It is critical to ensure in this type of data masking technique that most of the record is replaced by characters in order to preserve security. It is not suggested to mask a data point such as: 4297 8296 7496 87XX since the original data point can easily be determined by process of elimination. This research is based on the masked cost (Y_i) and technical data of thirteen MLG's presented in Table 8. Data scarcity in the aerospace industry is common and can hinder the accuracy of the results. Thus for both models under study, sub-samples of the data were randomly generated, where ten programs were used to generate the model and the three remaining programs were used for validation purpose. Three trials were conducted for this analysis. For the purpose of validation, the programs (6, 7, 13), (3, 7, 9) and (2, 8, 12) were removed from the data for trials 1, 2, and 3, respectively. Trial 1, omitting programs (6, 7, 13), is the selected evaluation to be presented in the following section and the remaining two trials will be demonstrated in the APPENDIX. Finally, a discussion pertaining to the findings of the thesis is presented.

4.5 Analysis Based on the Multiple Linear Regression Model

In the MLRM, each potential cost driver is represented by the variable Xm. From this point on and as it can be seen in the subsequent table, the following notation will be used. Moreover, for the ANOVA, the confidence interval for significance is set at 90%.

It should be noted that the data in Table 8 are contractually based and assumed to be normalized. Moreover, the actual costs and cost estimating relationships are masked to protect confidential information. While masking the data, it is important to maintain the characteristics of the original data, thus the number variance masking technique described by Edgar (2000) is utilized.

Table 8: Historical data for MLRM

	Cost Estimating Relationship			
Programs	X1: Weight (lbs)	X2: MTOW (lbs)	X3: Height (in)	Y: Cost (USD)
1	0.0027	0.2640	0.0009	0.5105
2	0.0027	0.2904	0.0009	0.5557
3	0.0031	0.3440	0.0009	0.5904
4	0.0039	0.5160	0.0010	1.0053
5	0.0016	0.3028	0.0003	0.6281
6	0.0021	0.3808	0.0003	0.9427
7	0.0026	0.4240	0.0003	0.8371
8	0.0027	0.4080	0.0003	0.8284
9	0.0043	0.5820	0.0004	0.8254
10	0.0043	0.6440	0.0004	0.9127
11	0.0048	0.6878	0.0004	0.8208
12	0.0042	0.7400	0.0006	0.8352
13	0.0042	0.7840	0.0006	0.8353

where,

X1, Weight of the MLG

X2, MTOW of the program

X3, Height of the MLG

Y, Masked cost of the MLG

For trial 1, the generated equations to estimate the target cost using (10) are summarized in Table 9. For the MLRM case, the MTOW (X2) is the only factor which has a p-value under 0.10, hence rendered as the only significant factor. The process utilized to eliminate the height and weight as possible CER's is as follows: by performing regression analysis it is important to validate several criteria. As per Table 6, the key attributes for regression analysis, such as: the P-value, the CV and the R-squared must be satisfied in order to consider a valid CER. Upon first analysis, the height variable (X3) did not satisfy the criterion listed above due to a high P-value of 0.9335. Hence, the height variable was discarded. The regression analysis process was replicated until all key criteria were satisfied.

Table 9: MLRM resulting equations

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.4058 - 16.2662X1 + 0.8125X2 + 20.2282X3$	0.5937
2	$\hat{Y} = 0.4167 - 9.4162X1 + 0.7673X2$	0.5932
1	$\hat{Y} = 0.4071 + 0.7201X2$	0.5925

For the MLRM, it can be observed that the coefficient of determination, R^2 remains similar when removing a CER. R^2 is a statistical measurement that provides information about the goodness of fit of a model. For every subsequent replication, a parameter which is rendered insignificant as explained by the p-values in below table is removed. In addition, the largest p-value gets omitted as indicated by the word "omit" in the following table.

Table 10: P-values for the MLRM

	P-value	P-value	P-value
	3 parameters	2 parameters	1 paramater
Intercept	0.0895	0.0239	0.0052
X1: Weight (lbs)	0.8992	0.9175	omit
X2: MTOW (lbs)	0.3174	0.1647	0.0092
X3: Height (inch)	0.9335	omit	omit

In order to comprehend which CER becomes significant, the highest correlation in Table 11 is taken and squared to understand the involvement of that specific CER to the cost model. The outcome of the procedure takes into consideration the correlation between the MTOW and the cost noted as 0.7698, then squared to become 0.5925, which can also be seen in Table 9. Therefore, 59.25% of the MLG equation is represented by the MTOW for the MLRM. According to ISPA, a coefficient of correlation between 35% and 70% is deemed marginal.

Table 11: Correlation matrix for the MLRM

	X1: Weight	X2: MTOW	X3: Height	Y: Cost
X1: Weight	1.0000			
X2: MTOW	0.8900	1.0000		
X3: Height	-0.0554	-0.3670	1.0000	
Y: Cost	0.6733	**0.7698**	-0.2817	1.0000

Table 11 represents the correlation factors between each CER for the MLRM. These correlation factors are integrated in equation (20) and (21) to generate the numerical results on the path analysis diagram shown in Figure 14. The path coefficients summarized below using the work of Wright (1934) serve as the interactions between each parameter, describing their interconnecting strengths.

Table 12: Path coefficients for the MLRM 3 parameters

Regression Coefficients		σ_i / σ_y	Path Coefficients
B1 =	-16.2662	0.0060	-0.0980
B2 =	0.8125	1.0690	0.8686
B3 =	20.2282	0.0016	0.0317

It can be deduced that the highest interconnecting strength is determined by the highest path coefficient. Recall that the path coefficients can be calculated by the multiplication between the regression coefficients and the ratio of the CER standard deviation and the cost standard deviation. In this regard, the regression coefficient B2, representing the MTOW, demonstrates the largest effect on the overall cost of the MLG.

Table 13: Results based on all three parameters for the MLRM

Programs	X1: Weight (lbs)	X2: MTOW (lbs)	X3: Height (in)	Y: Cost (USD)	Ŷ: Cost (USD)	*e*
1	0.0027	0.2640	0.0009	0.5105	0.5946	0.0841
2	0.0027	0.2904	0.0009	0.5557	0.6161	0.0603
3	0.0031	0.3440	0.0009	0.5904	0.6526	0.0623
4	0.0039	0.5160	0.0010	1.0053	0.7814	-0.2239
5	0.0016	0.3028	0.0003	0.6281	0.6329	0.0048
8	0.0027	0.4080	0.0003	0.8284	0.7008	-0.1276
9	0.0043	0.5820	0.0004	0.8254	0.8183	-0.0070
10	0.0043	0.6440	0.0004	0.9127	0.8687	-0.0439
11	0.0048	0.6878	0.0004	0.8208	0.8962	0.0755
12	0.0042	0.7400	0.0006	0.8352	0.9507	0.1155
6	0.0021	0.3808	0.0003	0.9427	0.6877	-0.2550
7	0.0026	0.4240	0.0003	0.8371	0.7142	-0.1229
13	0.0042	0.7840	0.0006	0.8353	0.9864	0.1512
mean	0.0034	0.4779	0.0006	0.7512	0.7512	0.0000
var	0.0000	0.0319	0.0000	0.0279	0.0166	0.0113
st dev	0.0010	0.1786	0.0003	**0.1670**	0.1287	**0.1065**

It must be noted that programs 6, 7 and 13 in the above table are not part of the calculation towards the mean, variance and standard deviation and only serve as validation points in this study. In order to calculate the uncorrelated value of the function according to all three parameters, Table 13 was generated. This table takes into account the standard deviation of the errors (*e*) with three variables along with the standard deviation of the cost (Y) which is essential in calculating the uncorrelated value of the function in the path analysis method. The PA methodology will be utilized to comprehend the strengths between the 3 potential cost drivers. At the same time the path analysis will reduce the number of variables while demonstrating the strongest link between the technical and commercial relationship. The uncorrelated value of the function is calculated using (21) and can be seen in Figure 14.

$$P_{ou} = 0.1065/\ 0.1607 = 0.6374$$

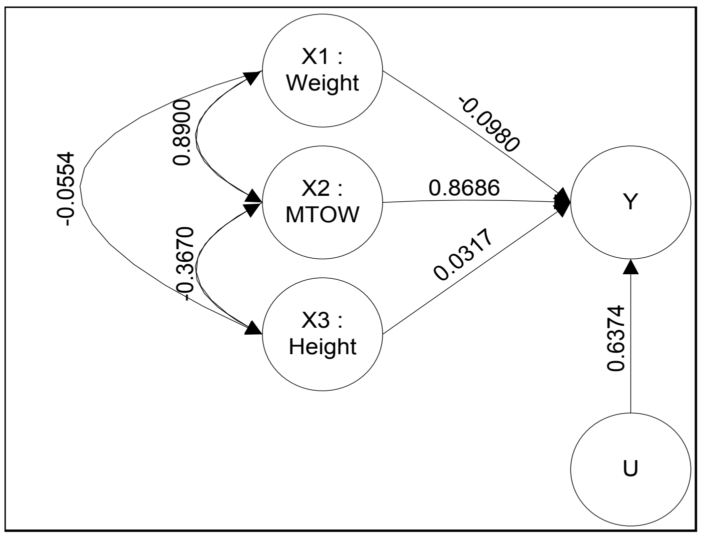

Figure 14: Path analysis diagram for MLRM with 3 parameters

If the path analysis process is repeated by removing insignificant CER's, the MTOW would result as the most contributing factor for the MLRM.

Table 14: Results based on the MTOW for the MLRM

Programs	X2: MTOW (lbs)	Y: Cost (USD)	Ŷ: Cost (USD)	*e*
1	0.2640	0.5105	0.5972	-0.0867
2	0.2904	0.5557	0.6162	-0.0605
3	0.3440	0.5904	0.6548	-0.0645
4	0.5160	1.0053	0.7787	0.2266
5	0.3028	0.6281	0.6252	0.0030
8	0.4080	0.8284	0.7009	0.1275
9	0.5820	0.8254	0.8262	-0.0008
10	0.6440	0.9127	0.8708	0.0418
11	0.6878	0.8208	0.9024	-0.0816
12	0.7400	0.8352	0.9400	-0.1048
6	0.3808	0.9427	0.6813	0.2613
7	0.4240	0.8371	0.7124	0.1246
13	0.7840	0.8353	0.9717	-0.1364
mean	0.4779	0.7512	0.7512	0.0000
var	0.0319	0.0279	0.0165	0.0114
st dev	0.1786	**0.1670**	0.1286	**0.1066**

Similarly to Table 13, the above table corresponds to the data based on the most contributing factor in the case study along with the cost associated to the MTOW, the predicted cost and the calculated error.

Table 15: Path coefficients for the MLRM 1 parameters

	Regression Coefficients	$\sigma_{i\,/}\,\sigma_y$	Path Coefficients
B2 =	0.7201	1.0690	0.7698

In the same way, the uncorrelated value of the function attributed to the MTOW is calculated using Table 15. Once the insignificant parameters have been disregarded,

Figure 15 represents the path of the final equation. The uncorrelated value of the function according to the MTOW is as follows:

$$P_{ou} = 0.1066/\ 0.1670 = 0.6383$$

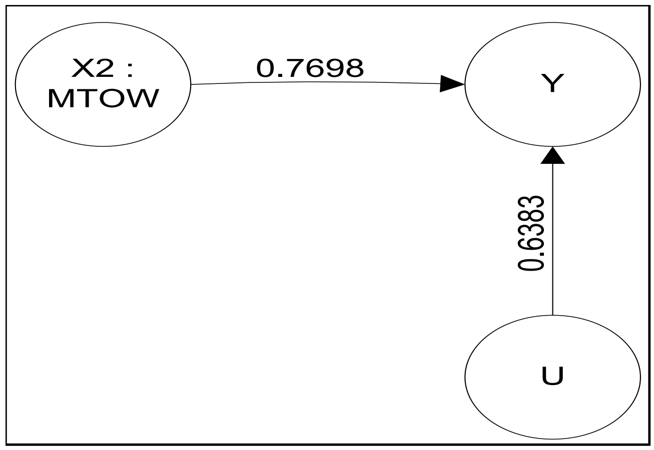

Figure 15: Path analysis diagram for MLRM with 1 parameters

In the end, for the MLRM, the MTOW is the most appropriate cost estimating relationship to predict the cost of the MLG. Figure 16, represents the actual and predicted cost versus the MTOW for the MLG. The actual data points are presented by the blue dots and the predicted cost are equally aligned on the linear regression curve. Based on the above analysis the final equation to be used to predict the cost of a new MLG is the following:

$$\hat{Y} = 0.4071 + 0.7201 * MTOW$$

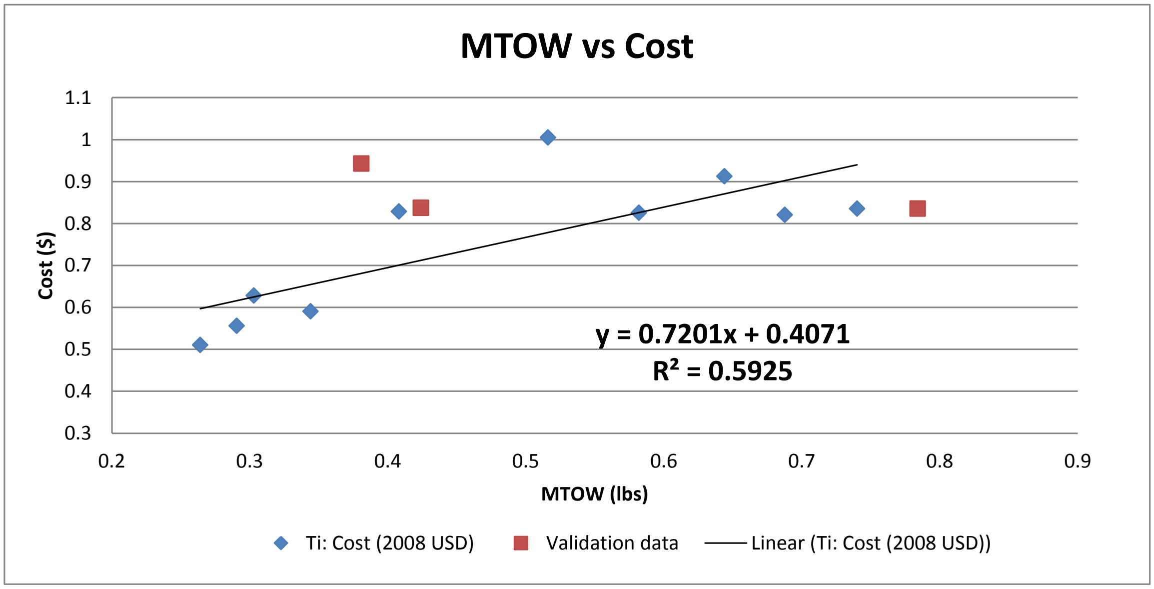

Figure 16: Cost versus the MTOW for the MLG (Linear Regression)

As can be seen from the above figure, it is clear that the linear regression model is lacking precision when it comes to estimation based on the coefficient of determination. Therefore, the necessity to develop a secondary estimation process based on non linear proprieties is eminent and will be discussed in section 4.6. The percent errors between the actual cost and the predicted cost are described below.

Table 16 presents the percent errors for the linear regression for all the 10 data points that were used to test the data, followed by the 3 validation programs. The model shows an average error of 1.84% based on ten programs, and an average error for the validation data of -8.76% based on 3 programs. The regression R-Square is 0.5925 which based on the international society of parametric analyst is deemed good. The overall regression is good for the level of confidence desired. In order to determine if the above data is linear, the linearity assumptions discussed in section 3.3 must be fulfilled.

Table 16: Percent errors of the MLRM based on the MTOW

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.5105	0.5972	16.9814
2	0.5557	0.6162	10.8890
3	0.5904	0.6548	10.9215
4	1.0053	0.7787	-22.5394
5	0.6281	0.6252	-0.4724
8	0.8284	0.7009	-15.3914
9	0.8254	0.8262	0.0990
10	0.9127	0.8708	-4.5813
11	0.8208	0.9024	9.9413
12	0.8352	0.9400	12.5443
6	0.9427	0.6813	-27.7240
7	0.8371	0.7124	-14.8908
13	0.8353	0.9717	16.3288
Ave. error data			1.8392
Ave. error (6,7,13)			-8.7620

4.5.1 Data Linearity Analysis

In order to employ the MLRM, the linearity assumptions must be satisfied. The first linearity assumption requires that the data set has a linear relationship between the dependant and independent variables. The scatter plot can test the linearity of a data set. In order to perform a scatter plot, the standardized residuals must be plotted against the non-standardized predicted values of the regression. The residuals, denoted by e, and the predicted values, denoted by $\hat{Y}$, for trial 1 based on all three parameters along with the 3 validation points are tabulated in Table 13.

If the errors are linearly distributed, approximately 95% of the standardized residuals should fall in the interval (Montgomery, 1984). Residuals that are far outside this interval may indicate the presence of an outlier, that is, an observation that is not typical of the rest of the data. Various rules have been proposed for discarding outliers. However, outliers sometimes provide important information about unusual circumstances of interest to experimenters and should not be automatically discarded. For further discussion of outliers, see Montgomery *et al.* (2001).

Using equation (6, 7), the standardized residuals can be calculated and the residual plot is displayed below.

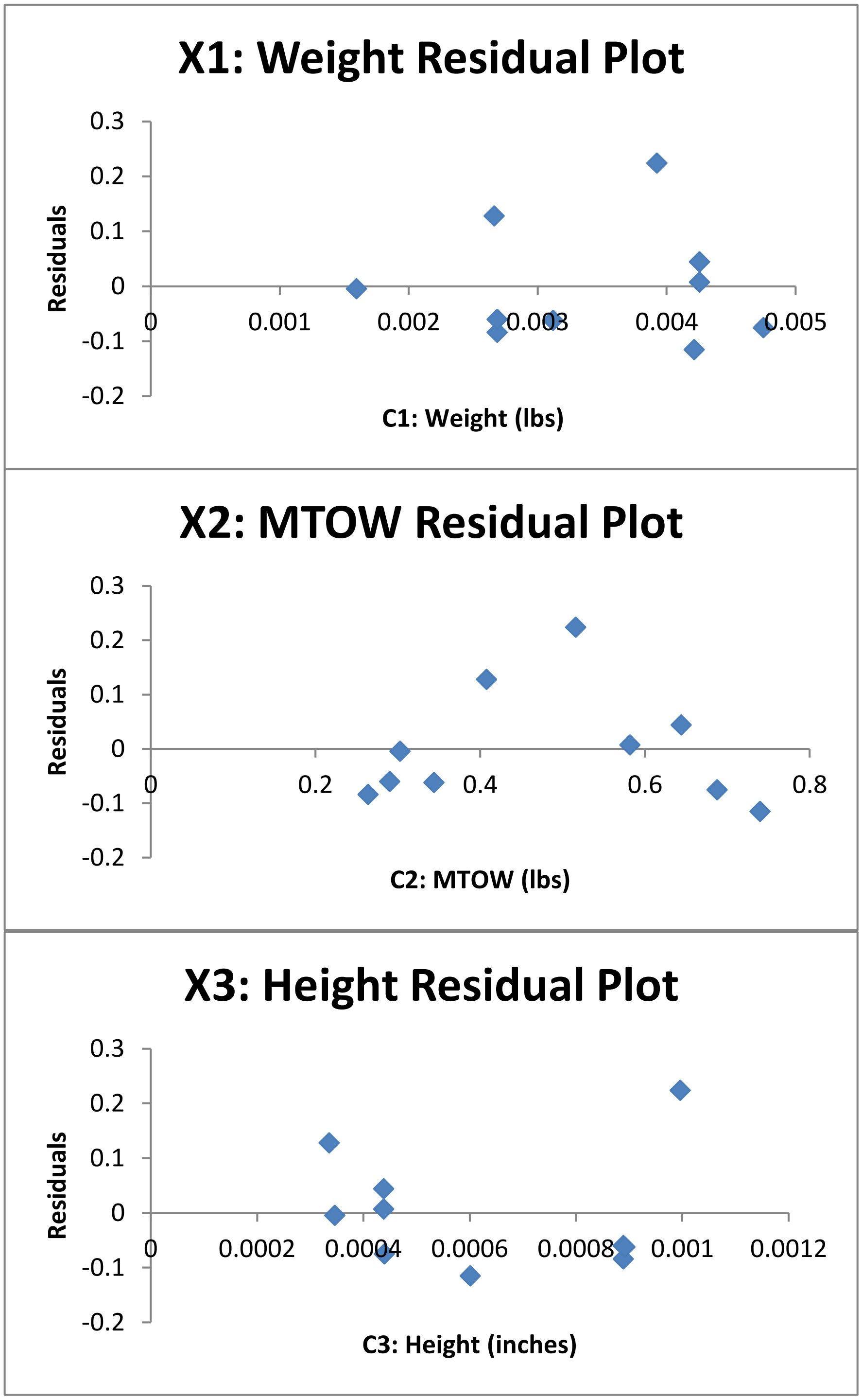

Figure 17: Residual plot of data set

Inspection of the residual plot indicates that there is a strong indication that the points lie scattered randomly around a straight line. Moreover, there is no evidence that the data follows any non linear patterns. Therefore, it is reasonable to assume that the mean of the random variable y is related to x by following a linear relationship.

The second linearity test must demonstrate that there is little or no multicollinearity in the data. As previously mentioned, multicollinearity occurs when two or more predictors in the model are correlated and provide redundant information about the response. In order to demonstrate the presence of multicollinearity in the data, the VIF test is utilized with an R^2 of 0.5925 for the multiple linear regression model.

$$VIF = 1/(1-R^2)$$

$$VIF = 1/(1-0.5925)$$

$$VIF = 2.4540$$

The VIF of the linear regression is equal to 2.4540. A general rule of thumb is that there is a presence of multicollinearity if the VIF > 5. According to Kutner (2004), a cut off VIF of 10 is acceptable. According to our proposed data there is no presence of multicollinearity in the data set. The VIF test results that the data under study is of a linear nature.

The third linearity test is called homoscedasticity. This test requires that the data set exhibits similar amounts of variance across the range of values for an independent variable. Each CER is plotted against the dependant variable in order to fulfill the linearity assumption and are displayed in Figure 18. As can be seen in the in the below figure, the data points are spread relatively scattered across a straight line, yielding similar variances across the data set.

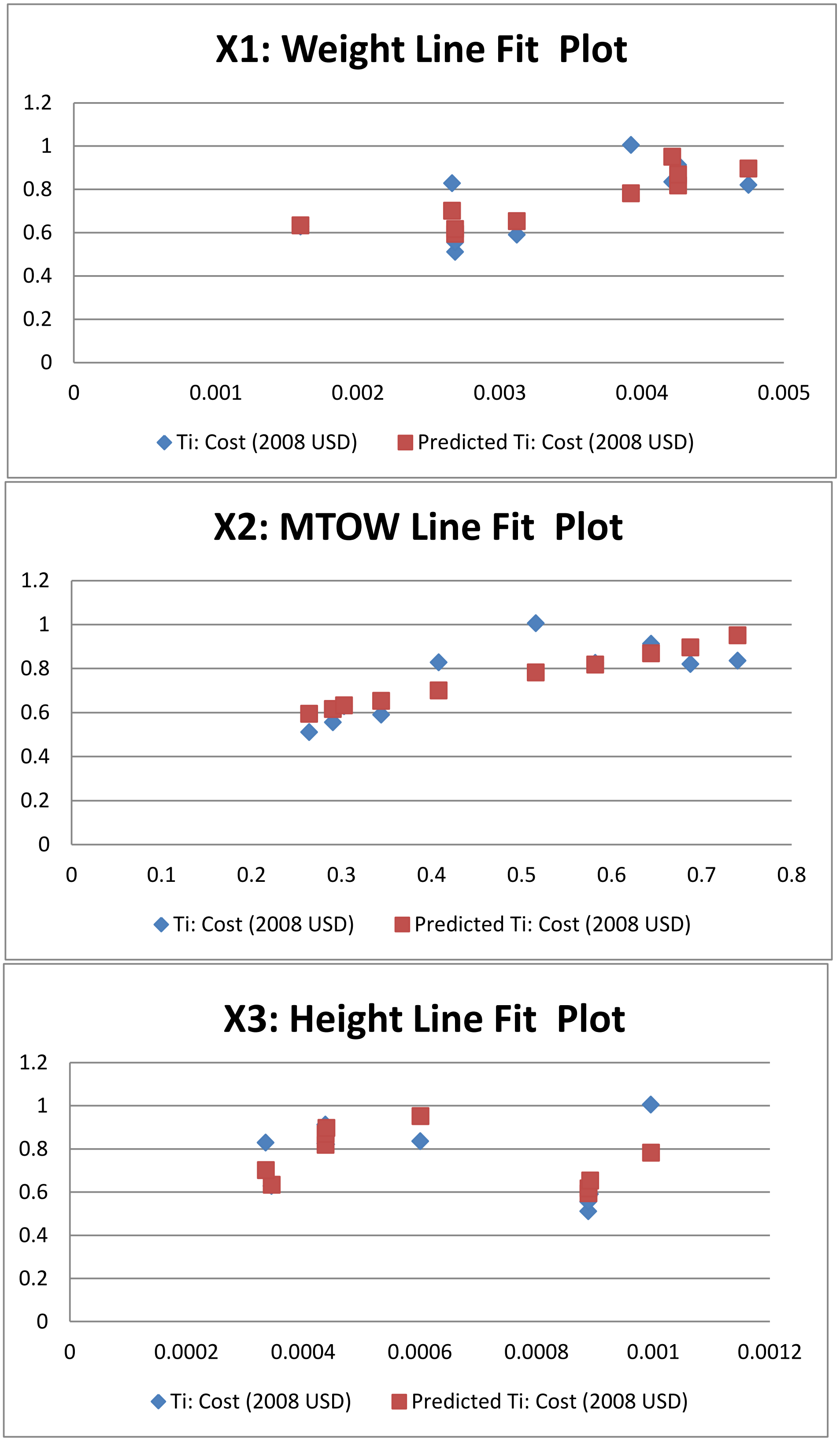

Figure 18: Scattered plot of CER versus dependant variable

In this study, numerous tests have been conducted on the data set of the main landing gear. It has been demonstrated that the data possess the desirable properties, in particular, these assumptions imply that the parameter estimates are unbiased, consistent, and efficient in the case of linear estimators.

4.6 Analysis based on a Non-Linear Model

A similar approach will be derived for the NLM. The comparison of the MLRM versus the NLM is addressed in the discussion section of the thesis. Table 17 represents the natural log of the MLG data gathered for all 3 parameters along with their reverent mean, variance and standard deviation. Similarly to the MLRM, the data was analyzed and their resulting equations along with their coefficient of determination are summarized in Table 18.

Table 17: Historical data for NLM

	Cost Estimating Relationship			
Programs	X1: ln (Weight) (lbs)	X2: ln (MTOW) (lbs)	X3: ln (Height) (in)	Y: ln (Cost) (USD)
1	-5.9196	-1.3318	-7.0251	-0.6723
2	-5.9196	-1.2365	-7.0251	-0.5875
3	-5.7700	-1.0671	-7.0217	-0.5270
4	-5.5405	-0.6616	-6.9111	0.0052
5	-6.4399	-1.1947	-7.9681	-0.4650
6	-6.1540	-0.9655	-7.9679	-0.0590
7	-5.9406	-0.8580	-8.0262	-0.1778
8	-5.9279	-0.8965	-7.9984	-0.1882
9	-5.4594	-0.5413	-7.7324	-0.1919
10	-5.4594	-0.4401	-7.7324	-0.0914
11	-5.3492	-0.3743	-7.7287	-0.1975
12	-5.4689	-0.3011	-7.4163	-0.1801
13	-5.4689	-0.2433	-7.4163	-0.1800

In the NLM, the same parameters were utilized and represented by the variable X_m.

where,

X1, ln (weight of the MLG)

X2, ln (MTOW of the program)

X3, ln (height of the MLG)

Y_i, ln (cost of the MLG)

$\hat{Y}$, ln (predicted cost of the MLG)

e, Difference between the ln (cost) and the ln (predicted cost)

Table 18: NLM resulting equations

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.4422\ X1^{-0.2229}\ X2^{0.6783}\ X3^{-0.0300}$	0.7380
2	$\hat{Y} = 0.4499\ X1^{-0.1743}\ X2^{0.6328}$	0.7368
1	$\hat{Y} = 1.1031\ X2^{0.5067}$	0.7177

For the NLM, the coefficient of determination, R^2 diminishes upon omission of insignificant parameters. Correspondingly to the MLRM, using the ANOVA process of rejecting irrelevant parameters, the height and the weight become invalid CER's due to the large p-value in Table 19 and therefore rendering an equation solely based on MTOW. The generated equations to estimate the target cost using (13, 14) are summarized in Table 18. The ANOVA process generates a coefficient of determination of 0.7177 for the equation based on the MTOW (X2), which according to the international society of parametric analysts is deemed good.

Table 19: P-values for the NLM

	P-value	P-value	P-value
	3 parameters	2 parameters	1 paramater
Intercept	0.5722	0.5477	0.3527
X1: Weight (lbs)	0.5862	0.4998	omit
X2: MTOW (lbs)	0.1012	0.0203	0.0020
X3: Height (inch)	0.8698	omit	omit

Likewise, Table 20 represents the correlation factors between each CER for the NLM which serves as basis for the path analysis procedure. A large correlation value between the ln MTOW and the cost implies the strongest relationship between all parameters. Yet again and observed in the ANOVA process of the MLRM, the highest correlation corresponds to the MTOW and is shown to be 0.8472.

Table 20: Correlation matrix for the NLM

	X1: ln (Weight)	X2: ln (MTOW)	X3: ln (Height)	Y: ln (cost)
X1: ln Weight	1.0000			
X2: ln MTOW	0.8366	1.0000		
X3: ln Height	0.1166	-0.3155	1.0000	
Y: ln (Cost)	0.6332	0.8472	-0.3400	1.0000

By taking the strongest correlation factor and squaring that value, the coefficient of determination can be determined, which is equal to 0.7177. The path coefficients are required for the path analysis which serves as the interactions between each parameter, describing their interconnecting strengths. The path coefficients for the NLR are as follows:

Table 21: Path coefficients for the NLM 3 parameters

Regression Coefficients		σ_i / σ_y	Path Coefficients
B1 =	-0.2229	1.4446	-0.3220
B2 =	0.6783	1.6719	1.1340
B3 =	0.0300	1.8403	0.0553

The path coefficients are determined by multiplying the regression coefficients by the standard deviation of the ratio between the potential cost driver (weight, MTOW, height) and the cost. Below is a table representing the residual values between the actual cost and predicted cost of the non linear parametric formula with all three parameters included.

Table 22: Results based on all three parameters for the NLM

Programs	X1: ln (Weight) (lbs)	X2: ln (MTOW) (lbs)	X3: ln (Height) (in)	Y: ln (Cost) (USD)	Ŷ: ln (Cost) (USD)	*e*
1	-5.9196	-1.3318	-7.0251	-0.6723	-0.6110	0.0613
2	-5.9196	-1.2365	-7.0251	-0.5875	-0.5464	0.0411
3	-5.7700	-1.0671	-7.0217	-0.5270	-0.4647	0.0623
4	-5.5405	-0.6616	-6.9111	0.0052	-0.2375	-0.2428
5	-6.4399	-1.1947	-7.9681	-0.4650	-0.4304	0.0346
8	-5.9279	-0.8965	-7.9984	-0.1882	-0.3431	-0.1549
9	-5.4594	-0.5413	-7.7324	-0.1919	-0.1986	-0.0067
10	-5.4594	-0.4401	-7.7324	-0.0914	-0.1300	-0.0386
11	-5.3492	-0.3743	-7.7287	-0.1975	-0.1098	0.0877
12	-5.4689	-0.3011	-7.4163	-0.1801	-0.0241	0.1560
6	-6.1540	-0.9655	-7.9679	-0.0590	-0.3386	-0.2796
7	-5.9406	-0.8580	-8.0262	-0.1778	-0.3151	-0.1372
13	-5.4689	-0.2433	-7.4163	-0.1800	0.0151	0.1951
mean	-5.7255	-0.8045	-7.4559	-0.3096	-0.3096	0.0000
var	0.1124	0.1506	0.1825	0.0539	0.0398	0.0141
st dev	0.3353	0.3881	0.4271	**0.2321**	0.1994	**0.1188**

The path coefficients shown in the below figure are taken form Table 21 and the uncorrelated value of the function is described in the following calculation of the ratio of the standard deviations for the MLG.

$$P_{ou} = 0.1188 / 0.2321 = 0.5118$$

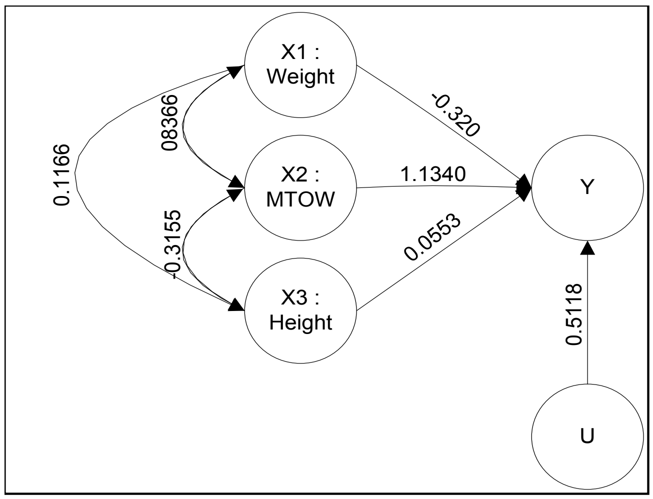

Figure 19: Path analysis diagram for NLM with 3 parameters

Upon discarding insignificant parameters, B2 which represents the MTOW, is the most representative CER for the NLM. Table 23 corresponds to the path coefficient calculation for the MTOW.

Table 23: Path coefficients for the NLM 1 parameters

	Regression Coefficients	σ_i / σ_y	Path Coefficients
B2 =	0.5067	1.6719	0.8472

Table 24 describes the predicted cost for the MLG based upon the MTOW along with their corresponding residuals. The associated uncorrelated value of the function is followed.

Table 24: Results based on the MTOW for the NLM

Programs	X2: ln (MTOW) (lbs)	Y: ln (Cost) (USD)	Ŷ: ln (Cost) (USD)	*e*
1	-1.3318	-0.6723	-0.5768	0.0955
2	-1.2365	-0.5875	-0.5285	0.0590
3	-1.0671	-0.5270	-0.4426	0.0844
4	-0.6616	0.0052	-0.2372	-0.2424
5	-1.1947	-0.4650	-0.5073	-0.0423
8	-0.8965	-0.1882	-0.3562	-0.1679
9	-0.5413	-0.1919	-0.1762	0.0157
10	-0.4401	-0.0914	-0.1249	-0.0335
11	-0.3743	-0.1975	-0.0916	0.1059
12	-0.3011	-0.1801	-0.0545	0.1256
6	-0.9655	-0.0590	-0.3911	-0.3321
7	-0.8580	-0.1778	-0.3367	-0.1589
13	-0.2433	-0.1800	-0.0252	0.1548
mean	-0.8045	-0.3096	-0.3096	0.0000
var	0.1506	0.0539	0.0387	0.0152
st dev	0.3881	**0.2321**	0.1966	**0.1233**

Similarly to the MLRM, the uncorrelated value of the function requires the standard deviation of the residual and the cost of the MLG. It is calculated by equation (21) and can be demonstrated in the below table.

$$P_{ou} = 0.1233 / 0.2321 = 0.5313$$

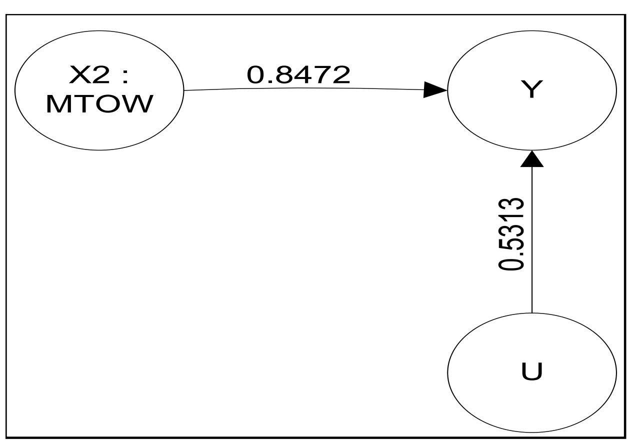

Figure 20: Path analysis diagram for NLM with 1 parameter

In this case, the U represents 51.18% of the uncorrelated value for the path based on three factors. Taken into consideration that this process is repeated until only one factor remains, the uncorrelated value symbolizes 53.13% represented in Figure 20. As seen in equation 14, the logarithm of the linear data was performed to conform to a non linear standard. Taking the exponential of the predicted cost data in Table 20 will bring these cost values back to a non linear form and can therefore be compared to the actual data in order to see the percent errors.

It has been proven, for the NLM, the MTOW is the most appropriate cost estimating relationship to predict the cost of the MLG. The final equation that symbolizes the predicted cost for future MLG platforms is the following:

$$\hat{Y} = 1.1031 * \text{MTOW}^{0.5067}$$

The associated plot of the actual versus predicted values are shown in the following graph. Moreover, the validation programs 6, 7 and 13 used in the research are represented by the red squares in the below figure.

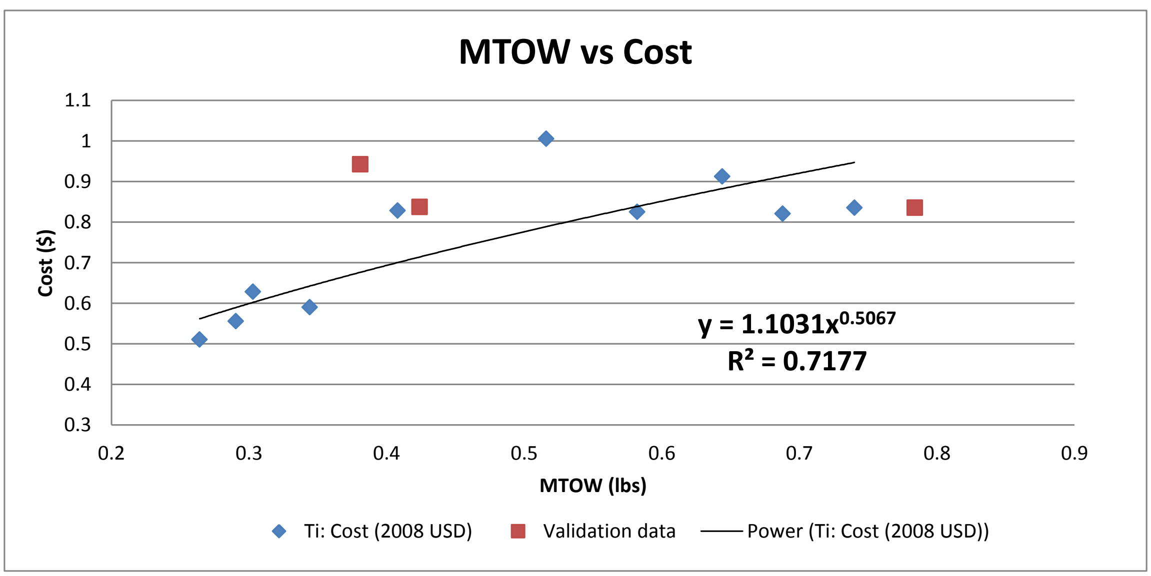

Figure 21: Cost versus the MTOW for the MLG (Non Linear Regression)

The percent error between the actual cost and the predicted cost for the NLM must be calculated in order to comprehend the estimation discrepancy. Moreover, it will be essential to understand which regression analysis has a higher degree of accuracy.

Based on the percent error table below, the non linear model shows an average error of 0.66%, based on 10 programs and an average error of -8.74% for the validation data. The regression R-Square is 0.7177 which based on the international society of parametric analyst is deemed good.

Table 25: Percent errors of the NLM based on the MTOW

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.5105	0.5617	10.0247
2	0.5557	0.5895	6.0788
3	0.5904	0.6423	8.8051
4	1.0053	0.7888	-21.5283
5	0.6281	0.6021	-4.1398
8	0.8284	0.7003	-15.4603
9	0.8254	0.8385	1.5832
10	0.9127	0.8826	-3.2952
11	0.8208	0.9125	11.1748
12	0.8352	0.9470	13.3818
6	0.9427	0.6763	-28.2593
7	0.8371	0.7141	-14.6885
13	0.8353	0.9751	16.7403
Ave. error data			0.6625
Ave. error (6,7,13)			-8.7358

4.7 Summary

The effects of three potential cost drivers and their associated masked costs were compared using the linear and non-linear regression model. The coefficient of determination based solely upon MTOW was determined to be 0.5925 and 0.7177 for the linear and non-linear regression, respectively.

The coefficients of determination presented in Table 9 and Table 18 remained very similar even though parameters such as height and weight have been omitted from the equation. It can be seen in Figure 14 and Figure 15, that for the MLRM, there is a minor increase in the U value when faced with a decrease of three parameters down to one parameter, from 0.6374 to 0.6383. The relative deviation when omitting two parameters (height and weight) can be interpreted as an insignificant change. Similar circumstances can be observed in the NLM, however with a smaller uncorrelated value attributed to the MTOW.

For the path analysis model, minor decreases in the coefficient of determination and minor increases in the uncorrelated value of the path coefficient, will have a relatively small effect in estimating the target cost for the MLG with three parameters. Based on the above information, the MTOW will suffice as the primary CER for the MLG.

For the MLG, both methodologies conclude that the MTOW is the most significant CER to estimate the target cost given the fact that in both cases, the resulting equation using the MTOW generates the best coefficient of determination. Moreover, from a technical point of view it is known that a higher MTOW will consequently be more expensive due to larger forgings, more material requirements, a higher complexity factor, hence rendering an equation solely on MTOW.

The relevant equation that pertains to this case study dealing with MLG's would be as follows:

$$\hat{Y} = 1.1031 * MTOW^{0.5067}$$

Referring to Table 26, the mean error between the predicted cost and the actual cost based solely on the MTOW for the ten trials is 1.84%, with a mean error -8.76% for the validation data of the MLRM. On the other hand, the mean error of the trials versus the validation data for the NLM is 0.66% and -8.74% respectively. Moreover, the following table demonstrates the coefficient of determination for all 3 trials omitting the validation data. The outcome of the validation data in terms of percent errors is relatively similar for both types of regression analysis.

The main finding of this study was that there was a positive relationship between the cost of the main landing gear and the maximum take off weight of the aircraft.

Table 26: Summary of results based on errors

	% errors		% errors		% errors	
	3 Parameters		2 Parameters		1 Parameter	
Programs	MLRM	NLM	MLRM	NLM	MLRM	NLM
1	16.4710	6.3218	16.3474	6.4956	16.9814	10.0247
2	10.8594	4.1985	10.5310	3.9170	10.8890	6.0788
3	10.5491	6.4294	10.3219	6.0836	10.9215	8.8051
4	-22.2682	-21.5558	-22.8370	-22.6381	-22.5394	-21.5283
5	0.7570	3.5226	0.9375	3.3649	-0.4724	-4.1398
8	-15.4058	-14.3499	-14.9357	-13.4322	-15.3914	-15.4603
9	-0.8522	-0.6711	-0.2637	0.2505	0.0990	1.5832
10	-4.8133	-3.7848	-4.5883	-3.3385	-4.5813	-3.2952
11	9.1974	9.1635	9.6163	9.9167	9.9413	11.1748
12	13.8269	16.8786	13.1251	15.5243	12.5443	13.3818
6	-27.0511	-24.3911	-26.9214	-24.2466	-27.7240	-28.2593
7	-14.6854	-12.8226	-14.3118	-12.0234	-14.8908	-14.6885
13	18.0985	21.5390	17.1584	19.8153	16.3288	16.7403
Mean Error data	1.8321	**0.6153**	1.8254	**0.6144**	1.8392	**0.6625**
Mean Error validation	-7.8793	**-5.2249**	-8.0249	**-5.4849**	-8.7620	**-8.7358**
R squared	0.5937	**0.7380**	0.5932	**0.7368**	0.5925	**0.7177**

From the results depicted in bold of the above table, it can be seen that the non linear parametric model outperforms the linear model in each case when estimating the target cost of the MLG for Bombardier Aerospace. According to the ISPA (2011), the coefficient of determination is deemed good for all non linear regression trials. However, the linear regression trials do not conform to the key standard of having a coefficient of determination above 70% as demonstrated in Table 6. That being the case, the NLM approach is deemed to be the more appropriate choice of regression method to use for estimation purpose compared to the NLM. It can be debated that the error analysis are slightly superior in the non linear regression model, however all key attributes are satisfied with the NLM. Thus, it can be concluded that in the case of the main landing gear at Bombardier Aerospace, the NLM is the prediction tool of choice to estimate the target costs. The case study determined a relatively small but statistically significant increase in the technique to estimate the target cost of a main landing gear. This improvement will increase the accuracy of the estimate when estimating the cost of a MLG for a new program. This study concludes that the weight and the height of the main landing gear do not statistically affect the price of the MLG. Both methodologies, analysis of variance and path analysis conclude that the ideal CER to estimate the target cost of the MLG at Bombardier Aerospace is the MTOW, disregarding the height and the weight as potential cost drivers.

In order to get a superior estimate for future applications, a procedure can be taken by excluding the program with the highest percent error and reproducing the entire analysis with 9 programs. Moreover, since 2009, Bombardier Aerospace has signed two new commercial agreements with new programs and renegotiations have taken place with several landing gear programs which were indentified as overpriced. By adding the new data points to the regression procedure, a new parametric equation will be determined for the MLG. The regression will minimize the error of the largest outlier and the coefficient of determination along with the percent error ought to yield a superior estimate.

The findings for this case study at Bombardier Aerospace are conclusive. The analysis points out that the MTOW is the only parameter required to estimate the target cost of the MLG. As represented by the final non linear regression formula, the higher the MTOW of the aircraft, the more the MLG will cost. Below is a high level summary table which depicts the mean error and the coefficient of determination. For each of the three cases, the non linear regression model outperforms the linear regression.

Table 27: Summary of results

	3 Parameters		2 Parameters		1 Parameter	
Regression	MLRM	NLM	MLRM	NLM	MLRM	NLM
Mean Error (%)	1.8321	**0.6153**	1.8254	**0.6144**	1.8392	**0.6625**
R squared	0.5937	**0.7380**	0.5932	**0.7368**	0.5925	**0.7177**

5.0 CONCLUSION AND FUTURE RESEARCH

When purchasing a new product, understanding its target cost is essential, as it will help the company in planning and during negotiations. Target costing focuses on managing cost and profitability in product development (Martin Carlsson-Wall, 2011) during the conceptual design phases, designers and decision makers often need to know accurate cost information to assess and compare multiple alternatives to determine a preferred design. Upper management needs to evaluate cost reduction possibilities and alternatives affecting system performance. Therefore, appropriate cost estimating models must portray accurate and robust cost estimates to support design to cost studies in the early conceptual phase of new programs.

This study focused on target cost models and adapting this knowledge to a case study in the aerospace industry. A case study involving the MLG at Bombardier Aerospace served as proof in demonstrating which type of regression analysis improves the cost estimating accuracy. Even though this empirical study is for a specific application, the methodology utilized can be applied to various industries, as the model has the flexibility of being generalized. The study is concluded in the following sections.

5.1 Major Findings

The main purpose of this research is to explore target costing methodologies and how the concept can be utilized in the early stages of product development. This will guide designers at the early phases of complex products. Moreover, parametric equations, a subset of target costing was analyzed in order to obtain higher predictive accuracy of cost estimation and guide designers at the early design phases of complex products. It is understood that accurately estimating cost is not an easy task at the early stage of complex product development when only a few conceptual attributes of the product are known.

This study shows a comparative parametric modeling technique, between linear and non-linear regression, to estimate the target cost of a major commodity at Bombardier Aerospace. Three parameters were initially considered as potential CER's for the MLG: Height, MTOW and Weight. As per the results shown in section 4, the MLG has a relatively high degree of correlation with the MTOW. The confidentiality of the data was appropriately masked to ensure that the data remains confidential.

In conclusion, the non linear regression model will generate a better accuracy to predict the target cost. This study shows that the overall performance of the NLRM is superior to that of the MLRM based on the following key criterion:

1. Lower percent error for the trial data (10 data points)
2. Lower percent error for the validation data (3 data points)
3. Acceptable coefficient of determination as per the ISPA

It has been proven by both methodologies; ANOVA and path analysis, that the maximum takeoff weight is the most predominant CER factor. This demonstrates that both methodologies will converge on the same findings.

It is important to point out that a cost model cannot reflect the reality to one hundred percent, but the goal is to build a model that is as close a possible. The regression analysis is limited by the scarcity of data points and that the contractual prices were assumed to be fair.

5.2 Limitations

One of the limitations of this thesis is the number of data points collected for the case study. Many other data points could have been considered from other aircraft manufacturers into the regression process, however due to sensitive contractual costs the data points were impossible to collect. Moreover, data obsolescence may be present due to the fact that the ten data points equalling ten programs at BA can take a decade to populate. In essence, a good database of technical and commercial data would lead to a more credible statistical significance and parametric relationship in order to extract the necessary cost drivers. However, it is extremely difficult to obtain a large number of data points in aerospace. The difficulty in accurate cost estimation should not be underestimated.

The current study found that in each case for the 10 programs under study excluding the validation points, the NLM outperforms the MLRM and can be seen in the below table. This is consistent with the work of Karadag *et al.* (2006), who describe that the non linear regression model has a better performance for analyzing experimental data in the field of ammonium exchange compared to the linear regression. This is also demonstrated in the work of Herman and Scherer (2003), where the non linear regression outperforms the linear regression in the first order degradation of pest control substances in soil.

5.3 Enterprise applicability

The development of target costing methods, more precisely parametric equations, can be extremely practical in any type of industry. Parametric equations give the user the practicality of understanding a high level estimate based on historical data. For example, in the supply chain when negotiating a new aircraft part or system, parametric equations give you the ability to determine if the initial commercial bid of a supplier is under or overpriced. This will enable the supply chain to follow with an appropriate commercial strategy. In the advanced design team, parametric equations give the ability to perform trade off studies in the early concept phase in order to determine whether the aircraft concept is profitable. Internal studies at Bombardier on the distribution of the aircraft recurring cost describe that 65% of the cost of the aircraft is associated with 6 major systems or structures.

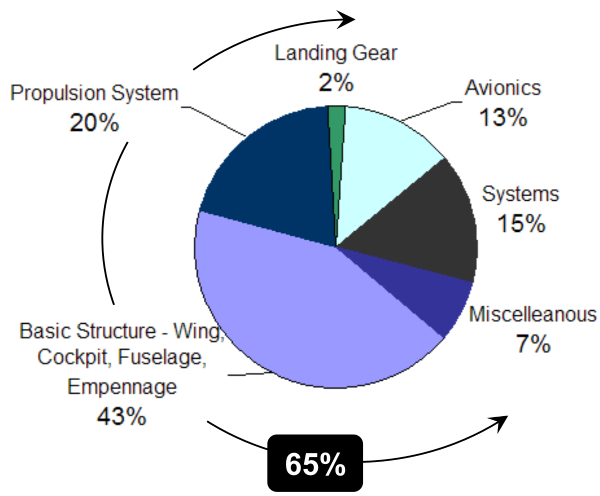

Figure 22: Distribution of aircraft recurring costs

With the success of the case study involving the MLG, BA is currently working on generating cost models for the entire aircraft. As discussed section 4.1, these cost models will complement the MDO tool at which the advanced design team will work on optimizing new products in the early conception phase. The accuracy of the cost models generated for new programs on the recurring cost varies within 15% at BA when comparing the cost models to actual contractual negotiations. Though the concept of target costing can provide substantial benefits in most cases, the aerospace reality can be very different. In the case where aircraft systems depend on software, only a couple of aerospace suppliers are certified and credible to deliver a quality product on time. Therefore, the few software developers that provide avionics suites, electrical systems and satellite systems are scarce and typically defy the target costing methodology. Listed below are several reasons why the application of target costing is not applicable in all aircraft commodities:

1. Software supplier have the power to set the price
2. Difficult to negotiate with software suppliers, lack of transparency
3. Unknown software cost drivers: certification, line of code, functionality
4. Software suppliers want to keep a high industry profitability

Furthermore, political issues often arise in some commodities due to the fact that one supplier holds the monopoly of the business or that companies sign partnership agreements with one another. This entitles the partnership supplier and aircraft manufacturer to have first right to bid before going to the public for quotation, which is named a request for pricing.

Further research pertaining to software cost estimation and other non linear equations can be studied in order to potentially increase the level of accuracy when estimating the target cost.

5.4 Future Research

The application of target costing methodologies in the early conception phase is essential because a large portion of a company's costs are decided interactively with customers and suppliers during product development (Wall, 2011).

There can be several future applications of this thesis. This thesis is limited to comparing linear and non linear regression analysis. However, other cost models can be developed, not limiting itself to a specific commodity, rather the model can be generalized. Moreover, this research can be utilized towards any type of industry, such as the automotive industry, the aerospace industry and many other manufacturers.

In order to seek the possibility of establishing a superior cost estimate, different types of cost models can be utilized to enhance the credibility of a cost estimate, such as: hierarchical regression models described in the literature review and the artificial neural networks (ANN). According to Suh (2005) and shown in the work of Anderson (1995), ANN is a mathematical model that adapts its structure based on the inflow of information through the network during the learning phase. Bayat *et al.* (2007) compared artificial neural networks, linear and non linear regression techniques in order to determine the penetration resistance in soil. ANN was recognized as the most powerful tool to predict the diverse conditions in soil. As mentioned throughout the thesis, it is known that there are numerous available cost estimating tools and techniques. These costing tools are used by different authors, each use a different method and reach results yielding different conclusions. In the literature review section, several costing techniques were presented and the tools used are all useful but sometimes contradictory with other results. For example, Bashir and Thomson (2004) select non linear regression; whereas Kahyani and Basiri (2011) use the linear regression. Moreover, Bayat *et al.* (2007) concludes that neural network cost estimating tool is the most powerful and precise costing tool. Without concretisation of which costing tool is the most powerful, what they do all have in common is the process used to build the model and determine conclusions, such as;

1. Potential cost drivers
2. Commercial data
3. Crossfunctional participation
4. Top management decision tool

Identifying and selecting cost drivers are very important in cost estimation. The ANOVA and path analysis proposed in this study are an effective way to identify the cost drivers of a product. ANOVA can eliminate the insignificant and irrelevant cost drivers under regression. However, most of the data analysis calculation were performed in excel where the analysis was conducted manually. This process was very time consuming and subject to human error. Moreover, the interpretation of the results in terms of p-values and the coefficient of determination are processed by the user and can lead to inconsistent results if conducted by other researchers. Future research could focus on reducing irrelevant cost drivers in less time and automating the regression analysis process using computer software. A preliminary trial using computer software named TableCurve2D and TableCurve3D was analyzed using the data in table 13. This computer software gives the ability to find the ideal costing model for complex data. This software has a built in library of thousands of linear and non linear equations and determines the best fit equation. By fitting the data of the case study the following preliminary conclusion were found. Moreover, the data set was inputted in TableCurve3D and the visual interpretation can be seen below.

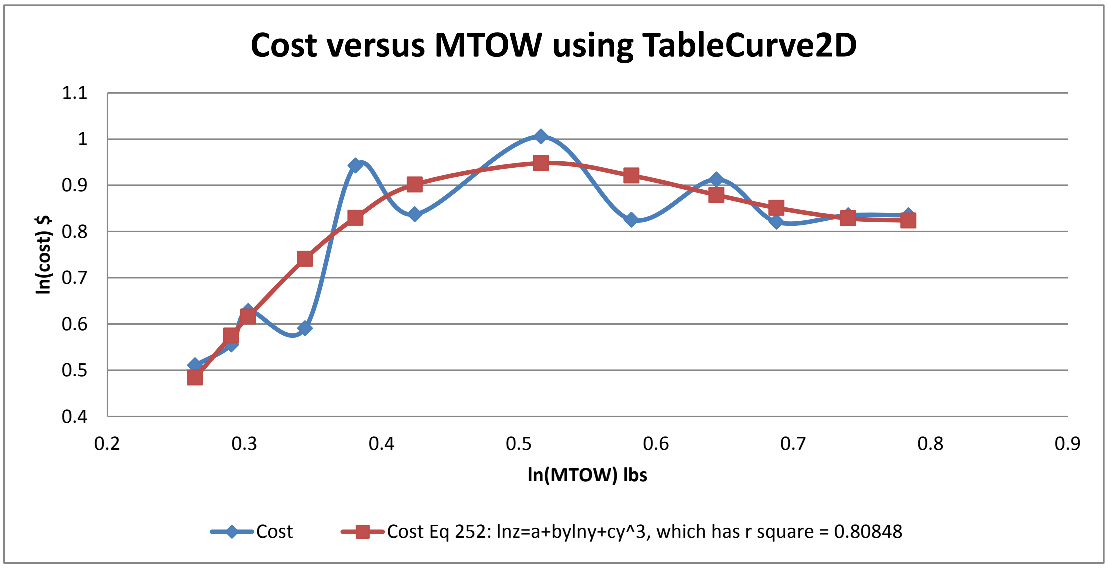

Figure 23: Cost versus MTOW using TableCurve2D

The formula yielded by the computer software yields a superior results based on a coefficients of determination of 80.85%. The parametric formula resulting from this analysis is as follows;

$$\text{Ln cost} = -\ 6.91\ -17.22{*}\text{MTOW} * \ln \text{MTOW}{*}\ 7.12\ \text{MTOW}^{\wedge 3}$$

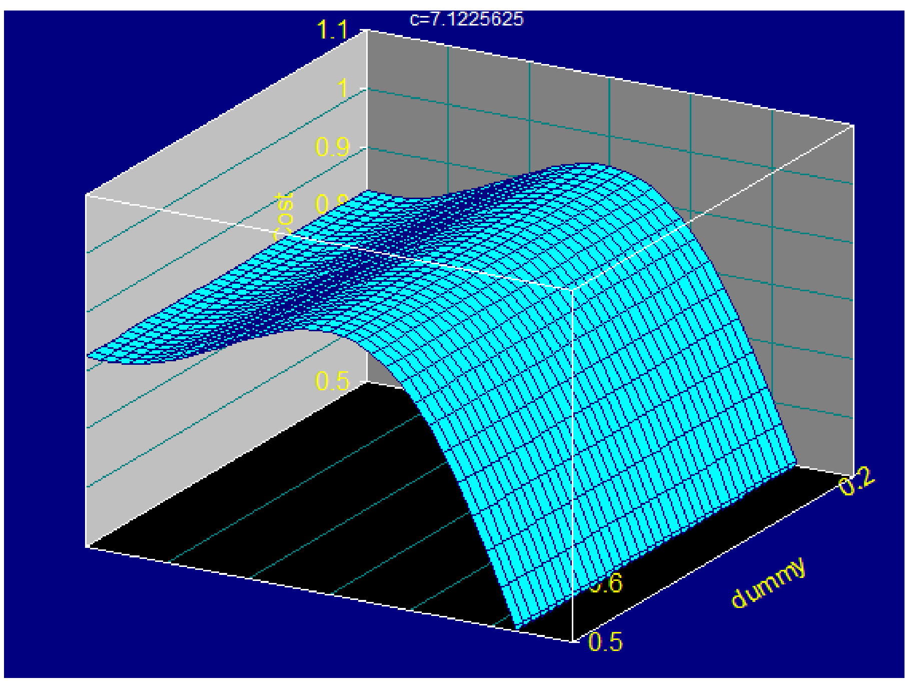

Figure 24: Cost versus MTOW using TableCurve3D

The integration of computer software in cost estimation will be beneficial in accelerating the estimate production, will eliminate human errors and it can be easily integrated in the concurrent engineering approach of product development.

Future research in the field of product development can pertain to an adequate accounting system for new product development. For example, Jörgensen and Messner (2010) developed practice theories in capturing the link between accounting and product development. Moreover, according to Brady and Davies (2004) seek that there is a good platform for further integration with between product development and accounting.

In the end, I believe that when applying target costing methodologies in the conception of a new product, for example and aircraft, with the intension of optimizing the product at the beginning of it's life cycle, that a staggering amount of savings can be perceived. Only a more sophisticated accounting system that captures inefficiencies in the product development cycle will allow organizations to see the benefit of target costing.

6.0 BIBLIOGRAPHY

H. Agndal, and U. Nilsson, "Interorganizational cost management in the exchange process," Management Accounting Research, Vol. 20, No 2, p. 85-101, 2009.

J. A. Anderson, "Introduction to neural networks," Cambridge, MA: MIT Press, p. 2-34, 1995.

S. L. Ansari and J. Bell, CAM-I Target Cost Core Group, Target Costing: The Next Frontier in Strategic Cost Management, Homewood, IL: Irwin, p. 130-135, 1996.

S. L. Ansari and J. Bell, Target costing: the next frontier in strategic cost management, New York: MacGraw-Hill, p. 20-40, 1997.

H. A. Bashir and V. Thomson, "Estimating design effort for GE hydro projects," Computers & Industrial Engineering, vol. 46, no. 2, p. 195-204, 2004.

H. Bayat, M. R. Neyshabouri, M. A. Hajabbasi, A. A. Mahboubi and M. R. Mosaddeghi. "Comparing Neural Networks, Linear and Nonlinear Regression Techniques to Model Penetration Resistance," Turkish journal of Agriculture and Forestry, vol. 32, p. 425-433, 2007

E. A. Bender, An introduction to mathematical modeling, New York: John Wiley & Sons, p. 36-40, 1978.

S. Bengtsson and T. Sjöblom, "A flow cost model – a case study at Volvo corporation," Lulea University of Technology, Gothenburg, p. 8-18, 2006.

N. Bhuiyan, D. Gerwin and V. Thomson, "Simulation of the new product development process for performance improvement," Management Science Vol 50 No 12, p. 1690-1703, 2004.

B. S. Blanchard and W.J. Fabrycky, Systems engineering and analysis, 3rd edition, Prentice-Hall Inc., p, 110-155, 1998.

S. A. Book, M. A. Broder and D. I. Feldman, "Statistical Foundations of Adaptive Cost-Estimating Relationships," Journal of Cost Analysis and Parametrics, Vol 4, No 1, p. 63-90, 2011.

T. Brady, and A. Davies, "Building project capabilities: from exploratory to exploitative learning," Organization Studies, vol 25, no. 9, p. 1601-1621, 2004.

R. J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro, E. Simoudis, "Mining business databases," Communications at the ACM Magazine, Vol. 39, Issue 11 p. 43-48, Nov. 1996.

E. Buckingham, “On physically similar systems; Illustrations of the use of dimensional equations,” Physical Review Vol IV, No. 4, p. 345-376, 1914.

W. Buggert and A. Wielpütz, “Target costing - Grundlagen und Umsetzung des Zielkostenmanagements,” München: Hanser Verlag, p. 5-10, 1995.

J. Burridge, Information preserving statistical obfuscation, Statistics and Computing, p. 321-327, 2003.

T. C. Chamberlin, “The method of multiple working hypotheses,” American Association for the Advancement of Science, vol. 148, New York, p. 754-759, 1965.

N. A. Constantine, “Regression analysis and causal inference: Cause for concern?,” Perspective on Sexual and Reproductive Health, vol. 44, no. 2, p, 134-137, 2012.

R. Cooper and W. B. Chew, “Control tomorrow's costs through today's design,” Harvard Business Review, p. 88-97, 1996.

R. Cooper and R. Slagmulder, “Target costing and value engineering,” Portland: Productivity Press, p. 49-75, 1997.

R. C. Creese and L. T. Moore, “Cost modeling for concurrent engineering,” Cost Engineering, vol. 32, Issue 6, p. 23-27, 1990.

R. Curran, A.K. Kundu, J. M. Wright, S. Crosby, M. Price, S. Raghunathan and E. Benard, “Modeling of aircraft manufacturing cost at the concept stage,” International Journal of Advanced Manufacturing Technology, Vol. 31, p. 407-420, 2006.

F. Dal-Ri, J. Alonso and C. Duarte, “Modeling the Subjectivity in the Target Costing Process: An Experimental Approach Based on the Fuzzy Logic Concepts,” The International Journal of Digital Accounting Research, Vol. 5, No 10, p. 203-222, 2005.

T. Davila, “An empirical study on the drivers of management control systems’ design in new product development,” Accounting, Organizations and Society, vol. 25, no. 4, p. 383-409, 2000.

T. Davila, and M. Wouters, “Designing cost-competitive technology products through cost management,” Accounting Horizons, vol. 18, no. 1, p. 13-26, 2004

Department of Defense, Parametric Estimating Handbook, 2nd edition, http://www.ispa-cost.org/PEIWeb/newbook.htm, 1999.

D. Edgar, “Data sanitization techniques,” A Net 2000 Ltd. White Paper, p. 2-5, 2000.

M. E. F. A. El-Salam, “An efficient estimation procedure for determining ridge regression parameter,”Asian Journal of Mathematics and Statistics, Vol. 4, p. 90-97, 2011.

L. T. Eriksson and F. Wiedersheim-Paul, “Att utreda, forska och rapportera (5th Ed),” Stockholm. Liber Ekonomi, 2001.

W. J. Fabrycky and B. S. Blanchard, “Life-Cycle Cost and Economic Analysis,” Prentice Hall, Englewood Cliffs, New Jersey, 1991.

J. J. Faraway, Practical regression and Anova using R, p. 123-130, 2000.

E. Fernanndez-Juricic, J. T. Erichsen, and A. Kacelnik, “Visual perception and social foraging in birds,” Trends in Ecology and Evolution, Vol. 19, p. 25–31, 2003.

N. Gohin, Superjet 100 Landing Gear, http://www.safran-na.com/spip.php?rubrique44, 2010.

F. A. Graybill, Theory and application of the linear model, MA: Duxbury Press, 1976.

C. G Hart, Z. He, R. Sbragio and N. Vlahopoulos, “An Advanced Cost Estimation Methodology for Engineering Systems,” Systems Engineering Vol. 15, Issues 1, p. 28-40, 2012

D. R. Helsel and R. M. Hirsch, Statistical Methods in Water Resources, USGS Techniques of Water Resources Investigations, Book 4, Chapter A3, p. 510-525, 2002.

R. A. Herman and P. N. Scherer, “Comparison of Linear and Nonlinear Regression for Modeling the First-Order Degradation of Pest-Control Substances in Soil,” Journal of Agricultural and Food Chemistry, Vol 51. No. 16, P. 4722-4726, 2003.

H. E. Huntley, Dimensional Analysis, Dover Publications, New York, p. 27-35, 1967.

International Society of Parametric Analysts, Parametric Estimating Handbook, 4th edition, ISPA/SCEA Joint Office, 2008.

R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis, NJ: Prentice Hall, 1992.

B. Jörgensen, and M. Messner, ”Management control in new product development: The dynamics of managing flexibility and efficiency,” Journal of Management Accounting Research, vol. 21, no. 1, p. 99-124, 2009

S. Kahyani and R. Basiri, “The Basic of Analytical of Simple Linear Regression in Forestry Studies Case Study: Relationship between Basal Area and Tree Coverage of

Quercus Brantii Lindl. in Absardeh, Chahar Mahale and Bakhtiari," International Journal of Applied Environmental Sciences, Vol. 6, Issue 1, p. 35-45, 2011.

R. S. Kaplan and A. A. Atkinson, Advanced management accounting, 3rd edition, Upper Saddle River: Prentice Hall, p. 20-35, 1998.

D. Karadag, Y. Koc, M. Turan and M. Ozturk, "A comparative study of linear and non-linear regression analysis for ammonium exchange by clinoptilolite zeolite," Journal of Hazardous Materials, vol 144, p. 432-437, 2006.

K. H. Kim and P. M. Bentler, "Tests of homogeneity of means and covariance matrices for multivariate incomplete data," Psychometrika Vol. 67, No. 4, p. 609-624, 2002.

M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, Applied Linear Statistical Models, New York: McGraw-Hill, p. 120-155, 2004.

H. L. Langhaar, Dimensional Analysis and the Theory of Models, New York: John Wiley & Sons, P. 1, 1951.

K. C. Land, "Principles of path analysis," San Francisco: Jossey-Bass, p. 3-37, 1969.

T. Leahy, "The target costing bull's eye," http://www.controllermag.com/issues/1998/January/targetbulleye.html, 1998.

C. C. Li, Path Analysis - A Primer, Boxwood Press, p. 110-131, 1975.

W. J. Marx, D. N. Mavris and D. P. Schrage, "A hierarchical aircraft life cycle cost analysis model," 1st AIAA Aircraft Engineering, Technology, and Operations Congress, Anaheim, CA, p.1-18, 1995.

B. H. Maskell and B. L Baggaley, "Lean Accounting: What's it all About?," Association for Manufacturing Excellence, Target Magazine Vol. 1, p. 2-10, 2006.

Y. Monden, and K. Hamada, "Target costing and kaizen costing in Japanese automobile companies," Journal of Management Accounting Research, vol. 3, p. 16-34, 1991.

D. C. Montgomery, Design and Analysis of Experiments (2nd ed.). New York: John Wiley & Sons, Inc. 1984

D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to linear regression analysis, 3rd edition, New York: John Wiley & Sons, p. 641-645, 2001.

D. C. Montgomery, Introduction to Statistical Quality Control, 4th edition, New York: John Wiley & Sons, p. 201-253, 2001.

D. C. Montgomery, and G. C. Runger, Applied statistics and probability for engineers, 4th edition, New York: John Wiley & Sons, p.386-436, 2007.

J. Mouritsen, A. Hansen, and C. Hansen, “Short and long translation management accounting calculations and innovation management,” Accounting, Organization and Society, vol. 34, no. 6, p. 738- 754, 2009.

O. E. Okereke, “Effect of transformation on the parameter estimates of a simple linear regression model: A case study of division of variables by constants,” Asian Journal of Mathematics and Statistics, Vol. 4, No, 4, p. 174-180, 2011.

L. Östman, “Towards a general theory of financial control for organisations,” Stockholm, The Economic Research Institute, p. 45-52, 2009

R. A. Parsa, K. Muralidhar and R. Sarathy, “Conditions for maximum security in linear data perturbation,” ASA Proceedings of the Section on Government Statistics and Section on Social Statistics, p. 70-72, 1998.

L. Pazmany, “Landing gear design for light aircraft,” 1st edition, San Diego: Pazmany Aircraft Corp, p. 40-65, May 1986.

E. J. Pedhazur, “Analytic methods in studios of educational effects,” Review of Research in Education, Vol. 3, p. 243-286, 1975.

A. Rajarathinam and R. S. Parmar, “Application of parametric and nonparametric regression models for area, production and productivity trends of castor crop,” Asian Applied Sciences, vol. 4, p. 42-52, 2011.

F. Rösler, “Target costing für die Automobilindustrie,” Wiesbaden: Dt. Univ.-Verl., p. 35-40, 1996.

P. Routio, “Developing an industrial product,” http://www2.uiah.fi/projekti/metodi /130.htm, 2007.

R. Roy, S. Forsberg, S. Kelvesjo, and C. Rush, “Risk analysis of parametric cost estimates within a concurrent engineering environment,” 7th ISPE International Conference on Concurrent Engineering: Research and Applications, Lyon, France, Technomic Inc., Pennsylvania USA, p. 23-30, 2000.

R. Roy, S. Forsberg, S. Kelvesjo, and C. Rush, “Quantitative and qualitative cost estimating for engineering design,” Journal of Engineering Design, vol. 12, p. 147-162, 2001.

S. M. Sajadifar and M. Allameh, “Some modifications to calculate regression coefficients in multiple linear regression,” Journal of Industrail and Systems Engineering, vol. 2, No1, p. 75-86, spring 2008.

A. Salam, N. Bhuiyan, G. J. Gouw, S. A. Raza, Estimating design effort for the compressor design department: A case study at Pratt & Whitney Canada, Design Studies, vol. 30 No. 3, p. 303-319, 2009.

M. Salehi and M Hematfar, “Comparing linear and non-linear relationships between accounting variables and dividend,” African Journal of Business Management, vol. 6, p. 2143-2151, 2011

A. A. Sani and M. Allahverdizadeh, “Target and Kaizen Costing,” World Academy of Science, Engineering and Technology, No. 62, p. 45-53, 2012,

R. Sarathy, K. Muralidhar and R. Parsa “Perturbing non-normal confidential attributes: The copula approach,” Management Science, vol. 48, no. 12, p. 1613-1627, 2002.

N. L. Smith and S. L. Murray, “The use of path analysis in program evaluation,” Research, evaluation and development paper series, p. 3-7, 1978.

J. Sudhir, Managerial Economics, Pearson Education, 2009.

W. Suh, “Web engineering: principles and techniques,” Idea Group Publishing, pp. 335-338, 2005.

D. Swenson, S. Ansari, J. Bell and I. W. Kim, “Best practices in target costing,” Management Accounting Quarterly, Vol 4, No. 2, p, 12-17, 2003.

T. Tanaka, “Target costing at Toyota,” Journal of Cost Management, Vol. 7, No. 1, p. 4-11, spring 1993.

D. Ting, S.E. Fienberg, and M. Trottini, Methodology for microdata release, Monographon Official Statistics – Work session on statistical data confidentiality, Geneva, p. 89-98, 2005.

D. Trump, “Think exist quotations, http://thinkexist.com/quotes/donald_trump/2.html,” 2010.

K. T. Ulrich and S. D. Eppinger, Product design and development, McGraw-Hill, 5th edition, p. 84-96, 2011.

G. A. Vignaux and J. L. Scott, “Theory & Methods: Simplifying regression model using dimensional analysis,” Australia & New Zealand Journal of Statistics, vol. 41, p. 31-41, 1999.

M. C. Wall, “Targeting target cost: cost management and inter-organizational product developmentof muti-technologyproducts,” Stockholm School of Economics, Sweden, 2011

C. E. Walston and C. P. Felix, “A method of programming measurement and estimation,” IBM System Journal, vol. 16, no. 1, p. 54-73, 1977.

S. Weisberg, Applied regression analysis, 3rd edition, New York: John Wiley & Sons Inc, p. 375-425, 2005.

J. P. Womack and D. T. Jones, “Lean Thinking – Banish waste and create wealth in your corporation; Free Press, New York, USA, 2003.

S. Wright, “The method of path coefficients,” Annals of Mathematical Statistics, vol. 5, no. 3, p. 161-215, 1934.

Y. Zengin and E. Ada, “Cost management through product design: target costing approach,” International Journal of Production Research, Vol. 48, No. 19, October 2010, p. 5593-5611, 2009.

7.0 DEFINITIONS

Adjusted R^2:	R^2 adjusted for the number of X-variables used to explain the variation in the Y-data
Coefficient of Determination:	Percent of the variation in the Y-data explained by the X-data.
Cost estimating relationship:	A mathematical expression of varying degrees of complexity expressing cost as a function of a cost driving variable
Coefficient of variation:	Measure of dispersion; produces a measure of average estimating errors. SE divided by mean of the Y-data, relative measure of estimating error
Degrees of Freedom:	Number of observations (N) less the number of estimated parameters (# of X-variables + 1 for the constant term "a"). Concept of parsimony applies in that a preferred model is one with high statistical significance using the least number of variables.

Dimensional analysis:	DA is a method employed to restructure the variables of a regression model into a set of independent dimensionless products
F-test:	Tests for trend in the data versus random dispersion. Tests whether the entire equation, as a whole, is valid.
Outliers:	Y observations that the model predicts poorly. This is not always a valid reason to discard the data.
P-value:	Probability level at which the statistical test would fail, suggesting the relationship is not valid. P-values less than 0.10 are generally preferred (i.e., only a 10% chance, or less, that the model is no good).
Residual	The deviation from the true value of a function to a sample set
R-squared:	Measures the percentage of variation in the pool explained by the CER or model; varies between 0% and 100%.
Standard Error (SE):	Average estimating error when using the equation as the estimating rule

Target Costing	The maximum amount of cost that can be incurred on a product that a company can earn a predetermined profit margin.
T-test:	Measures the significance of the individual components of the model; where there is only one independent variable (one 'base' variable), the significances of the t-test and of the F-test are identical. Tests whether the individual X-variable(s) is/are valid.
Value Engineering	Value Engineering is the systematic evaluation of all aspects of the value-chain business functions, with the objective of reducing costs while satisfying customers' needs."SANI"

8.0 APPENDIX A: DATA ANALYSIS ON PROGRAM OMIT 3, 7, 9

Table 28: MLRM resulting equations (Omit 3,7,9)

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.6145 + 2.7584X1 + 0.4589X2 - 112.9299X3$	0.3742
2	$\hat{Y} = 0.6145 + 0.4729X2 - 109.2529X3$	0.3741
1	$\hat{Y} = 0.5393 + 0.4945X2$	0.3474

Table 29: Path coefficients for the MLRM 3 parameters (Omit 3, 7, 9)

	Regression Coefficients	σ_{i} / σ_{y}	Path Coefficients
B1 =	2.7584	0.0065	0.0179
B2 =	0.4589	1.1920	0.5470
B3 =	-112.9299	0.0015	-0.1709

Table 30: Uncorrelated residual for MLRM with 3 parameters (Omit 3, 7, 9)

Uncorrelated residual 3 parameters	
r(y,U)= sigma e / sigma y	0.7911

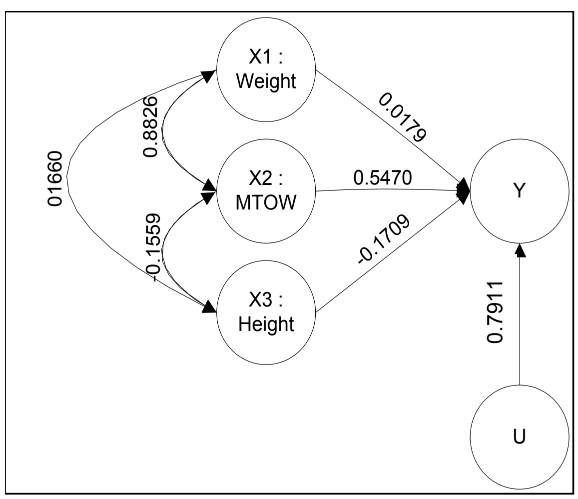

Figure 25: Path analysis diagram for MLRM with 3 parameters (Omit 3, 7, 9)

Table 31: Results based on the MTOW for the MLRM (Omit 3, 7, 9)

Programs	X2: MTOW (lbs)	Y: Cost (USD)	Ŷ: Cost (USD)	*e*
1	0.2640	0.5105	0.6699	-0.1593
2	0.2904	0.5557	0.6829	-0.1272
4	0.5160	1.0053	0.7945	0.2108
5	0.3028	0.6281	0.6891	-0.0609
6	0.3808	0.9427	0.7276	0.2150
8	0.4080	0.8284	0.7411	0.0873
10	0.6440	0.9127	0.8578	0.0549
11	0.6878	0.8208	0.8794	-0.0587
12	0.7400	0.8352	0.9053	-0.0701
13	0.7840	0.8353	0.9270	-0.0918
3	0.3440	0.5904	0.7094	-0.1191
7	0.4240	0.8371	0.7490	0.0881
9	0.5820	0.8254	0.8271	-0.0017
mean	0.5018	0.7875	0.7875	0.0000
var	0.0395	0.0278	0.0097	0.0182
st dev	0.1988	**0.1668**	0.0983	**0.1347**

Table 32: Path coefficients for the MLRM 1 parameter (Omit 3, 7, 9)

	Regression Coefficient	σ_i / σ_y	Path Coefficient
B2 =	0.4945	1.1920	0.5894

Table 33: Uncorrelated residual for MLRM with 1 parameter (Omit 3, 7, 9)

Uncorrelated residual 1 parameter	
r(y,U)= sigma e / sigma y	0.8078

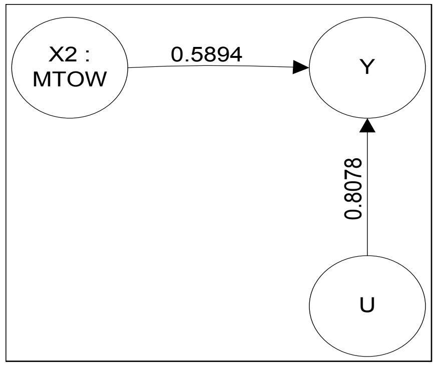

Figure 26: Path analysis diagram for MLRM with 1 parameter (Omit 3, 7, 9)

Table 34: Percent errors of the MLRM based on the MTOW (Omit 3, 7, 9)

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.2640	0.6699	31.2124
2	0.2904	0.6829	22.8912
4	0.5160	0.7945	-20.9660
5	0.3028	0.6891	9.7011
6	0.3808	0.7276	-22.8114
8	0.4080	0.7411	-10.5418
10	0.6440	0.8578	-6.0115
11	0.6878	0.8794	7.1485
12	0.7400	0.9053	8.3891
13	0.7840	0.9270	10.9858
3	0.3440	0.7094	20.1719
7	0.4240	0.7490	-10.5225
9	0.5820	0.8271	0.2118
Ave. error data			2.9997
Ave. error (3,7,9)			3.2871

Table 35: NLM resulting equations (Omit 3, 7, 9)

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.2487\ X1^{-0.1729}\ X2^{0.5161}\ X3^{-0.0702}$	0.5720
2	$\hat{Y} = 0.2360\ X1^{-0.2854}\ X2^{0.6054}$	0.5649
1	$\hat{Y} = 1.0427\ X2^{0.3968}$	0.5056

Table 36: Path coefficients for the NLM 3 parameters (Omit 3, 7, 9)

Regression Coefficients		σ_i / σ_y	Path Coefficients
B1 =	-0.1729	1.5629	-0.2702
B2 =	0.5161	1.7918	0.9248
B3 =	-0.0702	1.8259	-0.1282

Table 37: Uncorrelated residual for NLM with 3 parameters (Omit 3, 7, 9)

Uncorrelated residual 3 parameters	
r(y,U)= sigma e / sigma y	0.6542

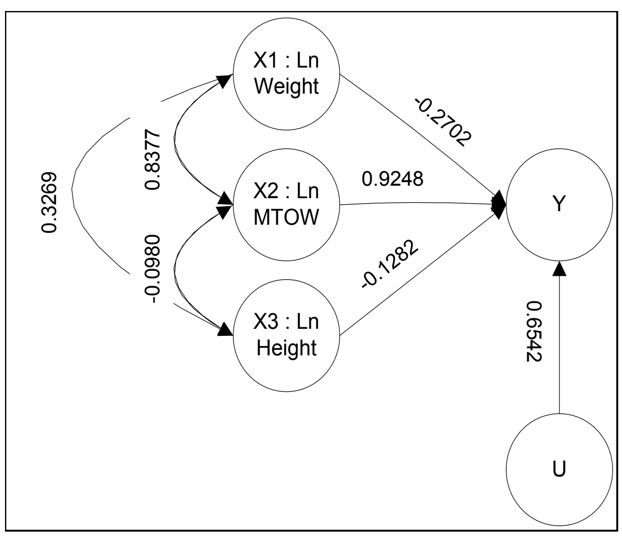

Figure 27: Path analysis diagram for NLM with 3 parameters (Omit 3, 7, 9)

Table 38: Results based on the MTOW for the NLM (Omit 3, 7, 9)

Programs	X2: Ln MTOW (lbs)	Y: Ln Cost (USD)	Ŷ: Ln Cost (USD)	*e*
1	-1.3318	-0.6723	-0.4867	0.1856
2	-1.2365	-0.5875	-0.4489	0.1386
4	-0.6616	0.0052	-0.2208	-0.2260
5	-1.1947	-0.4650	-0.4323	0.0327
6	-0.9655	-0.0590	-0.3413	-0.2823
8	-0.8965	-0.1882	-0.3139	-0.1257
10	-0.4401	-0.0914	-0.1328	-0.0414
11	-0.3743	-0.1975	-0.1067	0.0908
12	-0.3011	-0.1801	-0.0777	0.1024
13	-0.2433	-0.1800	-0.0548	0.1253
3	-1.0671	-0.5270	-0.3817	0.1454
7	-0.8580	-0.1778	-0.2987	-0.1208
9	-0.5413	-0.1919	-0.1730	0.0189
mean	-0.7645	-0.2616	-0.2616	0.0000
var	0.1713	0.0534	0.0270	0.0264
st dev	0.4139	**0.2310**	0.1643	**0.1624**

Table 39: Path coefficients for the NLM 1 parameter (Omit 3, 7, 9)

	Regression Coefficient	σ_i / σ_y	Path Coefficient
B2 =	0.3968	1.7918	0.7111

Table 40: Uncorrelated residual for LRM with 1 parameter (Omit 3, 7, 9)

Uncorrelated residual 1 parameter	
r(y,U)= sigma e / sigma y	0.7031

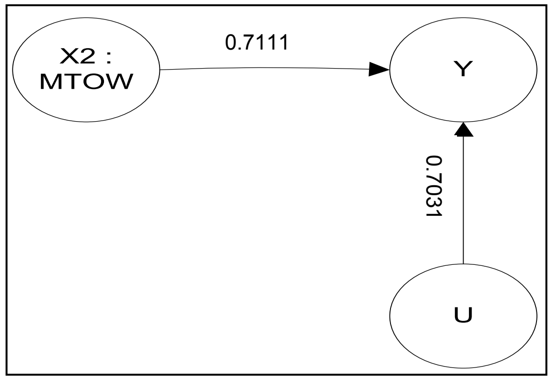

Figure 28: Path analysis diagram for NLM with 1 parameter (Omit 3, 7, 9)

Table 41: Percent errors of the NLM based on the MTOW (Omit 3, 7, 9)

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.5105	0.6147	20.3954
2	0.5557	0.6383	14.8682
4	1.0053	0.8019	-20.2278
5	0.6281	0.6490	3.3270
6	0.9427	0.7108	-24.5945
8	0.8284	0.7306	-11.8127
10	0.9127	0.8756	-4.0574
11	0.8208	0.8988	9.5046
12	0.8352	0.9253	10.7836
13	0.8353	0.9467	13.3435
3	0.5904	0.6827	15.6477
7	0.8371	0.7418	-11.3830
9	0.8254	0.8411	1.9098
Ave. error data			1.1530
Ave. error (3,7,9)			2.0582

Table 42: Summary of results based on errors (Omit 3, 7, 9)

	% errors		% errors		% errors	
	3 Parameters		2 Parameters		1 Parameter	
Programs	MLRM	NLM	MLRM	NLM	MLRM	NLM
1	25.8791	11.6312	25.7831	11.7970	31.2124	20.3954
2	17.8225	7.7244	17.8007	8.8055	22.8912	14.8682
4	-25.4333	-25.5595	-25.4323	-23.5514	-20.9660	-20.2278
5	14.4297	13.8422	14.5986	14.5302	9.7011	3.3270
6	-19.8015	-18.7356	-19.7279	-19.1947	-22.8114	-24.5945
8	-6.9124	-7.6507	-6.9669	-10.1179	-10.5418	-11.8127
10	-4.4237	-3.9706	-4.5516	-5.9084	-6.0115	-4.0574
11	8.8689	8.3538	8.6352	5.5028	7.1485	9.5046
12	7.4982	10.4448	7.6040	12.1438	8.3891	10.7836
13	9.9074	13.7783	10.0868	16.1253	10.9858	13.3435
3	15.2217	7.8174	15.1274	8.7380	20.1719	15.6477
7	-6.8857	-6.3872	-6.9058	-8.6230	-10.5225	-11.3830
9	2.2349	0.7776	1.9884	-2.1437	0.2118	1.9098
Mean Error data	2.7835	**0.9858**	2.7830	**1.0132**	2.9997	**1.1530**
Mean Error validation	3.5236	**0.7359**	3.4033	**-0.6762**	3.2871	**2.0582**
R squared	0.3742	**0.5720**	0.3741	**0.5649**	0.3474	**0.5056**

9.0 APPENDIX B: DATA ANALYSIS ON PROGRAM OMIT 2, 8, 12

Table 43: MLRM resulting equations (Omit 2, 8, 12)

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.5892 - 36.8291X1 + 0.7109X2 - 44.1148X3$	0.3700
2	$\hat{Y} = 0.5686 - 49.8480X1 + 0.7901X2$	0.3672
1	$\hat{Y} = 0.5287 + 0.5316X2$	0.3388

Table 44: Path coefficients for the MLRM 3 parameters (Omit 2, 8, 12)

Regression Coefficients		σ_i / σ_y	Path Coefficients
B1 =	-36.8231	0.0066	-0.2433
B2 =	0.7109	1.0948	0.7783
B3 =	-44.1148	0.0016	-0.0704

Table 45: Uncorrelated residual for MLRM with 3 parameters (Omit 2, 8, 12)

Uncorrelated residual 3 parameters	
r(y,U)= sigma e / sigma y	0.7937

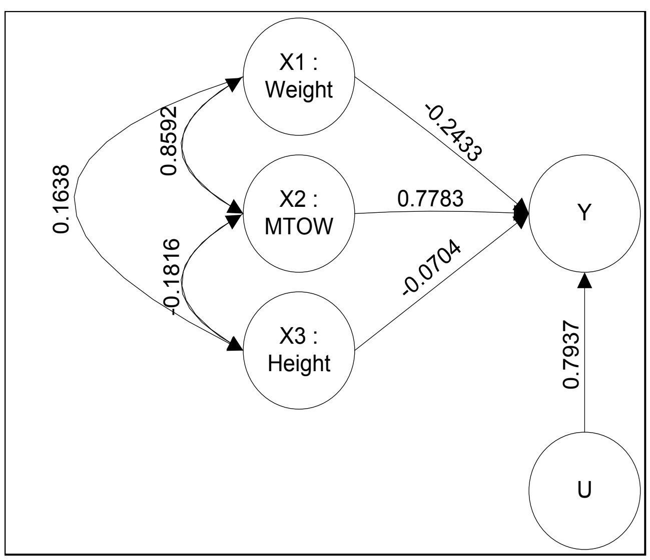

Figure 29: Path analysis diagram for MLRM with 3 parameters (Omit 2, 8, 12)

Table 46: Results based on the MTOW for the MLRM (Omit 2, 8, 12)

Programs	X2: MTOW (lbs)	Y: Cost (USD)	Ŷ: Cost (USD)	*e*
1	0.2640	0.5105	0.6691	-0.1586
3	0.3440	0.5904	0.7116	-0.1213
4	0.5160	1.0053	0.8031	0.2022
5	0.3028	0.6281	0.6897	-0.0616
6	0.3808	0.9427	0.7312	0.2115
7	0.4240	0.8371	0.7542	0.0829
9	0.5820	0.8254	0.8382	-0.0128
10	0.6440	0.9127	0.8711	0.0415
11	0.6878	0.8208	0.8944	-0.0736
13	0.7840	0.8353	0.9456	-0.1103
2	0.2904	0.5557	0.6831	-0.1274
8	0.4080	0.8284	0.7457	0.0828
12	0.7400	0.8352	0.9222	-0.0870
mean	0.4929	0.7908	0.7908	0.0000
var	0.0313	0.0261	0.0089	0.0173
st dev	0.1770	**0.1617**	0.0941	**0.1315**

Table 47: Path coefficients for the MLRM 1 parameter (Omit 2, 8, 12)

	Regression Coefficient	σ_{i} / σ_{y}	Path Coefficient
B2 =	0.5316	1.0948	0.5820

Table 48: Uncorrelated residual for MLRM with 1 parameter (Omit 2, 8, 12)

Uncorrelated residual 1 parameter	
r(y,U)= sigma e / sigma y	0.8132

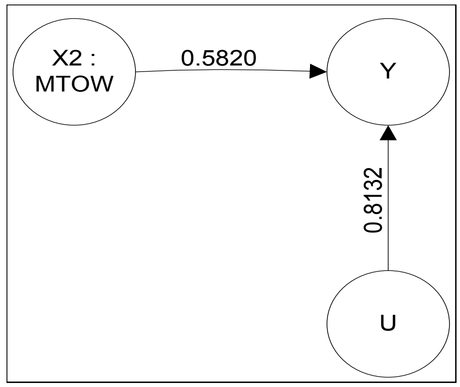

Figure 30: Path analysis diagram for MLRM with 1 parameter (Omit 2, 8, 12)

Table 49: Percent errors of the MLRM based on the MTOW (Omit 2, 8, 12)

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.5105	0.6691	31.0595
3	0.5904	0.7116	20.5428
4	1.0053	0.8031	-20.1129
5	0.6281	0.6897	9.8061
6	0.9427	0.7312	-22.4342
7	0.8371	0.7542	-9.9060
9	0.8254	0.8382	1.5477
10	0.9127	0.8711	-4.5511
11	0.8208	0.8944	8.9704
13	0.8353	0.9456	13.2038
2	0.5557	0.6831	22.9271
8	0.8284	0.7457	-9.9906
12	0.8352	0.9222	10.4117
Ave. error data			2.8126
Ave. error (2,8,12)			7.7827

Table 50: NLM resulting equations (Omit 2, 8, 12)

# of Parameters	Resulting Equations	R^2
3	$\hat{Y} = 0.2110\ X1^{-0.3025}\ X2^{0.6577}\ X3^{-0.0087}$	0.5741
2	$\hat{Y} = 0.2104\ X1^{-0.3160}\ X2^{0.6697}$	0.5740
1	$\hat{Y} = 1.0682\ X2^{0.4189}$	0.4890

Table 51: Path coefficients for the NLM 3 parameters (Omit 2, 8, 12)

Regression Coefficients		σ_i / σ_y	Path Coefficients
B1 =	-0.3025	1.6143	-0.4883
B2 =	0.6577	1.6691	1.0978
B3 =	-0.0087	1.9308	-0.0168

Table 52: Uncorrelated residual for NLM with 3 parameters (Omit 2, 8, 12)

Uncorrelated residual 3 parameters	
r(y,U)= sigma e / sigma y	0.6526

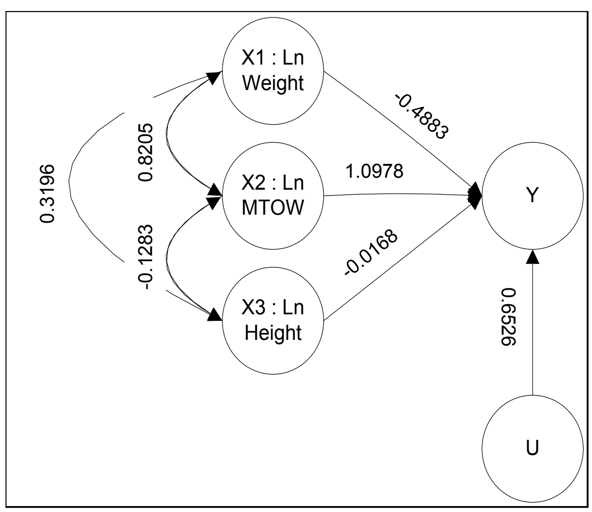

Figure 31: Path analysis diagram for NLM with 3 parameters (Omit 2, 8, 12)

Table 53: Results based on the MTOW for the NLM (Omit 2, 8, 12)

Programs	X2: ln (MTOW) (lbs)	Y: ln (Cost) (USD)	Ŷ: ln (Cost) (USD)	*e*
1	-1.3318	-0.6723	-0.4920	0.1803
3	-1.0671	-0.5270	-0.3811	0.1460
4	-0.6616	0.0052	-0.2112	-0.2165
5	-1.1947	-0.4650	-0.4345	0.0305
6	-0.9655	-0.0590	-0.3385	-0.2795
7	-0.8580	-0.1778	-0.2935	-0.1157
9	-0.5413	-0.1919	-0.1608	0.0311
10	-0.4401	-0.0914	-0.1184	-0.0270
11	-0.3743	-0.1975	-0.0908	0.1067
13	-0.2433	-0.1800	-0.0360	0.1440
2	-1.2365	-0.5875	-0.4520	0.1354
8	-0.8965	-0.1882	-0.3096	-0.1214
12	-0.3011	-0.1801	-0.0602	0.1199
mean	-0.7678	-0.2557	-0.2557	0.0000
var	0.1374	0.0493	0.0241	0.0252
st dev	0.3707	**0.2221**	0.1553	**0.1588**

Table 54: Path coefficient for the NLM 1 parameter (Omit 2, 8, 12)

	Regression Coefficient	σ_i / σ_y	Path Coefficient
B2 =	0.4189	1.6691	0.7111

Table 55: Uncorrelated residual for LRM with 1 parameter (Omit 2, 8, 12)

Uncorrelated residual 1 parameter	
r(y,U)= sigma e / sigma y	0.7149

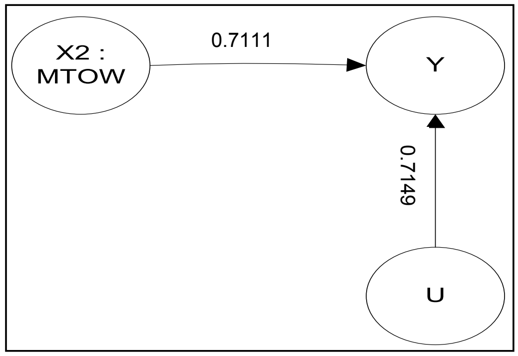

Figure 32: Path analysis diagram for NLM with 1 parameter (Omit 2, 8, 12)

Table 56: Percent errors of the NLM based on the MTOW (Omit 2, 8, 12)

Programs	Y: Actual Cost (USD)	Ŷ: Predicted Cost (USD)	*e* (%)
1	0.5105	0.6114	19.7622
3	0.5904	0.6831	15.7141
4	1.0053	0.8096	-19.4639
5	0.6281	0.6476	3.0954
6	0.9427	0.7128	-24.3816
7	0.8371	0.7457	-10.9216
9	0.8254	0.8515	3.1597
10	0.9127	0.8883	-2.6633
11	0.8208	0.9132	11.2572
13	0.8353	0.9647	15.4913
2	0.5557	0.6363	14.5049
8	0.8284	0.7337	-11.4288
12	0.8352	0.9416	12.7389
Ave. error data			1.1050
Ave. error (2,8,12)			5.2716

Table 57: Summary of results based on errors (Omit 2, 8, 12)

	% errors		% errors		% errors	
	3 Parameters		2 Parameters		1 Parameter	
Programs	MLRM	NLM	MLRM	NLM	MLRM	NLM
1	25.1082	9.6743	26.0101	9.6855	31.0595	19.7622
3	15.0984	7.8811	16.0180	8.0205	20.5428	15.7141
4	-23.6490	-22.9058	-22.3384	-22.5947	-20.1129	-19.4639
5	16.2770	15.1205	15.9453	15.1879	9.8061	3.0954
6	-18.7032	-18.2001	-18.9982	-18.2430	-22.4342	-24.3816
7	-6.8997	-7.2695	-7.7130	-7.5129	-9.9060	-10.9216
9	0.1799	-0.1198	-1.0969	-0.3964	1.5477	3.1597
10	-4.5702	-3.4523	-5.1867	-3.6022	-4.5511	-2.6633
11	7.6711	8.4213	6.6295	8.1805	8.9704	11.2572
13	15.5045	20.0783	17.0823	20.5229	13.2038	15.4913
2	18.3115	7.2734	19.5164	7.4075	22.9271	14.5049
8	-7.4971	-9.0117	-8.4756	-9.2862	-9.9906	-11.4288
12	11.7681	15.6112	12.9286	15.9586	10.4117	12.7389
Mean Error data	2.6017	**0.9228**	2.6352	**0.9248**	2.8126	**1.1050**
Mean Error validation	7.5275	**4.6243**	7.9898	**4.6933**	7.7827	**5.2716**
R squared	0.3700	**0.5741**	0.3672	**0.5740**	0.3388	**0.4890**

10.0 APPENDIX C: Interview Questionnaire for Cost Analysis

Category	Questions

Recurring Cost	▪ What is the most expensive part of the system?
	▪ According to your knowledge what drives the cost of the system?
	▪ What technical features does this system have?
	▪ What technical feature is known about the system in the early design phases?
	▪ What do you base a new system on?
	▪ What is the breakdown level of the cost?
	▪ Is there any learning included in the cost?
	▪ What is the planning base assumed when negotiated cost of the system?
	▪ Is there any non recurring amortized in the cost?
	▪ What is the percentage of the cost/BOM that is made in house?
	▪ What is the percentage of the cost/BOM that is made in emerging country?
	▪ Which country do you do business with Emerging Country or High cost country?
	    ▪ What type of contract agreement do you have?
	    ▪ How long have you been doing business with this supplier?
	    ▪ Do you own this company or any other company within this country?
	    ▪ What are the payment terms with your sub tiers?
	    ▪ What is the price adjustment agreement with your sub tiers?
	    ▪ Does any of your sub tiers have de-escalation?
	▪ Do you have an umbrella agreement with your suppliers?
	▪ How similar is your product from your clients?
	▪ Any discount if we surpass the forecasted planning base?
	▪ What is the percentage of material vs labor?

Non Recurring	▪ What is the breakdown level of the cost? ▪ Can you breakdown the Non Recurring in terms of Man Hours? ▪ Can you breakdown the Non Recurring in terms of Man Power/Year? ▪ How much effort/NRC can be leveraged from other programs? ▪ Over how many platforms are you amortizing the NRC cost?