A case study to estimate design effort for Pratt & Whitney canada *

Adil M. Salam1, Nadia F. Bhuiyan1, Gerard J. Gouw1, Syed Asif Raza2

1 Department of Mechanical and Industrial Engineering, Concordia University, Montreal, Canada
2 Centre de recherche sur les transports, Université Montréal, Montréal, Canada

(Received December 25 2007, Accepted March 10 2008)

Abstract. The design effort required to complete a project is an important aspect of a project. It impacts the final cost, as well as the lead-time of a project. In this paper, a case study, which is carried out at Pratt & Whitney Canada, a global leader in the design and manufacture of aircraft engines, is presented. A Parametric model is proposed to estimate the design effort required in for a particular department to complete their design phase of an integrated blade-rotor low-pressure compressor fan. In a sensitivity analysis, the model estimation is compared with the actual estimates and the comparison demonstrates that the parametric model results in a good estimation. The analysis further explores the impact of various factors used to develop the parametric model, as well as demonstrates the significance of the proposed modeling methodology.

Keywords: design effort, parametric model, estimation, sensitivity analysis

1 Introduction

The primary issues in a project are the lead-time and the cost it may require. To accurately estimate the time required to complete a project would resolve a lot of problems related to forecasting, scheduling, bidding and reputation. Even though, there have been studies to control a project according to plan (Thamhain and Wilemon, 1986; Grant et al., 2006)[10, 22], there exists a need to further investigate this matter. A case study conducted by Bounds (1998)[5], stated that only 26% percent of the projects completed in the US were on time and within budget.

Frimpong (2000)[9] found that only 25% of construction projects were within budget and completed on time. Assaf and Al-Hejji (2006)[1] conducted a survey and found that only 30% of the construction projects were completed on time. The delays for the projects ranged from 10-30%. Moreover, the research of Norris (1971)[19] and of Murmann (1994)[18] pointed out that the unexpected or underestimated cost of projects was between 97-151% more than the original estimate. It is even more drastic in schedule, running from 41-258% later than originally estimated.

According to Bronikowski (1986)[6] these inaccurate estimates would sometimes lead to the termination of projects resulting in the company incurring huge costs and waste of effort of their resources. If a new product is being launched, time to market is critical. According to Ulrich and Eppinger (2003)[23], missing the target schedule could result in failing to launch the new product with competitors taking control of the market. There are several studies reported in the literature that show that the underestimation of design effort is a major cause of delays and budget cost errors (Colmer et al., 1999; Bashir and Thomson, 2004)[3, 7].

Colmer et al. (1999)[7] studied design effort and costing improvement issues in a new product development project. There paper presented a discussion on why these new projects are vulnerable to cost overruns.

* The authors are indebted to Mr. Yvan Beauregard from Pratt & Whitney Canada for his support and assistance towards this research.
† Corresponding author. E-mail address: a_salam@encs.concordia.ca.

Published by World Academic Press, World Academic Union
They discussed how the limitations of cost estimates lead to poor decision making, which contributes to cost and schedule overruns. They discussed some issues regarding project and risk management, which will help projects meet their own needs and the need of the customers.

Bashir and Thomson (1999)\cite{2} developed a model to estimate design effort based on the product complexity (PC). PC was assumed to be a metric that depends on the number of functions and the depth of their functional trees. Later, Bashir and Thomson (2004)\cite{3} analyzed fifteen designs and developed a parametric model to estimate the time required to design hydroelectric generators for General Electric.

The work presented in this paper extends the studies carried out in Bashir and Thomson (2004)\cite{3}. The model presented in this research is also parametric, however, the work presented in paper differs from an earlier study presented in Bashir and Thomson (2004)\cite{3} since it is a case study which is specialized for aerodynamic department at Pratt and Whitney Canada (PWC). In order to develop parametric model for the estimation of the design effort in an integrated blade-rotor low-pressure compressor fan manufacturing process at the department. New factors specific to aerodynamic department at PWC are identified that are used in the proposed parametric model, which contribute to improve the design effort estimation. Bashir and Thomson (2004)\cite{3} proposed a generalized parametric model for estimating design effort for different project and thus used common factors such as product complexity (PC). PC remains the same for a single type of product, thus it is not considered in the case study specific to a single type of product development process at particular department at PWC. Next, we propose a parametric model to estimate the design effort for an integrated blade-rotor low-pressure compressor (IBR LPC) fan designed and manufactured at PWC. The company is interested in quantifying the design effort of various departments for engine components.

The rest of the paper is organized as follows: Section 2 presents the model of design effort estimation, including a discussion of the prominent factors that are used in parametric modeling, followed by an overview of the jackknife and data masking techniques utilized in the modeling. Section 3, presents the use of multiple linear regression (MLR) method to analyzed aforementioned parametric model. The outcomes of the MLR based analysis is reported in section 4. Finally, the conclusions and suggestions for future research are identified in section 5.

## 2 Design effort estimation for the aerodynamics department

In this section, a parametric modeling approach is presented. The proposed model enables estimation of the effort measured in human-hours required to design an IBR LPC fan in aerodynamic department at PWC. The data information used in the model development and analysis is from aerodynamic department PWC. In order to maintain the confidentiality of the equipment manufacturer’s information, the engines of the compressor fans being studied will not be disclosed. Furthermore, the actual data received from PWC are masked using the general additive data perturbation (GADP) method suggested in Muralidhar et al. (1999)\cite{17}. The GADP method maintains the characteristics of the attributes of the original data in terms of linear combinations and the correlation of factors while using the masked values for these attributes. Among the other essentially important features of the GADP method is that it is free from several types of biases and maintains security of the masked attributes.

Bashir and Thomson (2004)’s work suggested a parametric model built using the data for design time recorded for hydro projects at General Electric. Equation (1) states parametric model for estimating design effort used in the work. The model involves several factors that are believed to be the significant factors along with the product complexity (PC) factor.

$$\hat{E} = aPC^b D_1^{c_1} D_2^{c_2} \ldots D_m^{c_m}$$  \hspace{1cm} (1)

where,

- $\hat{E}$: Estimated design effort in hours
- $PC$: Product complexity
- $D_m$: Effort driver (factor $m$)
- $a, b, c_m$, Constants (weights) estimated from historical data
This research presents methodology to use the parametric modeling dedicated to the estimation of the design effort for IBR LPC fan in aerodynamic department at PWC. Whereas an earlier model suggested in Bashir and Thomson (2004)\textsuperscript{[3]} uses PC as a driver of effort, it is omitted in this work because this model is developed solely to a specific type of component in the same engine family for PWC; hence product complexity is not a factor. It is also expected a model specific to a department would improve estimation capability and would also give better understanding about the department from the parametric model. The proposed parametric model is shown in equation (2):

$$\hat{E} = a_0 D_1^{a_1} D_2^{a_2} \ldots D_M^{a_M}$$ (2)

The design effort estimation model was developed using data provided on seven specific design jobs (DJs) by PWC for the IBR LPC fan for a certain class of turbo fan engines. It was essential to identify the principal design parameters (factors) that may have significant impact on the design effort estimation. After holding several interviews and discussions with managers, designers, and project engineers at PWC, the following four factors, listed below, were selected as the parameters to be used in this study:

- Type of design (TD)
- Degree of change (DC)
- Concurrency (Con)
- Experience of departmental personnel (DE)

2.1 Type of design

When designing a component, the effort required to complete the will highly depend on the type of design (TD). An initial design is not usually expected to require the same amount of time as a redesign. For this reason, each DJ was assigned one of the following attributes:

- Initial Design: 1
- Redesign: 2

2.2 Degree of change

Another factor considered to be of importance was the degree of change (DC). The purpose of this factor is used to attribute a value to the level of rework created from the initial design to a redesign, or from a redesign to a subsequent redesign. Supposing there is a major change to a given design, the amount of rework generated would be expected to be greater than if only a minor design change is required. In fact, as will be seen in the data, in some cases, a major change resulted in more effort than the initial design. Thus, different values were attributed to the designs as shown below:

- Initial design: 1
- Redesign with minor modifications: 2
- Redesign with major modifications: 3

2.3 Concurrency

It is noticed that product development teams practice concurrent engineering (CE) at PWC. The impact of CE in reducing the lead time is well established in the works of many research such as Loch and Terwiesch (1998)\textsuperscript{[14]}, Yassine et al. (1999)\textsuperscript{[26]}, Joglekar et al. (2001)\textsuperscript{[11]}, Yassine and Braha (2003)\textsuperscript{[25]}, Bhuiyan et al. (2004)\textsuperscript{[4]} that CE reduces the overall lead-time to design components. On the other hand, it is equally important to understand the level or amount of concurrency involved in the design, because it has also been pointed out by numerous researchers such as Bhuiyan et al. (2004)\textsuperscript{[4]}, Wang and Yan (2005)\textsuperscript{[24]}, and Jun et al. (2005)\textsuperscript{[12]} that CE tends to create rework, which could add to the total effort required to complete the design of the component. Thus concurrency of activities was considered an important factor and was in turn included in the model. The concurrency values for each of seven design jobs (A to G) under study are estimated. To

\textit{MSEM email for contribution: submit@msem.org.uk}
accomplish this, a concurrency matrix was first created as shown in Tab. 1. For each of the $j$ periods (nine in this case), a weight from zero to one was given for the concurrency value, in a given period ($c_j$), that the compressor aerodynamics department (D1) had with the other departments (D2-D9), in designing the component. Equation (3), shown below is used to calculate the net concurrency value for each DJ.

$$Con_i = \frac{\sum_{j=1}^{n} c_j}{\text{Number of Periods}}$$

As can be seen from the Tab. 1, the net concurrency value for DJ A would be:

$$Con_A = \frac{0.33 + 0.50 + 0.75 + 0.75 + 1.00 + 0.75 + 0.88}{9} = 0.73$$

This shows that there is a high level of concurrency (73%) with DJ A and all the other DJ’s. The net concurrency values of all of the seven DJs for D2 are shown in Tab. 2.

<table>
<thead>
<tr>
<th>DJ</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td>0.73</td>
<td>0.67</td>
<td>0.64</td>
<td>0.83</td>
<td>0.73</td>
<td>0.71</td>
<td>0.66</td>
</tr>
</tbody>
</table>

2.4 Experience of departmental personnel

The departmental personnel experience is another factor of significant consideration. Since the experience of the person working on the job plays an important role when estimating the required effort they will need to complete the task. This phenomenon can be easily understood since it is obvious that a person having more experience in working on the design or analysis of the component will be expected to complete the task more quickly than a person with very little or no experience at all. The attributes for design experience (DE) were assigned for different ranges of years of experience as can be seen below:

- 0-2 years of experience: 1
- 3-4 years of experience: 2
- 5+ years of experience: 3

In order to quantify the experience of departmental personnel, it is also assumed that if there is more than one person in a department having different levels of experience, then the experience level for a particular job will be calculated from the weighted average of experience from all the $n$ persons working on the design job. This can be seen from the equation (4).

$$DE = \sum_{j=1}^{n} (\% \text{ hours of } i) \times (\text{experience of } i)$$

At PWC, there are seven observations of DE contain the actual design effort available for the IBR LPC fan for the analysis as well as a quantitative measure of the aforementioned factors; TD, DC, and DE.
2.5 Parametric modeling

The suggested parametric modeling approach uses multiple linear regression (MLR) technique for the estimation of design effort. Due to the scarcity of the data, the jackknife technique is also utilized with the MLR. This technique is commonly used not only to improve the problem of biased estimation due to small sample size, but also in situations where the distribution of the data is hard to analyze (Efron and Tibshirani 1993)[8]. In principal this technique is based on sub-sampling rule in which the data are divided into sub-samples, and the sub-samples are obtained by deleting one observation at a time. The calculations are carried out for each sub sample. In a data set \( x = (x_1, x_2, x_3, \ldots, x_n) \), the \( i^{th} \) jackknife sample \( x_{-i} \) is defined to be \( x \) with the \( i^{th} \) data point removed. The pseudo-values, \( P_{s_i} \), are determined using equation (5).

\[
P_{s_i} = ns\hat{\beta} - (ns - 1)\hat{\beta}_{-i}
\]

where,

- \( P_{s_i} \): Pseudo-value for the entire sample, omitting sub-sample \( i \).
- \( ns \): Number of sub-samples
- \( \hat{\beta} \): Least-squares estimator of the whole sample
- \( \hat{\beta}_{-i} \): Least-squares estimator for the entire sample, omitting sub sample \( i \)

The jackknife estimator \( \tilde{\beta} \) is determined as follows with equation (6).

\[
\tilde{\beta} = \frac{\sum_{i=1}^{ns} P_{s_i}}{ns}
\]

Furthermore to the jackknife technique, GAPD model proposed in Muralidhar et al. (1999)[17] was also considered in this study to mask the confidential data used in the parametric model.

3 Parametric model analysis

Data was gathered from PWC for the seven DJs. The values for TD, DC, DE and the actual value (ACT) can be seen in Tab. 3.

<table>
<thead>
<tr>
<th>DJ</th>
<th>TD</th>
<th>DC</th>
<th>Con</th>
<th>DE</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>0.73</td>
<td>3</td>
<td>182.45</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
<td>0.67</td>
<td>2.80</td>
<td>784.87</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
<td>0.64</td>
<td>3</td>
<td>825.71</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>0.85</td>
<td>2</td>
<td>218.56</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>2</td>
<td>0.73</td>
<td>2.07</td>
<td>816.78</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>3</td>
<td>0.71</td>
<td>2.43</td>
<td>864.88</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>0.66</td>
<td>1.70</td>
<td>228.07</td>
</tr>
</tbody>
</table>

As discussed earlier, the GAPD model of Muralidhar et al. (1999)[17] was used to mask the confidential attribute, in this case the data on design effort (referred as ACT) measured in person-hours. Thus, whenever the term ACT is used in this paper, it refers to the masked values of actual design effort. It is also important to mention that this analysis was carried out on four departments at PWC. However, the analysis for only the compressor aerodynamics department is presented in this paper. The analysis details with the rest of the departments can be found in Salam (2007)[21].

The linearization of the proposed parametric model comprised of aforementioned factors. The linearization enables the use of MLR technique and resulting model is presented in Equation (7). The parametric proposed model is similar to the model suggested in Bashir and Thomson (2004)[3], however, with distinct factors to estimate the design effort.
\[ \ln E = \ln(a_0) + a_1 \ln(TD) + a_2 \ln(DC) + a_3 \ln(Con) + a_4 \ln(DE) \] (7)

The jackknife technique (Efron and Tibshirani, 1993)\(^8\) is used and sub-samples of the data are generated. The technique estimates the jackknife regression coefficients (COEFJACK) associated with each factor and these are reported in Tab. 4:

**Table 4. Regression coefficients**

<table>
<thead>
<tr>
<th>Constants</th>
<th>All</th>
<th>JackA</th>
<th>JackB</th>
<th>JackC</th>
<th>JackD</th>
<th>JackE</th>
<th>JackF</th>
<th>JackG</th>
<th>COEFJACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\ln(a_0))</td>
<td>5.583</td>
<td>5.504</td>
<td>5.644</td>
<td>5.598</td>
<td>5.428</td>
<td>5.644</td>
<td>5.598</td>
<td>5.605</td>
<td>5.575</td>
</tr>
<tr>
<td>(a_1)</td>
<td>1.762</td>
<td>1.727</td>
<td>1.604</td>
<td>1.789</td>
<td>1.814</td>
<td>1.908</td>
<td>1.732</td>
<td>1.690</td>
<td>1.752</td>
</tr>
<tr>
<td>(a_2)</td>
<td>0.223</td>
<td>0.180</td>
<td>0.339</td>
<td>0.200</td>
<td>0.191</td>
<td>0.257</td>
<td>0.231</td>
<td>0.147</td>
<td>0.221</td>
</tr>
<tr>
<td>(a_3)</td>
<td>-0.065</td>
<td>-0.050</td>
<td>0.023</td>
<td>-0.035</td>
<td>-0.401</td>
<td>0.023</td>
<td>-0.035</td>
<td>-0.441</td>
<td>-0.131</td>
</tr>
<tr>
<td>(a_4)</td>
<td>-0.337</td>
<td>-0.180</td>
<td>-0.383</td>
<td>-0.346</td>
<td>-0.315</td>
<td>-0.383</td>
<td>-0.346</td>
<td>-0.466</td>
<td>-0.345</td>
</tr>
</tbody>
</table>

From the jackknife values above, the predicted (PRED) \(\ln\) of ACT (actual masked values) would be as follows:

\[
PRED \ln(\text{ACT}) = 10.36 - 3.361 \times \ln(TD) + 2.112 \times \ln(DC) - 0.955 \times \ln(Con) - 4.972 \times \ln(DE)
\]

An essential assumption in the use of MLR is linearity. As discussed in Kutner et al. (2005)\(^{13}\), we carried out two tests in order to validate the linearity assumption. The first test of linearity was carried out by generating the scattered plot of the standardized residuals against the predicted values for the regression, which can be seen as Fig. 1.

**Fig. 1. Standardized residual plot**

Fig. 1 shows that the standardized residual values fall within the \(\pm 1\) threshold, and there are no curvilinear patterns, indicating a normal linear behavior. The residual plots were generated for all of the seven jackknife samples and they also behaved in a similar manner. The second test to be carried out is the normality of error test. The test requires the value of \(r\), which is the coefficient of correlation. The value of \(r\) is calculated from equation (8).

\[
r = \pm \sqrt{R^2}
\] (8)

The \(r\)-value has to be less than the critical value, \(r_L\) of 0.898 (Looney and Gulledge, 1985)\(^{15}\). The coefficient of determination, \(R^2\) values of the entire sample and all jackknife samples were calculated. The \(r\)-value was calculated from the minimum \(R^2\) value, assuming if it passes the test in the worst case, it would pass in the rest of the cases. The \(r\)-value calculated was 0.999, greater than \(r_L\) showing normal behavior of
the error estimates. Thus the developed parametric model (equation (9)) to estimate design effort, \( \hat{E} \) is shown below:

\[
\hat{E} = 2.63 \times 10^2 TD^{1.75} DC^{0.221} Con^{-0.131} DE^{-0.345}
\]

(9)

Next, we study the estimation accuracy of the model by calculating the relative error for each of the DJs. Tab. 5 shows the actual (masked) hours, the estimated (predicted) hours, and their respective relative errors for each of the DJs.

<table>
<thead>
<tr>
<th>Table 5. Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DJ</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>G</td>
</tr>
</tbody>
</table>

The study of Tab. 5 reveals that the model built performs very well as the maximum relative error found to be only 2.99%. The direct impact of one factor on another factor is determined by the estimating the correlation coefficient. Tab. 6 shows the correlation between all factors. However, the original data is masked, but with the use of the GAPD model for masking, the correlation coefficients describe same relationship as it would be in the case of the unmasked data. Furthermore, the correlation coefficient between two factors also determines the standardized regression coefficient performed between these two factors. Thus, as shown in Tab. 6, the correlation coefficient between TD and DC is high \((r = 0.945)\), therefore it states that with an increase in the measure one factor the measure of the other factor also increases.

<table>
<thead>
<tr>
<th>Table 6. Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>ln(TD)</td>
</tr>
<tr>
<td>ln(DC)</td>
</tr>
<tr>
<td>ln(Con)</td>
</tr>
<tr>
<td>ln(DE)</td>
</tr>
<tr>
<td>ln(ACT)</td>
</tr>
</tbody>
</table>

Even though the model is quite accurate, nevertheless it is important to verify if the factors are statistically significant. The t-test was used for this purpose and an analysis of variance (ANOVA) table is also reported. The following criterion defined by Kutner et al. (2005)\[13\] is used to determine if the factors are deemed significant.

If \((Pr > |t|) \leq 0.05\), the factor is considered statistically significant

Else

Factor is considered statistically insignificant

The summary of statistics, showing the \(Pr > |t|\) for the entire sample, and for all jackknife samples can be seen in Tab. 7.

The studies of Jogelek ar et al. (2001)\[11\], Bhuiyan et al. (2004)\[4\], Jun et al. (2005)\[12\] amongst many others, show that concurrency does impact the design effort, nevertheless due to the fact that the DJs are highly concurrent, resulting in concurrency being a statistical insignificant factor, as can be noticed from ANOVA.

Previously, the proposed parametric model also assumes the concurrency as a potential factor, however, in a statistical reported earlier, it is found insignificant. Thus, it seems more appropriate to reduce the factors
by excluding the concurrency and to reconstruct the model with the remaining factors. The new model is also subjected to statistical testing and has successfully passed aforementioned tests suggested for model validation. The new parametric model is reported in equation (10) which estimates design effort for the aerodynamic department at PWC is as follows:

\[ \hat{E}_{\text{No Con}} = 2.69 \times 10^2 TD^{-1.76} DC^{0.225} DE^{-0.325} \]  

(10)

Furthermore, the error remained minimal at a maximum value of 3.30%, while using the model reported in equation (10).

### 4 Sensitivity analysis

In this section, we report a sensitivity analysis to show the main effect of an individual factor’s contribution on the design effort. The results are reported in Fig. 2 ~ 4:

[Fig. 2. Impact of the type of design on design effort]

From Fig. 2, it can be seen that if the design were an initial one, it would take longer than that of a redesign. Fig. 3 identifies that the degree of change being major will require more time than a minor one. Finally, Fig. 4 shows that the higher the net level of experience of the departmental personnel working on the design job, the less effort is required.

### 5 Conclusions and future research suggestions

The importance of determining design effort is essential to estimate the lead-time, cost, and manpower needed. The product development process is a very complex process and it is dependent on many factors involved in the design process. Even though it is complicated to estimate the design effort needed for the development of a product, it is essential for the concerned personnel to know it precisely.

Depending upon the product complexity and development strategies, a list of potential factors can be identified that can help to improve such estimations. In this thesis, a case study focusing on four departments
The study utilizes a parametric modeling technique to estimate the design effort for each of the four departments. Four design factors are considered: type of design, degree of change, concurrency, and experience of departmental personnel. An analysis of each department initially considers all the above-mentioned factors and reduces to keep only the statistically significant factors. Since the data used in this thesis is confidential, coming from a leading aerospace company, an appropriate method for data masking was required. The model performed well according to a number of accuracy tests suggested in this paper. Comparison of the design effort estimation determined by the model is made with the actual design effort reported by PWC for each of the seven design jobs. The maximum error observed was 3.30%.

Even though the model has promising results there are limitations to this model. Even though the methodology can be used for other major components such as the turbine or combustor, the model is specific to the IBR LPC fans for a certain class of engines. Using this approach models will have to be developed for each type of component. Also, factors will have to be selected and validated for each type of component. Even though the factors are general, there may be other factors that will be relevant to other components being studied.

There can be several future applications of this thesis. As mentioned in the limitations, the model is specific to the component, other models can be developed which can be applied not limiting itself to a particular type of component or class of engine, rather the model can be more general. Another possible application could be to study the length of time (lead-time) to design the components. Knowing the lead-time and phasing of hours will significantly help management in scheduling their tasks and in assigning priorities.

**MSEM email for contribution:** submit@msem.org.uk
References


MSEM email for subscription: info@msem.org.uk