SCIREA Journal of Energy

http://www.scirea.org/journal/Energy

September 25, 2020

Volume 5, Issue 1, February 2020

GAS TURBINE ENGINE PRICE ESTIMATION USING REGRESSION ANALYSIS

David Olusina Rowlands[1], Mark Savill[2]

[1]School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedford MK43 0AL, UK

[2]School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedford MK43 0AL, UK

Email: d.o.rowlands@cranfield.ac.uk (David Olusina Rowlands), mark.savill@cranfield.ac.uk (Mark Savill)

ABSTRACT

The economic climate for any process, product or industry is influenced by numerous variables. Any organisation willing to thrive must make deliberate effort to adequately understand the interactions between the elements, factors and variables influencing its economic climate. Economic analysis is a tool which determines how effectively a system is operating, or will operate, from an economic standpoint. The insight obtained from economic analysis provides useful information required for informed decision making. However, the reliability of any economic analysis is greatly influenced by accuracy in the adopted price of capital assets. This is especially true for investments demanding high capital such as power plant projects which require largely capital intensive assets like gas turbines as prime movers.

In this study, a model is developed which applies regression analysis to estimate the acquisition cost of gas turbine units from a dataset of historical records of gas turbine engine performance parameters and acquisition costs. As a validation to the implemented approach, the developed model is applied to estimate the acquisition cost of known gas turbine units. Results obtained from model predictions reveal an estimating accuracy between 72% and 98% with a coefficient of determination (R^2) of 94% and strong positive correlation (r) of 0.97 between the considered dependent and independent variables.

Keywords: Gas Turbine, Regression Analysis, Price Estimation, Acquisition Cost, Power Plant

NOMENCLATURE

E	Random error or residuals
K	Degrees of freedom
N or n	Number of Samples
R or r	Coefficient of correlation
R^2 or r^2	Coefficient of determination
X	Independent variable
y	Dependent variable
f	function
$\widehat{B}_0$ or $\hat{b}_0$	Y intercept
$\widehat{B}_1$ or $\hat{b}_1$	Slope
i_{rate}	inflation rate
S_1	Standard error of slope
X_{PWR}	Power output
$\widehat{y}_{AQC}$	Estimated acquisition cost
$\widehat{y}_0$	Target/ Best estimate
ANN	Artificial Neural Network

ANOVA	Analysis of variance
CV	Coefficient of Variation
PER	Price estimating relationship
PI	Prediction Interval
PPI	Producer price index
RA	Regression Analysis
REM	Repurposed engine model
SE	Standard error of estimate
max	Maximum

Symbols

e	exponential
Σ	Summation
α	Significance level
ln	Natural log

Subscripts

AQC	Acquisition cost
cyear	Current year
Estimate	
norm	Normalised
Power	
PWR	Power
pyear	Previous year
Rate	

INTRODUCTION

The investment of an organisation in any capital-intensive project constitutes a significant threat to the organisations financial stability. It is therefore extremely important for the right decisions to be made at every stage in an organisation investment plan. An ill-informed or wrong investment decision could spell irreversible disaster for an organisations debt/equity ratio. This is particularly true for organisations investing in power generating platforms utilizing gas turbines as prime movers in both simple cycle, combined cycle and combined heat and power applications.

Economic analysis is necessary to provide the insight and information needed for consistent informed decision making over the life of a project or investment option. Primarily, it provides an overview of all costs and benefits associated with an investment over the life of the investment. Several elements interact to influence the outcome of economic analysis. One of such elements, which is critical in defining a reliable analysis is the price of capital assets. It is essential that an analyst obtains price estimates that reflect very closely, if not exactly, the price of each capital asset considered in an economic analysis. Error in this area will results in wrong economic estimates and conclusions which can eventually lead to major crisis in an investment or project. An unreliable economic analysis will results in wrong decision making and planning.

The acquisition costs of capital assets like gas turbine units vary widely. This variation can be attributed to a number of factors some of which are listed below.

- Engine Performance Parameters
- Engine Performance requirements
- Engine Manufacturer and model
- Year of manufacture
- Implemented Technology
- Geographical location e. t. c.

The disparity in the acquisition cost of both similar and dissimilar gas turbine units coupled with the difficulty in obtaining accurate and consistent pricing information for gas turbines usually complicates the job of the economic analysts.

This study utilizes regression analysis to estimates gas turbine engine acquisition cost from inputs of engine performance parameters. A dataset of historical gas turbine engine records consisting of engine price, power output, power turbine rpm, heat rate and year of manufacture is used to evaluate a regression coefficient from which estimates are made. This method of parametric cost estimation provides an accurate first hand prediction of gas turbine engine acquisition cost suitable for conducting reliable economic analysis, necessary for guided and informed decision-making.

REGRESSION ANALYSIS (RA)

Regression analysis consists of a collection of statistical methods for determining the association between variables. These include techniques for modelling and analysing relationships between a dependent variable and one or more independent variables. Dependent variables refer to the predicted values in a regression model while independent variables are factors, which influence the predicted value (dependent variable).

Fundamentally, regression models consists of factors and elements which characterise the relationship between two or more variables. The interaction between these parameters and variables often express a deterministic relationship that describes a given regression model [1]. Outlined below are the parameters and variables associated with a regression model.

- Dependent Variables Y
- Independent Variables X
- Unknown Variables β
- Number of unknown variables K
- Number of Samples or data points N

Regression models relate a dependent variable Y to a function of independent variables X and unknown variables β as described by Equation (1). The unknown variable β represents a scalar or vector quantity. The relationship described in Equation (1) is implemented for each sample point in a dataset set of N sample points.

$$Y \approx f(X,\beta) \qquad (1)$$

Regression Model Characterisation

When the number of samples points **N,** in a dataset, is **less than** the number of unknown variables **K**, in a regression model, the system of equations defining the regression model is termed **underdetermined**.

$$(N < K)\ underdetermined \qquad (2)$$

When the number of samples points **N,** in a dataset, is **equal to** the number of unknown variables **K**, in a regression model, the function ***f*** is identified as linear. Thus Equation (1) can be solved exactly rather than approximately.

$$(N = K)\ linear\ function \qquad (3)$$

When the number of samples points **N,** in a dataset, is **greater than** the number of unknown variables **K**, in a regression model, there is enough information in the dataset to estimate the unknown variable β (described in Equation (1)). Thus, the system of equations defining the regression model are **overdetermined**.

In an overdetermined regression model, the excess information available can be used to provide statistical information about the unknown parameter β and the dependent variable Y [2].

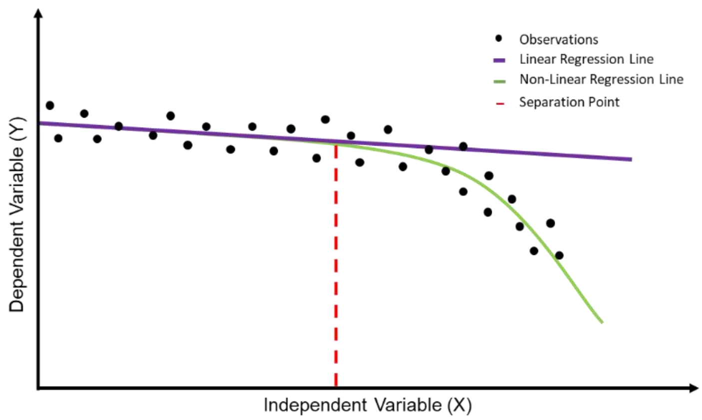

Figure 1: Graphical Representation of Regression Analysis

In Figure 1, the observations (data samples) are the results of random deviations from an underlying relationship represented by the linear and non-linear regression lines. A separation point also called the transition point is the point at which the random observations become non-linear. In such a situation, a non-linear regression model should characterize the underlying relationship between the dependent (Y) and independent variables (X) [3].

Regression analysis helps to identify the interaction between the dependent and independent variables and provides insight into the degree of influence that changes in various independent variables, associated with an investigated system or scenario, have on a dependent variable.

This study applies regression analysis to estimate the acquisition cost of gas turbine engines based upon engine performance parameters from a dataset of engine historical records.

Applications of Regression Analysis

Regression techniques are excellent methods for estimating relationships between variables in a system. Once developed, estimating relationships become excellent mechanisms for rapid evaluation of scenarios to guide decisions making. With regression analysis, estimators are able to statistically evaluate and quantify their level of confidence in the accuracy of estimates they provide.

However, it is often difficult for non-analysts to really understand the statistics associated with certain developed estimating relationships. Furthermore, apart from the complexity involved in gathering appropriate and relevant data and developing statistically correct estimating relationships, regression models tend to lose credibility outside their relevant data range [4]. It is therefore obligatory for an analyst to fully describe and document information associated with data preparation (including data selection, gathering, and normalization), estimating relationship development, statistical outcomes and conclusions. This is to enable adequate validation and acceptance of regression analysis results.

Regression analysis techniques have found application in a variety of fields. Some of these applications include:

- Design-to cost-trade studies
- Data-driven risk analysis
- Sensitivity analysis
- Software development
- Long-range planning
- Architectural studies
- Cross-checking

Presented below are some studies which apply regression analysis in the gas turbine field and in cost estimation.

- Rowlands and Theoklis employ linear and non-linear regression analysis to predict the remaining useful life of two single spool gas turbine engines. The study uses a dataset of recorded engine performance and operating parameters overtime to estimate future engine health. Conclusions from the study suggest that quality and quantity of data available for analysis as well as the considered prediction intervals influenced the accuracy of model predictions [2].

- Liu et al predicted the performance of gas turbine units from a dataset of recorded gas turbine operating parameters using multiple linear regression analysis. The study revealed that based on assumptions made, scenarios considered and data used, gas turbine performance can be effectively predicted from input of initial import conditions. Furthermore, application of multiple linear regression for predicting power output gave better results than for predicting turbine efficiency. The study concluded that the presented approach can be used to provide some reference values for larger data analysis and economic operation of power plants [3].
- Tsoutsanis and Meskin developed a data-based method for gas turbine performance prognostics taking into account dynamic engine operating modes. They apply linear regression to locally fit detected variation (i.e. degradation) in engine performance and predict gas turbine engine behaviour. The authors report that results obtained from the study demonstrate an improved accuracy for prognostics of gas turbines under dynamic modes in comparison to other methods [4].
- Kumar et al apply a statistical multivariate linear regression technique to predict the exhaust gas temperature (EGT) of a small gas turbine engine using input variables of air temperature, fuel control valve angle and gas pressure. They reported that results obtained from the study are indicative of an anomalous situation within a considered operating period and supports earlier findings from similar analysis conducted using artificial neural network (ANN) [5].
- Hamilton and Wormley employ regressions analysis to estimate hardware costs of turbine aircraft engines similar to the T3/5 engines. A linear regression relationship is developed and used to estimate hardware costs based on engine shaft horse power and shaft revolutions per minute (rpm). Statistical methods were used to validate results obtained from the analysis but the report, does not provide any other form of validation for the analysis conducted. This challenges the reliability of the implemented models which may be statistically accurate but not reflect actual and realistic relationships between the variables considered. [6].

In this study, regression analysis has been applied to predict gas turbine engine acquisition cost from inputted engine performance parameters.

METHODOLOGY

Presented in Figure 2, is the regression analysis methodology adopted in this study. This section describes each element of the methodology and how it is applied to estimate gas turbine engine acquisition cost.

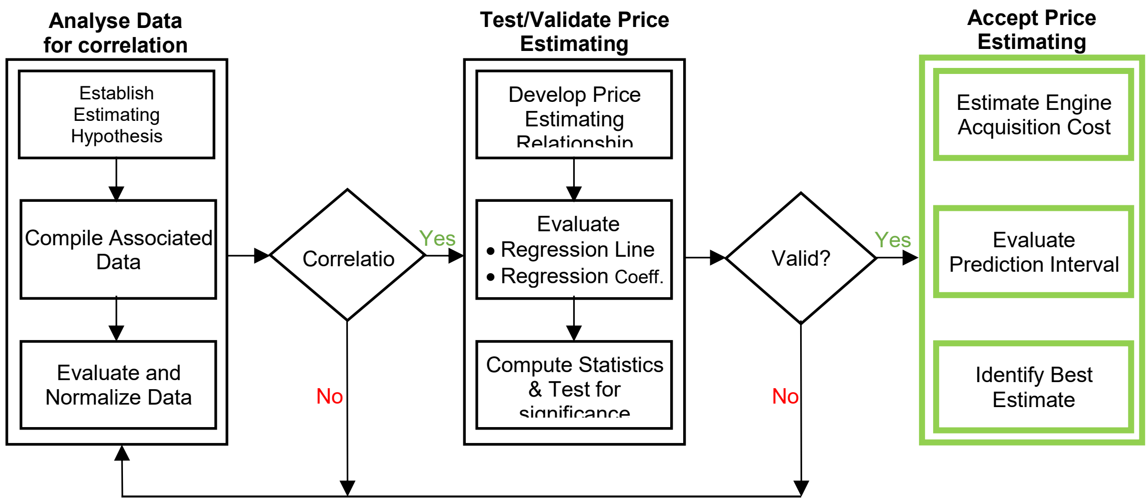

Figure 2: Regression Analysis Methodology for Estimating Acquisition Cost

Data Gathering and Analysis for correlation

This section of the methodology identifies correlation between the variables and parameters of a dataset to be used for regression analysis. A positive or negative correlations favours application of a dataset for analysis. However, a correlation value of zero or closer to zero indicates there is little or no relationship between variables and discourages application of a dataset for regression analysis (RA). The data gathering and analysis component of the presented methodology comprises three step described below.

Establish estimating Supposition

This is the first step in implementing the RA estimating methodology. It is concerned with developing an estimating supposition from which appropriate estimating variables are identified. This is important in guiding the nature and type of data to compile. In this study, gas turbine acquisition cost is estimated based on an underlying relationship between engine rated performance parameters and associated acquisition cost.

Compile Associated Data

With a supposition defined, appropriate data is compiled from historical records of engine acquisition costs and performance parameters retrieved from various sources including the Gas Turbine World Handbook, Turbomachinery International and the Nye Corporation Database [10, 11, 12] . The compiled dataset consists of records of engine price and associated engine power output, engine rpm and engine heat rate.

Normalize and Evaluate Data

The compiled data is normalized to ensure consistency and homogeneity. Normalization of the data in this study involves standardization of data units and data scaling by accounting for inflation in engine price using Equation 4 to Equation 6. The producer price index (PPI) industry data provided by the U.S. Bureau of Labour Statistics is used to evaluate inflation [13].

$$inflation\,rate,\, i_{rate} = \frac{PPI_{cyear}}{PPI_{pyear}} \qquad (4)$$

$$Price_{cyear} = \mathrm{Price}_{pyear} \times (1 + i_{rate}) \qquad (5)$$

$$X_{norm} = \frac{X}{max(X)} \qquad (6)$$

Where PPI_{cyear} is the PPI in the current year and PPI_{pyear} is the PPI in the previous year. X is the input before scaling and X_{norm} is the scaled input.

The normalized dataset is evaluated using a scatter plot and the Pearson’s Correlation Coefficient (also called Pearson’s r) to identify correlation between variables. The correlation coefficient is a single summary number that gives a good idea as to how closely a variable relates to another variable [14] . Equation 7 and Equation 8 describes the Pearson’s correlation coefficient.

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{(SS_X)(SS_Y)}} \qquad (7)$$

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\left(\sum X^2 - \frac{(\sum X)^2}{n_X}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n_Y}\right)\right]}} \qquad (8)$$

Where X and Y are the variables in a dataset and n is the number of samples for each variable in the dataset.

The degree of influence which one variable has over another is quantified by Pearson's correlation coefficient, which always falls between -1.00 and +1.00 (Higgins, 2005). A value of -1.00 indicates a perfect negative correlation and +1.00 indicates a perfect positive correlation. The closer the value of the correlation coefficient to -1.00 or +1.00 the stronger the relationships between variables and the greater the predictability, of the influence, of one variable over the other.

A positive relationship means that an increase in the value of one variable initiates a predictable increase in the value of another variable. A negative relationship implies that an increase in the value of a variable results in a predictable decrease in the value of another. No correlation indicates that a change in the value of one variable does not influence any change in the value of another variable [15]. The correlation coefficients for the dataset compiled in this study are presented in the following section. A strong positive correlation is observed between engine power output and engine acquisition cost (Figure 3).

Certain assumptions have been made to validate the application of the Pearson's Correlation Coefficient in this study. These assumptions are outlined below.

- Variables are measured on an interval or ratio scale
- The dependent variable (acquisition cost) is normally distributed in the population.
- The interactions between variables are characterized by a linear relationship.
- The values of the dependent variable are normally distributed across each value of the independent variable (Assumption of Homoscedasticity).

Development and Validation of Price Estimating Relationship

After establishing correlation between variables, regression analysis is conducted. This involves developing, testing and validating a price estimating relationship (PER) using analysis of variance (ANOVA) techniques and other statistical methods. The following sections describe these steps.

Develop price-estimating relationship (PER)

A price estimating relationship PER is developed with acquisition cost as the dependent variable, Y and Power output as the independent variable, X. Engine Power output parameter is the only independent variable implemented in this analysis due to the strong positive correlation observed between power output and engine acquisition cost. The evaluation of other variables revealed a very weak correlation with acquisition cost as shown in Figure 4 and Figure 5.

Since analysis of the compiled dataset reveals a linear functional form, represented by a strong positive correlation between power output and acquisition cost, a simple linear regression (SLR) model is initially applied, Equation 9. Data samples are generated from the compiled dataset using the simple linear regression model for a sample, Equation 10.

$$Y = \beta_0 + \beta_1 X + E \tag{9}$$

Where $\boldsymbol{\beta_0}(Y intercept)$ and $\boldsymbol{\beta_1}(the\ slope)$ are the population regression coefficients, and $\boldsymbol{E}$ is the random error or residual disturbance term. $\boldsymbol{Y}$ is the dependent variable and $\boldsymbol{X}$ is the independent variable.

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X \tag{10}$$

Where $\boldsymbol{\hat{b}_0}(Y intercept)$ and $\boldsymbol{\hat{b}_1}(the\ slope)$ are the population regression coefficients, $\boldsymbol{\hat{Y}}$ (acquisition cost) is the dependent variable and $\boldsymbol{X}$ is the independent variable (power output).

Evaluate regression line and regression coefficient

The regression coefficients are evaluated from Equation 11 and Equation 12.

$$\hat{b}_1 = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \tag{11}$$

$$\hat{b}_0 = \frac{n\sum Y - \hat{b}_1(\sum X)}{n} \tag{12}$$

With the regression coefficients evaluated, a simple linear regression model for price estimation is defined as shown in Equation 13. Thus, a predicted average value for the dependent variable $\hat{Y}_{AQC}$ can be determined for any value of the independent variable X_{PWR}.

$$\hat{Y}_{AQC} = \hat{b}_0 + \hat{b}_1 X_{PWR} \tag{13}$$

Figure 9 and Figure 11 show plots of the calculated sample data points (regression line) superimposed with the actual data from the compiled data set. To ensure the validity of the implemented price estimating approach and the analysed dataset, the following assumptions are made:

- The database obtained for analysis is homogeneous. This implies that the items in the database are of the same category. All engines are used for similar applications and are operated under similar conditions (at ISA SLS).

- A linear relationship exists between power output (independent variable, X) and engine acquisition cost (dependent variable, Y).
- The independent variable (power output, X) has been measured without error. Thus, any deviation in the data analysis is restricted to the dependent variables (cost, Y).
- Residuals, E, are normally distributed about the regression line.
- The residuals, E, come from an identical and independently distributed random distribution with mean zero and constant variance. Thus, residuals cannot be predicted from knowledge of the independent variable (power output, X).
- All acquisition costs have been adjusted for inflation to a common base year (2017).
- Engine acquisition costs include the cost of installation and accessories.

Compute statistics and test of significance

Tests are conducted to measure the validity of the regression line. This is essentially to determine how well the developed price estimating relationship (PER) predicts the dependent variable and whether or not a trend exists. Ultimately, these tests provide a basis for selecting or discarding the developed regression equation. Six techniques have been used to assess the regression analysis performed. These include:

- Coefficient of determination (R^2 or r^2)
- Coefficient of correlation (R or r)
- Standard error of estimate (SE)
- Coefficient of variation (CV)
- Graphical Residual Analysis
- Hypothesis test

Evaluate Coefficient of Determination (R^2)

Coefficient of determination (R^2) measures the accuracy of the regression line fit to the sample data points. R^2 ranges from 0 to 1 (0 to 100 percent). An R^2 equal to or greater than 90 percent is typically desired. An R^2 of 97 percent means that the regression model explains 97 percent of deviations leaving only 3 percent to chance. An R^2 of 100 means the regression model perfectly explains the deviations in considered variables. Equation 14 describes the coefficient of determination.

$$R^2 = \frac{[\sum XY - n(\bar{x})(\bar{Y})]^2}{[\sum X^2 - n(\bar{X})^2][\sum Y^2 - n(\bar{Y})^2]} \qquad (14)$$

Evaluate Coefficient of correlation (R or r)

The coefficient of correlation (R) is similar to R^2 but yields one piece of information not provided by R^2. This is the direction of the slope, whether negative or positive. In additions, it is a singular value, which gives information on how closely a considered variable relates to another. The coefficient of correlation is the square root of R^2 as described in Equation 15. R can also be evaluated from Equation 8. Typically, a value of R closer to -1.00 or +1.00 is desired.

$$\mathrm{R} = \sqrt{R^2} \qquad (15)$$

Evaluate Standard error of estimate (SE)

Standard error of the estimate (SE) is an absolute measure of the deviation of sample point from the regression line. It is calculated for each estimate using Equation 16.

$$\mathrm{SE} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-k-1}} \qquad (16)$$

Where n is the sample size, k is the number of independent variables in the price estimating relationship (PER). SE should be as small as possible.

Evaluate Coefficient of variation (CV)

The Coefficient of variation (CV) is a statistic that employs the standard error of estimate (SE) to evaluate the validity of a regression line. It is desirable that CV be less than 20 percent. CV is calculated from Equation 17.

$$\mathrm{CV} = \frac{SE}{\bar{Y}} \qquad (17)$$

Graphical Residual Analysis

With respect to the assumption that the Residuals, E, are independent, normally distributed, random variables, with a mean of zero and constant variance, another examination is required to determine the appropriateness of the developed PER for the regression analysis. This is to check whether the computed errors satisfy the assumption.

A random scatter plot of residuals against the independent variable is generated for each estimate. Visually analysing the plots should reveal that all points fall approximately within an equal band above and below a zero line. This would indicate that the errors are independent and normal as earlier assumed. If this is not observed, an alternative PER is considered.

Hypothesis Test

A hypothesis test is conducted to determine if the slope (the coefficient of the independent variable) is significantly different from zero. This is essential because:

- A slope of zero would indicate that the regression relationship is purely by chance.
- A slope that is not zero but also not significantly different from zero means, knowing the independent variable is of no use in estimating (predicting) the dependent variable.
- A slope significantly different from zero indicates that a knowledge of the independent variable is significant in estimating the dependent variable. Thus, the developed regression relationship (Equation 13) is regarded as significant and can thus be retained.

To conduct this test, it is necessary to calculate the standard error of the slope (S1) given by:

$$S_1 = \frac{SE}{\sqrt{\sum X^2 - n \sum \bar{X}^2}} \qquad (18)$$

A null and an alternative hypothesis is defined. The hypotheses considered in this study are as follows:

a) Null hypothesis

$$H_0{:}b_1 = 0$$

b) Alternative hypothesis

$$H_1{:}b_1 \neq 0$$

A significance level (α) of 0.05 is chosen for the test conducted in this study. Due to the composition of an alternate hypothesis, the test is defined as a two-tailed test. Variable t_p (from t-table) and t_c (Equation 20) are computed and a comparison of the evaluated values of t_c and t_p either accept or reject the established hypothesis.

$$t_p = t_{(1-\,(\alpha/2)\,,n-2)} \qquad (19)$$

$$t_c = \frac{\hat{b}_1}{S_1} \qquad (20)$$

In summary, If $|t_c| > |t_p|$, null hypothesis rejected [4] . If $|t_c| \leq |t_p|$, null hypothesis is NOT rejected [4] . If the hypothesis is accepted, the coefficients are statistically insignificant and the regression equation discarded. Alternatively, a rejected hypothesis indicates that the coefficients are statistically significant and the initial regression equation is retained.

If the statistical tests and analysis conducted, reveal that developed price estimating relationship, PER, adequately predicts the dependent variable *Y*, then the adopted regression model and computed regression line is identified as valid. Alternatively, if the results are unfavourable, an alternative regression model (e.g. Simple nonlinear regression) is considered. Equation 21 to Equation 25 describe the simple nonlinear regression model implemented in this study. *It is worth mentioning that even though a regression analysis provides valid outcomes from a statistical perspective, the causality of the equation is still analysed for a logical relationship.*

Simple Nonlinear Regression Model

When applying nonlinear regression analysis, the non-linear functional form (Equation 22) of the dataset is expressed in linear form (transformed) as described in Equation 21, by taking the natural log of the considered variables (dependent and independent variables). This yields the equation for a straight line in fit space where the transformed variables exist with slope $\widehat{B}$ and intercept $ln\widehat{A}$ [4].

$$ln\widehat{Y} = ln\widehat{A} + \widehat{B}\,lnX \tag{21}$$

$$\widehat{Y} = \widehat{A}X^{\widehat{B}} \tag{22}$$

The method of least squares is then applied to calculate the predictive values of slope $\widehat{B}$ and intercept $ln\widehat{A}$ from Equation 23 and Equation 24.

$$\widehat{B} = \frac{n\sum(lnX\,lnY) - (\sum lnX)(\sum lnY)}{n\sum(lnX)^2 - (\sum lnX)^2} \tag{23}$$

$$ln\widehat{A} = \frac{\sum lnY - \widehat{B}\sum lnX}{n} \tag{24}$$

$$\widehat{A} = e^{ln\widehat{A}} \tag{25}$$

Computing the statistics and tests of significance, to measure the validity of the nonlinear regression curve, is done using similar approach as described above. However, calculations for the coefficient of determination R^2 are performed using the transformed dependent and independent variables.

Estimation of Acquisition Cost

With a price estimating relationship, PER, tested, validated and accepted, an adopted regression model can be applied to estimate the acquisition cost for any gas turbine within the scope of the evaluated dataset, and at 95% confidence level, based on the engine's rated power output. The interval within which the estimates are expected to fall is computed using Equation 26.

$$PI = \widehat{Y}_0 \pm t_p \, SE\sqrt{\frac{n+1}{n} + \frac{(X_0 - \overline{X})^2}{\sum X^2 - n\overline{X}^2}} \qquad (26)$$

Where *PI* is the prediction interval, *SE* is the standard error or standard deviation of the estimate, $\widehat{Y}_0$ is the target or best estimate and t_p is obtained from the t table.

IMPLEMENTATION OF ESTIMATING APPROACH

Data compilation and Normalization

A dataset containing historical records of engine performance parameters (including engine power output, engine rpm and engine heat rate) and associated acquisition costs is compiled. The compiled dataset is normalized using the approach and equations described in the methodology section and the dataset variables are evaluated for correlation by means of scatter plots.

Figure 3 to Figure 5 show the scatter plots of power output, engine rpm and heat rate (independent variables) against acquisition cost (the dependent variable). The Pearson's Correlation formula in Equation 8, is used to identify and quantify the correlation between the considered variables.

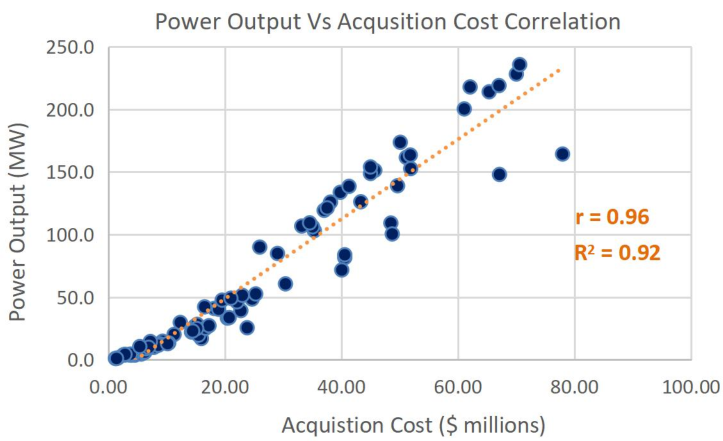

Figure 3: Scatter Plot of Power Output Vs Acquisition Cost

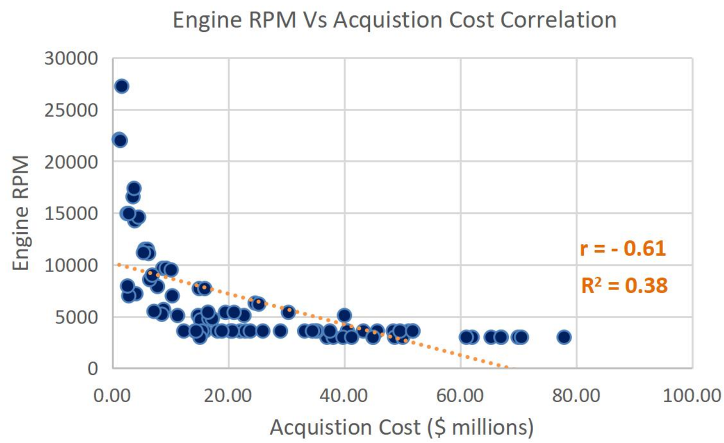

Figure 4: Scatter Plot of Engine RPM Vs Acquisition Cost

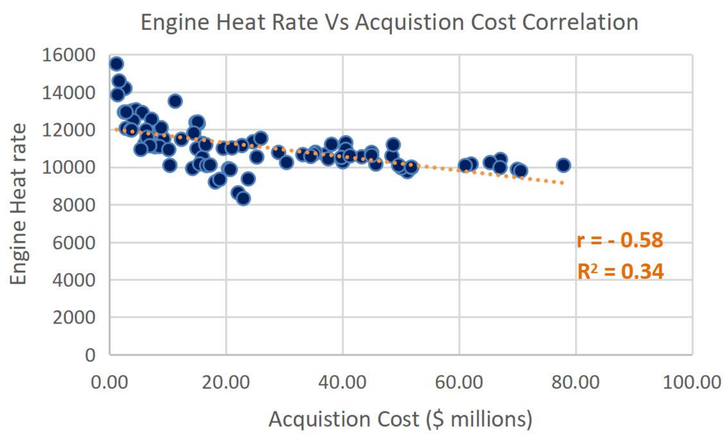

Figure 5: Scatter Plot of Engine Heat rate Vs Acquisition Cost

Evaluation of the dataset for each variable reveals the following:

- Power output has a strong positive correlation with acquisition cost (*r=0.96*). This means a strong relationship exists between power output and acquisition costs. The high R^2 value of *0.92* indicates that more variance can be accounted for in estimates and thus more accurate predictions of gas turbine acquisition costs are possible with accurate knowledge of gas turbine power output. Furthermore, a positive correlation suggests that an increase in gas turbine rated power output can influence an increase in gas turbine acquisition cost.

- RPM has a weak negative correlation with acquisition cost (*r=-0.61).* This means a weak relationship exists between engine rpm and acquisition costs. The very low R^2 value of 0.38 indicates lesser accuracy in the prediction of gas turbine acquisition cost from the knowledge of engine rpm. A negative correlation suggests that engines with lower values of rated rpm are likely to have higher acquisition costs.

- Heat rate has a weak negative correlation with acquisition cost (*r=-0.58*). This means a weak relationship exists between engine Heat rate and acquisition costs. The very low R^2 value of *0.34* implies lesser accuracy in the prediction of gas turbine acquisition cost from knowledge of engine heat rate. A negative correlation suggests that engines with lower values of heat rate are likely to have higher acquisition costs.

Further evaluation of the coefficient of correlation and determination between the datasets of considered independent variables is presented in Figure 6 to Figure 8.

In Figure 6 and Figure 7 power output has a weak negative correlation with engine rpm and heat rate. This suggests a weak relationship between them. The low values of R^2 suggest that any predictions made, based solely on the identified relationships, may not be accurate. Figure 8 shows a moderately strong positive relationship between rpm and heat rate. This suggests that there is some relationship between rpm and heat rate with the potential to make moderately accurate predictions. It also suggests that engines with higher rpm may have higher heat rates.

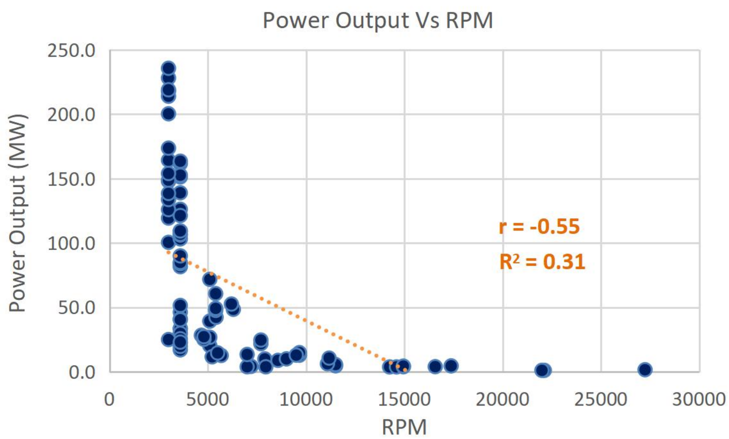

Figure 6: Scatter Plot of Power Output against RPM

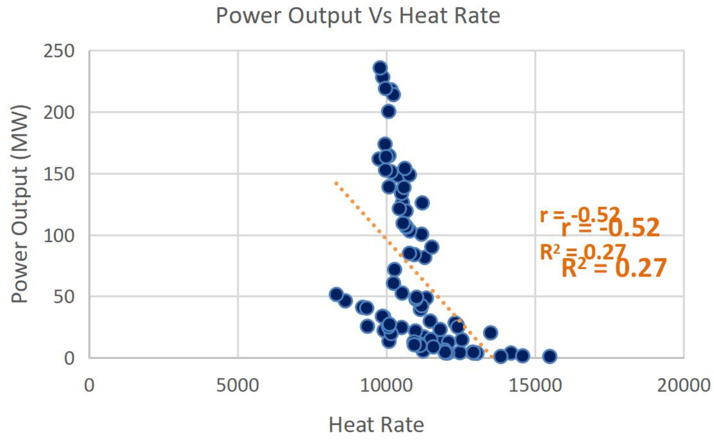

Figure 7: Scatter Plot of Power Output against Heat Rate

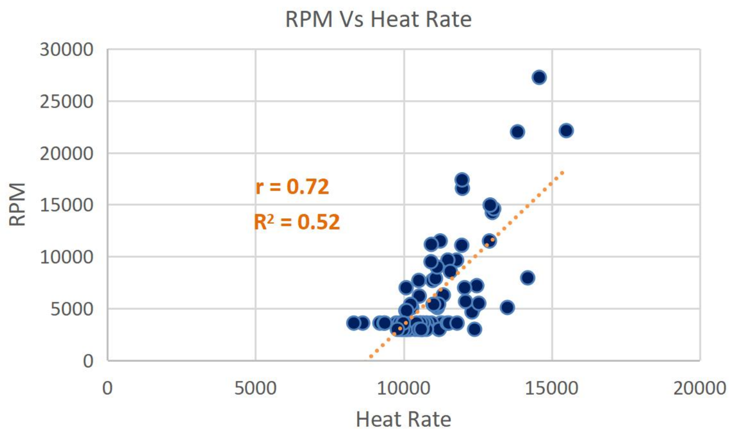

Figure 8: Scatter Plot of RPM against Heat Rate

Evaluation of the compiled dataset identified a very strong relationship between engine power output and acquisition cost and a weak relationship with other variables. The very low values of R^2 observed across the considered variable interactions suggests that less accurate predictions are more likely to emerge from said relationships. However, observed interactions between engine power output and acquisition cost dataset reveals a high R^2 value. This favours the development of a statistically valid and accurate regression model.

Thus, due to the very low values of coefficient of determination, (R^2) and the relatively weak correlations observed, engine rpm and heat rate variables have not been included in the development of a regression model for this study.

Development of Regression model

Upon evaluating and establishing correlation, a linear estimating relationship (Equation 27) is developed based on Equation 9 and Equation 10. The estimating relationship is used to evaluate a regression line as shown in Figure 9.

$$\widehat{Y}_{AQC\,estimate} = \widehat{b}_0 + \widehat{b}_1 X_{Power} \qquad (27)$$

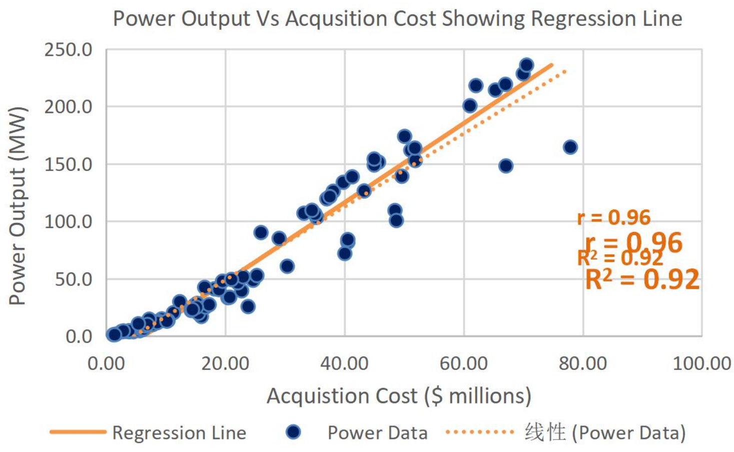

Figure 9: Scatter Plot of Power Output Vs Acquisition Cost Showing Regression Line

Analysis of the plot in Figure 9 reveals an increasing spread in data samples from left to right. It also reveals that the data samples are not normally distributed, equally, around the regression line. Even though a strong positive correlation has been identified between the power output and acquisition cost dataset, offset in the regression line suggests that residuals may not be normally distributed as previously assumed. This could mean that the applied regression model is not appropriate for the analysis. Therefore, in order to validate the developed regression model, graphical residual analysis is performed to identify if the previous assumption (*residuals are independently and normally distributed, random variables, with a mean of zero and constant variance*) is valid. The residuals are evaluated using Equation 28 and a plot of residuals against power output is analysed.

$$E = Y_{AQC} - \widehat{Y}_{AQC\,estimate} \qquad (28)$$

Where E is the residual (that is, the error between each data sample and regression estimate).

Observation of the plot in Figure 10 shows that the residuals are **not** normally distributed. Therefore, the implemented linear regression model is discarded and a simple non-linear regression model applied.

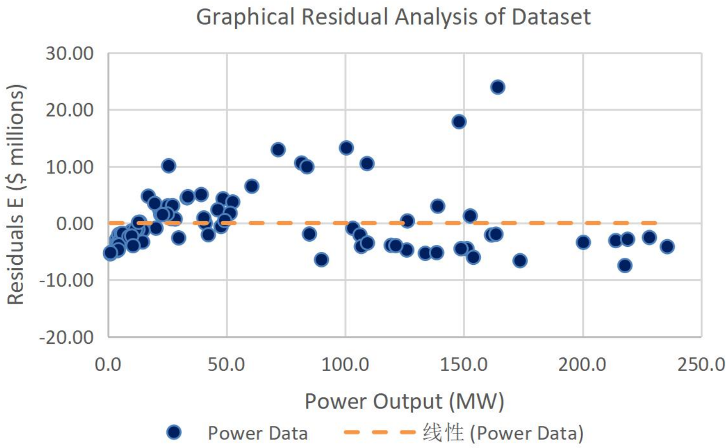

Figure 10: Graphical Residual Analysis with respect to Power output

A simple non-linear regression model (Equation 21) is implemented. Since the dataset has been identified as not having a linear functional form, the dataset set is transformed (expressed in linear form) by taking the natural log of the dependent and independent variables. This yields a nonlinear regression model of the form described in Equation 29. The transformed dataset is used to evaluate a new regression line. Figure 11 shows the non-linear regression line superimposed over the transformed dataset.

$$ln\widehat{Y}_{estimate} = ln\widehat{A} + \widehat{B}\, ln\, X_{Power} \qquad (29)$$

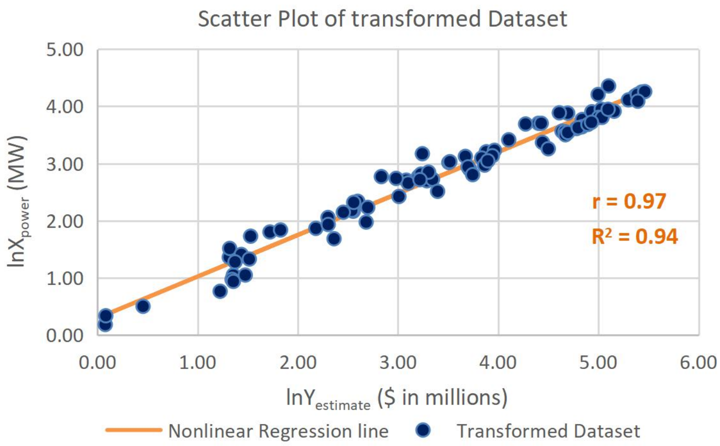

Figure 11: Scatter Plot of Transformed Dataset

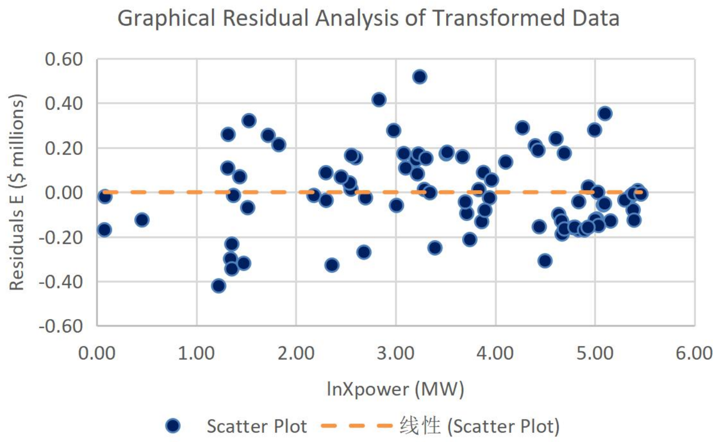

Figure 12: Graphical Residual Analysis of Transformed Dataset

Already, a much more evenly distributed dataset is observed from the scatter plot and graphical residual analysis, of the transformed data, shown in Figure 11 and Figure 12. For evaluation of actual values in unit space, the nonlinear regression model expressed in unit space form is used to evaluate the dependent variable (acquisition cost) from Equation 30.

$$\widehat{Y}_{estimate} = \widehat{A} X_{power}{}^{\widehat{B}} \qquad (30)$$

Variables $\widehat{A}$ and $\widehat{B}$ are evaluated using Equations 23 to Equations 25.

Statistics and Significance Test

In order to validate the implemented non-linear regression model, several statistical tests, are conducted.

Graphical Residual Analysis

Graphical residual analysis is performed to identify if the assumption of normality and independence of residuals is valid. This assumption has been validated in Figure 12 by an observed even distribution of residuals about a mean zero line.

The coefficient of determination, R^2

The coefficient of determination, R^2 is calculated from Equation 14 to measure the accuracy of the evaluated regression line fit to the transformed sample data points. An R^2 of **0.94 (94 percent)** is obtained. This means that the implemented nonlinear regression model adequately explains 94 percent of deviations leaving only 6 percent to chance.

The coefficient of Correlation, R or r

Evaluation of r gives information on the type and direction of the relationship between the dependent and independent variable. The coefficient of correlation was evaluated using Equation 15 as **0.97**. This implies that a very strong positive correlation exists between the dependent and independent variables. Identification of correlation between the variables, supports the application of the regression model for estimation of acquisition cost.

Standard error of estimate, SE

Evaluation of the standard error of estimate, SE, gives an absolute measure of the deviation of sample points from the nonlinear regression line (i.e. a measure of unexplained error). SE was evaluated using Equation 16 and gave a value of $5.1 million. The standard error of estimate on its own cannot be used to evaluate a regression line but can only be used for comparison between regression lines and for evaluation of the coefficient of variation (CV) [4].

Coefficient of variation, CV

The calculated standard error of estimate, SE, is used to compute a coefficient of variation, CV, from Equation 16. The coefficient of variation is a relative standard error computed in order to validate a regression line based on a desired limit, less than or equal to 20 percent [4]. The CV for the developed regression model is evaluated as **0.196 (19.6 percent)**. This means that the adopted nonlinear regression model can be identified as valid since its CV falls below 20 percent.

Hypothesis Test

Hypothesis test is conducted to identify if the coefficient of the independent variable (slope) is significantly different from zero. Outcomes from the test (presented in table 1) reveal that the slope coefficient is statistically significant ($t_c > t_p$). This means that a knowledge of the independent variable (Power output) is significant in estimating the dependent variable (acquisition cost). Therefore, the developed nonlinear regression relationship is significant and thus retained.

Table 1: Hypothesis/t-test results

α	tp	tc	Hypothesis Status	Description
0.05	1.99	36.78	Rejected	Slope Coefficient is Statistically Significant (H_0:$b_1 \neq 0$)

Outcome from the statistical tests conducted in this study favour the retention and application of the developed non-linear regression equation as a predictive model for estimating gas turbine acquisition cost.

Evaluation of Prediction Interval

With the nonlinear regression model validated, a prediction interval (PI), within which estimates may fall, is computed with respect to the evaluated regression line. Equation 26 is applied to evaluate the prediction interval. Figure 13 shows the upper and lower prediction intervals obtain from the analysis.

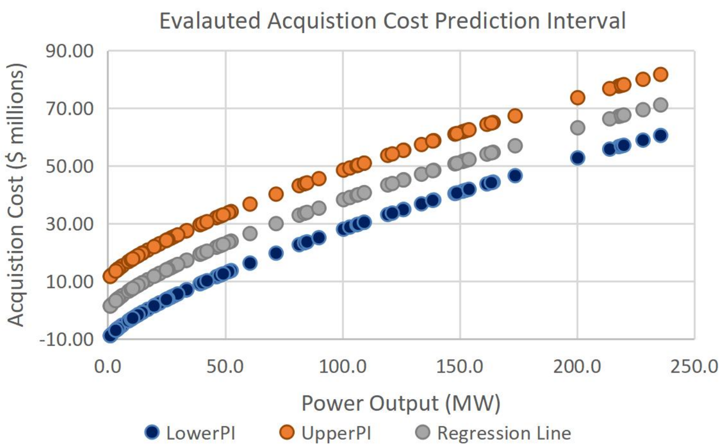

Figure 13: Acquisition Cost Prediction Interval

Validation of Regression Model Estimates

The developed regression analysis model has been applied to predict the price of known gas turbine engines based on rated engine power output. Comparison is made between the regression model estimates and actual engine acquisition costs obtained from literature. Results are presented in the following section.

APPLICATION, RESULTS AND DISCUSSION

The developed regression analysis model has been applied to estimate the price of an aero-derivative engine model. This engine model is derived from a turbojet engine similar to the TUMANSKY R-25-300 Turbo Jet engine. The entire endeavour is part of a study which seeks to provide an alternative profitable use for the power plants of grounded fighter aircrafts in electrical power generation. To determine the economic feasibility of adopting the **repurposed engine model (REM)** for electrical power generation, economic analysis is necessary for which the acquisition cost of the engine model is required.

In order to evaluate the accuracy of the regression model predictions, the acquisition cost for four know engine models have also been estimated using the developed regression model. Table 2 presents the engine performance parameters, associated with the considered engines.

The RR Trent 60 engine specification and pricing is referenced from an order made by Qatar in 2006 [16]. The engines were ordered to complement gas boost duty. In 1977, the West German Navy ordered six LM2500 engines at $4.2million *($17million in 2017 dollars)* each. This was to power six new frigates [11]. An advert in 2013 placed a 26MW GG4/FT4C-1D gas turbine for sale at $14million *($14.7million in 2017 dollars)* [17]. A Siemens SGT-800 was priced in 2016 at $17million *($17.3million in 2017 dollars)* [16].

For acquisition cost estimates conducted in this study, all retrieved engine prices have been adjusted for inflation to 2017 dollars.

Table 2: Engine Performance Parameters [11, 16, 17]

Engine	Power Output (MW)	rpm	Heat Rate (KJ/KWh)	PR	Exhaust Temp. (K)	NOx (ppm)
REM	17.00	11149	9749	9.6	647	N/A
GG4/FT4C-1D	26.00	3600	12397	N/A	N/A	N/A
LM 2500+G4 $_{DLE}$	32.50	3600	9867	24.2	825.2	25.0
Siemens SGT-800	47.50	6608	9547	20.1	814.2	15.0
RR Trent 60$_{DLE}$	58.29	3600	8798	36.0	702.2	25.0

Comparing the acquisition cost estimates computed using regression analysis (presented in Table 3) with the known actual engine acquisition costs, retrieved from literature, it is evident that the regression model predictions are relatively close to the actual engine costs. Results show RA estimating accuracy between 71% and 99%. This suggests that the developed regression model estimates are reasonably accurate and can be applied to estimate gas turbine engine acquisition costs for preliminary economic analysis and planning.

The slight overestimation and underestimation of acquisition costs observed can be attributed to uncaptured elements which influence gas turbine price. There are certain influencing factors in a system which cannot be adequately modelled by a regression equation. The effect of these unknowns on RA predictions can be minimised by employing larger quantity of accurate data samples in a regression analysis. Also, the implementation of multivariate regression analysis can enhance model prediction accuracy by taking into account, the influence of other variables in the regression equation. Figure 14 and Figure 15 shows the results obtained from this study.

Further investigation is performed to compare acquisition cost predictions made using the implemented regression analysis model with those obtained from artificial neural network estimates [18]. Results reveal that the artificial neural network predictions are much more accurate than the regression analysis predictions.

However, considering the uncertainty and variability in the acquisition cost of GTs, the interval of uncertainty captured by the regression analysis estimates, presents a much more realistic representation of the variability associated with the pricing of gas turbine units. An interval within which the acquisition cost of a GT may fall, due to variability in certain influencing factors, is captured by the regression analysis predictions shown in Figure 13 and Figure 16.

Table 3 and Table 4 show the accuracy of the estimates in relation to the target cost value for each unit.

Table 3: Accuracy of Regression Analysis Estimate **(2017 dollars)**

Engine	Target Value ($ in millions)	RA Estimate ($ in millions)	Accuracy (%)
REM	N/A	11.00	N/A
GG4/FT4C-1D	14.71	14.30	97.21
LM 2500+G4$_{DLE}$	17.04	16.86	98.94

Siemens SGT-800	17.34	22.22	71.84
RR Trent 60_{DLE}	23.25	25.70	89.46

Table 4: Accuracy of ANN Estimate (2017 dollars)

Engine	Target Value ($ in millions)	ANN Estimate ($ in millions)	Accuracy (%)
REM	N/A	10.00	N/A
GG4/FT4C-1D	14.71	14.72	99.94
LM 2500+$G4_{DLE}$	17.04	17.15	99.34
Siemens SGT-800	17.34	17.92	96.62
RR Trent 60_{DLE}	23.25	24.41	95.02

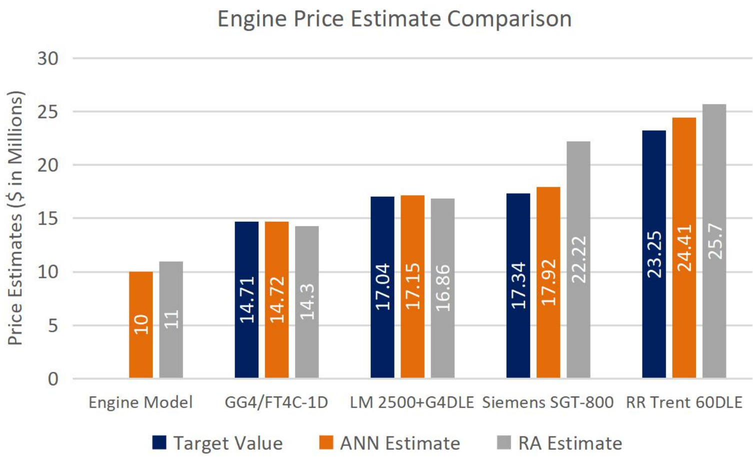

Figure 14: Acquisition Cost Estimate Comparison

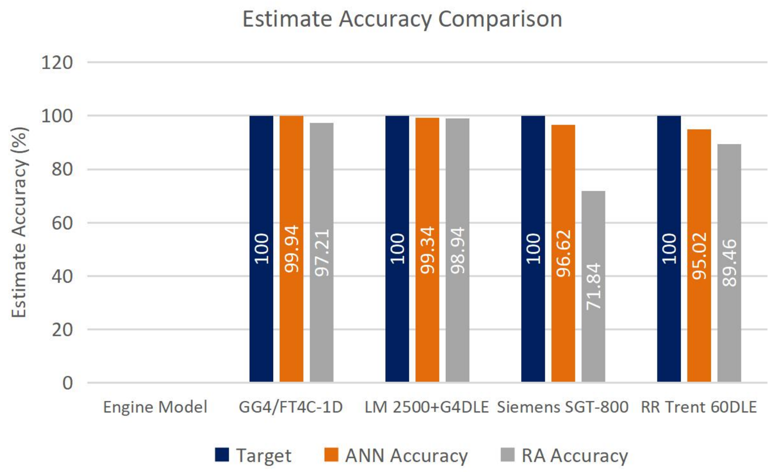

Figure 15: Estimating Accuracy Comparison

The prediction interval within which estimates for the acquisition cost of the repurposed engine model (REM) may fall is shown in Figure 16.

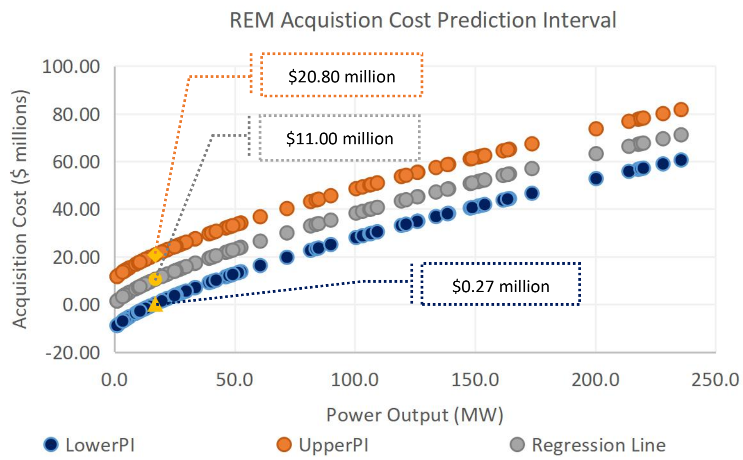

Figure 16: REM Acquisition Cost Prediction Interval

CONCLUSION

This paper presents the application of regression analysis to gas turbine engine price estimation. The developed approach, utilizes engine performance parameter of power output and associated engine acquisition cost to predict the cost of gas turbine engines. A dataset of historical records of gas turbine engine performance and acquisition cost is used to develop a price estimating relationship from which predictions are made. Evaluation of the relationship between the engine acquisitions cost (dependent variable) and power output (independent variable) dataset revealed a strong positive correlation with an R value of 0.97 and an R^2 of

94%. The computations performed in this study are based on dataset records, of gas turbine engines, for industrial and marine applications.

Linear and non-linear regression analysis techniques have been used to develop a regression model capable of estimating acquisition cost at 95% confidence level. The implemented regression analysis model has been applied to predict the acquisition costs of four know gas turbine engines with accuracy between 71% and 99%. Results obtained from the study indicate that reasonably accurate estimates are obtainable from the presented regression analysis approach. This suggests that the approach can be applied in predicting gas turbine engine acquisition costs for preliminary economic analysis and planning activities. Comparison of regression analysis estimates for the investigated engines with estimates from artificial neural network reveals higher accuracy for predictions with artificial neural network.

Further investigation is recommended into the application of multivariate regression analysis approach for estimating gas turbine acquisition costs. A multivariate approach enables additional variables such as engine pressure ratio, exhaust temperature and emissions data, to be considered in the regression model. This will potentially improve the accuracy of the model estimates. Introduction of additional performance parameters to improve the accuracy of the implemented estimating approach in this study, has been restricted by the difficulty in accessing sufficient and complete data required to adequately represent the quality, quantity and type of data needed for the investigation.

The scope of this study has been limited to industrial and aero derivative engine price estimation. Further study can be conducted into the application of regression analysis for predicting the acquisition costs of aero-engine.

REFERENCES

[1] O. Akinyemi, “Gas Turbines Performance Prognostics,” Cranfield University, Bedfordshire, 2008.

[2] M. A. Saint-Germain, “Research methods-Simple Regression,” Carlifornia State University, Long Beach, Carlifornia, 2001.

[3] Y. G. Li and P. Nilkitsaranont, “Gas turbine performance prognostic for condition-based maintenace,” Journal of Applied Energy, vol. 86, no. 10, pp. 2152-2161, 2009.

[4] National Aeronautics and Space Administration, NASA, “Cost Estimating Methodologies - Parametric Cost Estimating,” in NASA Cost Estimating Handbook, NASA, 2015, pp. 16-18.
[5] D. O. Rowlands, “Application of Gas Turbine Performance Prognostics to Two Single Spool Gas Turbine Engines,” Cranfield University, Bedfordshire, 2014.
[6] C. Liu, R. Jin, E. Gong, Y. Liu and M. Yue, “Predicition for the performance of Gas trubine Units Using Multiple Linear Regression,” in Proceeding of the Chinese Society of Electrical Engineering, China, 2017.
[7] E. Tsoutsanis and N. Meskin, “Derivative-driven window-based regression method for gas turbine performance prognostics,” Energy, vol. 128, pp. 302-311, 2017.
[8] A. Kumar, A. Banerjee, A. Srivastava and N. Goel, “Prediction of Exhaust Gas Temperature in GTE by Multivariate Regression Analysis And Anomaly Detection,” in IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), Toronto, ON, Canada, 2014.
[9] J. L. Hamilton and J. T. Wormley, “Application of Regression Analysis to Cost Analysis,” US Army Materiel Command, McLean, Virginia, 1968.
[10] Nye Thermodynamics Corporation, “Gas Turbine Prices $ per KW,” Nye Thermodynamics Corporation, Ontario, 2016.
[11] Gas Turbine International, “Latest News - General Electric Sells,” Sawyer's Gas Turbine International, vol. 18, no. 4, p. 56, 1977.
[12] Pequot Publishing Inc, Gas Turbine World Handbook, Southport: Pequot Publishing Inc, 1973-2017.
[13] U.S. Bureau of Labor, “Statistics, U.S. Bureau of Labor Statistics- Databases, Tables & Calculators by Subject,” U.S. Bureau of Labor, Washington, DC, 2017.
[14] J. Higgins, The Radical Statistician, 5th ed., Carlifornia: The Management Advantage, Inc., 2005.
[15] D. Rowlands, “Techno-Economic Research Analysis for Effective Power Generation from Aero-Engines with Minimal Emissions,” Cranfield University, Cranfield, 2018.
[16] Forecast International, “Industrial Trent 60 / MT30 Marine Engine,” PowerWeb - A Forecast International Inc. Subsidiary, Newtown, USA, 2018.