Vulnerability in Determining the Cost of Information System Project to Avoid Loses

Kholid Haryono\textsuperscript{1} and Zulfa Amalia Ikhsani

Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Jl. Kaliurang KM. 14.5 Sleman Yogyakarta Indonesia

\textsuperscript{1} Corresponding author: kholid.haryono@uii.ac.id

Abstract. Context: This study discusses the priority of cost required in software development projects. Objectives: To show the costing models, the variables involved, and how practitioners assess and decide the priorities of each variable. To strengthen the information, each variable also confirmed the risk if ignored. Method: The method is done by two approaches. First, systematic literature reviews to find the models and variables used to decide the cost of software development. Second, confirm and take judgments about the level of importance and risk of each variable to the software developer. Result: Obtained about 54 variables that appear on the 10 models discussed. The variables are categorized into 15 groups based on the similarity of meaning. Each group becomes a variable. Confirmation results with practitioners on the level of importance and risk. It shown there are two variables that are considered very important and high risk if ignored. That is duration and effort. Conclusion: The relationship of variable rates between the results of literature studies and confirmation of practitioners contributes to the use of software business actors in considering project cost variables.

1. Introduction

One of the success parameters software development project is the economic aspect in which the costs incurred can be controlled and the revenue is in line with expectations\textsuperscript{1}. The difference is also called profit or loss. Unfortunately, there are still many software developers hit by iceberg phenomenon where the problem seems small when in fact the problem is big, broad, and deep. As a result the costs can be out of control and be among the causes of project failure.

Knowledge of the costing factor to control it is important because most software developers cannot quickly estimate the costs involved\textsuperscript{2}. The hasty question is often asked by the user is what the price of software offered. Answering these questions require knowledge of the costs involved in software development. This is needed so that practitioners can use it to determine prices based on the cost variables involved. These variables have been shown in various models of cost estimation that have been widely used\textsuperscript{2}.

This study focuses on four issues, namely (a) what model is used to estimate software development costs; (b) what factors are taken in each model; (c) categorize the factors taken and prioritization of each category; and (d) how software developers perceive each category and provide a risk assessment of neglected categories. The purpose of this study is to show the factors that influence in determining cost of software, comparison between literature and assessment of practitioners, and to determine the level of risk of these variables that could potentially cause project losses.

The results provide new insights and knowledge for stakeholders in using the advantages and disadvantages of a factor in determining project costs. This knowledge is expected to improve the quality of decisions related the project costing, thus avoiding project losses. The quality of decisions taken by
software business practitioners about efforts to reduce project costs to achieve project success. To support the knowledge in the decision-making, any factors found will be confirmed. There are two confirmed, i.e, important level and risk level to software developers.

2. Motivation
We have two motivations for this research. First, provide important information about the factors or variables that affect the cost of software development for the perpetrators. Second, it gives confirmation of the influencing factors to the software business practitioners mainly related to the cost and the risks.

3. Research methodology
This study focuses on three stages, namely (a) literature review to look for software costing models. Each model is deepened by looking for its cost variables. The most variables appearing on the model are given the highest priority and order to the lowest. (b) empirical variable collection, all variables are confirmed to IT business practitioners. Correspondents are asked to assess the importance of each variable with the degree of importance 1 to 5. The higher the value is the more important the variable has an effect on the variable costing software. These variables are also confirmed about the level of risk posed if the variable is ignored. The risk rating of scale 1 to 5 also with value 5 has the greatest degree of risk (c) discusses the findings to discover new knowledge contributions:

3.1. Systematic Literature Review
Systematic Literature Review (LRS) has been widely used to observe, categorize, and evaluate various literature to reveal facts related to research areas undertaken by researchers by applying two techniques namely inclusion and exclusion. Systematic literature can be used for various fields. Systematic literature in accordance with the field of software engineering has been reviewed by Kitchenhan. Most researchers use three phases in LRS, that are planning, conducting, and reporting.

3.1.1. Planning the review.
Planning starts from determining the question for literature search. Search queries include: estimation model costing software development. The found models are then deepened by searching for the variables used each model with questions according to the model variables concerned.

The first source of data from general search through google. Some of the models found are LOC, COCOMO, COSYSMO, SLIM, PRICE-S, SEER-SEM, PROBE, EBS, REVIC. Each model is in the search for variables that are taken into account each variable.

Early models were found on Wikipedia with the keyword "Software development effort estimation" or "Cost estimation in software engineering". The models found are used as keywords to search for variables through scientific publications (journals and proceedings). This further study is used to determine the context of the variables used in each model. Selection of paper using inclusion and exclusion criteria to determine the appropriate paper. To make sure the paper is searched properly, keyword searching and paper collection are determined. The selected papers are selected through the title and abstract content

3.1.2. Conducting the review.
Based on data searches conducted using scientific data sources 2016 and 2017 with the filter "original research" obtained the number of papers such as Table 1.

| No | Sources      | Found | Relevant |
|----|--------------|-------|----------|
| 1  | Science Direct | 131   | 19       |
| 2  | Google Scholar | 160   | 11       |
|    | **Amount**   | **291** | **30**   |

3.1.3. Reporting the review.
Based on the literature, the variables of each model are shown in Table 2.

Table 2. The variables in model

| No | Model Name | The variables involved |
|----|------------|------------------------|
| 1  | COCOMO     | Staffing; Productivity; Effort; Duration; Maintenance |
| 2  | COSYSMO    | Personal Team; Size; Maintenance; Migration; Complexity; Process Business; Stakeholder Cohesion; Risk; Documentation; Tool Support |
| 3  | REVIC      | Staffing; Productivity; Effort; Duration; Maintenance; Integration; Testing Complexity |
| 4  | SEER-SEM   | Staffing; Productivity; Effort; Duration |
| 5  | SLIM-Putnam Model | Productivity; Size; Effort; Duration; Function |
| 6  | PRICE-S    | Scope; Effort; Scheduling; Cost |
| 7  | Evidence Based Scheduling | Time; Activity; Budget |
| 8  | ITK Method-CETIN | Labor maintenance; Labor middleware; Function Size; Middleware; Complexity; Design; Budget |
| 9  | PROBE      | Size; Time; Quality; Effort |
| 10 | Software Equation | Productivity; Effort; Line of Code; Skill Factor; Time |

Furthermore, the variables of each model are grouped by their similarity. For example staffing variables are owned by five models namely COCOMO, COSYSMO, REVIC, SEER-SEM, and CETIN\(^1\). The variables arranged according to the priority are shown in Table 3.

Table 3. List of variables from each group

| No | Variable | Amount Model | Term on the model |
|----|----------|--------------|-------------------|
| 1  | Duration | 8            | Duration; time; scheduling |
| 2  | Effort   | 7            | Effort             |
| 3  | Complexity | 7          | Complexity to understanding; Process business capability; Testing complexity; Skill factor; Design complexity; Stakeholder cohesion |
| 4  | Staffing | 6            | Staffing; Personal team; Labor |
| 5  | Size     | 5            | Size; Scope; Function Size |
| 6  | Productivity | 5          | Productivity platform; Productivity |
| 7  | Budget   | 3            | Cost; Budget       |
| 8  | Maintenance | 3           | Maintenance        |
| 9  | Function | 2            | Function; Task activity |
| 10 | Integration | 2           | Integration; Migration |
| 11 | Quality  | 1            | Quality            |
| 12 | Middleware | 1           | Middleware         |
| 13 | Line of Code | 1          | LOC                |
| 14 | Tools Support | 1          | Tools Support      |
| 15 | Documentation | 1          | Documentation      |
3.2. Empirical data collection
This stage confirms the variables found in the literature study to software business actors. There are two confirmed, which are related to the degree of importance and the degree of risk that can occur if those variables are ignored.

Findings on the degree of importance will be compared with the priorities found in the literature study. This comparison becomes important information especially if there is a significant difference because the difference can be new knowledge that is very useful for decision-making in the future. Risk assessment can be used as knowledge of potential losses and strong suggestions on the use of variables in project costing.

Wherever possible try to ensure that the size of the text in your figures (apart from superscripts/subscripts) is approximately the same size as the main text (11 points).

3.2.1. Respondents.
Conditions of respondents in this study there are two, the first person who is active in the software company, and second on the company, this person has a position to calculate the cost of production of IT projects and or determine the selling price of the software. Based on the distribution of questionnaires, the participating respondents came from 14 companies with the number of respondents 15. The names of the companies are: CV.Ainan; Partikelir; Prima solusi hexa; PT. Krafthaus Indonesia; Maylab; Adhi jasa informatika; CV. SAK; Ngrembel Asri; PT Skyshi Digital Indonesia; Smallstep Labs; Folarium; Orion Techno; Softwareseni; and ATSoft.

![Figure 1](image1.jpg)
**Figure 1.** The product type of these companies

![Figure 2](image2.jpg)
**Figure 2.** Age of respondents ranged from 22-43

The customer category of the company is divided into three, government (6), private (14), and abroad (4). Based on the educational background, the average of the last education of the respondent is S1. Distribution of educational background is 80% Bachelor degree, 13% Master, and 7% High school. The sex of the respondents is predominantly male. Of the 15 respondents, only one is female.

3.2.2. Survey and analysis.
The questionnaires that were distributed included two assessments namely the degree of importance and the degree of risk. The list of questions and number of respondents is shown in Table 4.

| Variable | Degree of Importance | Degree of Risk |
|----------|----------------------|---------------|
| Middleware | High | High |

The variables are sorted according to degree of importance and degree of risk. To calculate the highest score is by multiplying each variable with weight (1-5) the result is divided by the number of respondents it will produce a score of 1 - 5. The highest score shows the high degree of importance. The graph of degree of importance and degree of risk is shown in Figure 3.

Comparison of importance and risk levels are at the same level. It shows the perception of respondents to each variable is in accordance with the meaning of the variable. Only middleware variables alone have too high a perception gap. This happens because middleware includes the unpopular in determining the cost of developing an IT project.
**Table 4.** List of questions and amount answers respondents

| No | Element of Project Costing | Degree of importance | Degree of risk |
|----|-----------------------------|----------------------|---------------|
|    |                             | 5 is most important  | 5 is most risky|
|----|-----------------------------|----------------------|---------------|
| V1 | Duration                    | 11 4 3 2 1           | 11 4 3 2 1    |
| V2 | Effort                      | 8 7 6 0 0            | 8 6 0 1 0     |
| V3 | Complexity                  | 3 10 4 1 0           | 4 6 3 2 0     |
| V4 | Staffing                    | 5 5 4 1 0            | 6 6 3 0 0     |
| V5 | Size                        | 3 8 2 2 0            | 2 7 2 4 0     |
| V6 | Productivity                | 2 7 4 2 0            | 1 7 4 3 0     |
| V7 | Budget                      | 6 7 2 0 0            | 6 8 0 1 0     |
| V8 | Maintenance                 | 4 7 2 2 0            | 7 6 1 1 0     |
| V9 | Function                    | 8 5 1 1 0            | 9 4 0 2 0     |
| V10| Integration                 | 2 10 2 1 0           | 4 9 1 1 0     |
| V11| Quality                     | 7 6 2 0 0            | 7 6 2 0 0     |
| V12| Middleware                  | 3 4 5 2 1            | 6 6 2 1 0     |
| V13| Line of Code                | 5 7 3 0 0            | 4 7 2 2 0     |
| V14| Tool Support                | 7 7 1 0 0            | 8 5 1 1 0     |
| V15| Documentation               | 4 6 4 1 0            | 4 7 3 1 0     |

**Figure 3.** Graph of of degree of importance and degree of risk

**4. Finding**

Based on the data obtained from the various models and empirical data, found a shift in the variable priority between the model, importance level, and risk level of each variable. The shift is shown in Figure 4.

The findings show two consistent variables occupying the topmost level. Namely, duration and effort. This shows that the highest priority in determining cost of software development is the length of time required and the amount of effort that must be prepared and done.

Complexity variables dropped the 3rd priority position into the 8th importance and the level of risk further dropped to the order of 13. This means, according to the respondents varied complexity is on the middle priority in considering the cost of software development. Respondents assessed the risk level for this variable is also low if ignored so that this risk level data reinforce the decline in importance level.
In contrast, the variable quality rises from the order of the 11th model to the 5th importance level with the 4th risk level. This shows that quality is an important priority that has a high enough risk if ignored. Similarly, other variables.

5. Discussion

A discerning look at the variables involved in determining the cost of a software product becomes important because it involves the success of a project. In business, the success of a project is determined by the amount of the difference in cost and income received. Failure to see these variables can cause losses due to costs that should not be issued.

Based on empirical data collection, the priority level of the variable interest can be divided into three parts. First, the top priority variables of order 1-5 are duration, effort, tool support, function, and quality. This section is the top priority that is considered important in determining the cost of software. This section is supported by the highest risk level assessment. Rankings 1-5 on risk assessment are also occupied by these variables although the sequence is different.

Second is the five medium priorities. They are budget, LoC, complexity, staffing, and maintenance. Budget is client budgeted cost or client purchasing power to software project. Line of Code (LoC) is the number of program lines each function. Complexity is the uncertainty of problems and business procedures that will be resolved. A small visible problem shows the complexity of a project because the real problem will be found in the course of project work. Staffing is a team requirement, including experts who need to be hired. Maintenance is to consider the maturity level of the company/client. The less maturity IT will be the greater the cost of maintenance budgeted because the business is still full of uncertainty.

Budget, staffing, and maintenance at this intermediate level, risk level variable according to the variable of importance level. The other two variables are middleware and integration that outperform LoC and complexity variables. Middleware is a service that bridges between software. This section is strongly associated with integration so that two of these by the respondents considered quite risky.

Third, the variables that occupy the lower interest level are integration, documentation, size, productivity, and middleware. Although in the lower group, the average importance of this level is 3.7 on a scale of 1 to 5. The same lower risk level is documentation, size, and productivity. Documentation is generally attached to the record of each process so that it becomes the usual priority. Size is the project boundary. The size limitations are difficult to measure for current problems due to the rapidly changing developments and competition. While productivity is more on the ability of the team formed to complete the project of each feature.
6. Future work

The findings discussed in this study are important to be tested directly on the practice of software development projects. The order of variables can be a guide to the determination of weights on costing. The composition of weights can be tailored to the project and the policies used. Priority risk variables can also be used as a calculation of risk management weights in software development. The costing of this paper is more directed to software development companies based on orders. This is to be tailored to the needs of the company. For cloud computing companies with subscription system is not covered in this research.

7. Conclusion

The models that can be used in determining the cost of software development are at least 10 models as shown in Table 3. Each model has variables used for calculation. The total variables collected were 54. The variables are grouped/categorized based on similarity of meaning. The result is 15 main groups called 15 main variables. These variables are shown in Table 5. The variables have been sorted by the number of variable members in the model. The more models that have a particular variable then based on the higher the priority of the variable. The emerging variables are confirmed to the software business through questionnaires. There are two confirmations, namely the level of importance and the level of risk. Two consistent variables at the top are duration and effort. Both have the highest level of importance and are considered very risky if ignored. The sequence of variables can be used as an important consideration in calculating the cost of software development in practice.

References

[1] Niazi M, Mahmood S, Alshayeb M, and Majid A 2016 Toward successful project management in global software development *JPMA* **34** pp 1553–67.
[2] Potdar S M, Ingle A, Puri M, and Potdar M 2014 Factors influencing on cost estimation *Glob. J. Adv. Eng. Technologies* **3** pp 119–23.
[3] Barbara K et al 2009 Systematic literature reviews in software engineering – A systematic literature review *Inf. Softw. Technol.* **51** pp 7–15.
[4] Kitchenham B 2004 *Procedures for Performing Systematic Reviews*.
[5] Živadinović J, Medić Z, Maksimović D, Damnjanović A, and Vujčić S 2011 Methods of effort estimation in software *Int. Symp. Eng. Management And Competitiveness* **2011** pp 417–22.
[6] Ziauddin S, Kamal T, and Ziauddin Z 2012 An effort estimation model for agile software development *Adv. Comput. Sci. its Appl.* **2** pp 314–24.
[7] Rajper S and Shaikh Z A 2016 Software Development cost estimation : a survey *Indian J. Sci. Technol.* **9** pp 1–5.
[8] Besimsee W B 1996 *The PSP Size Estimating Approach* **1**.
[9] Spolsky A J 2008 *More Joel on Software* **320** (Apress). doi:10.1007/978-1-4302-0988-1
[10] Strickland D 1999 REVIC 11 : Converting the REVIC Model to COCOMO II *Center for Systems and Software Eng.* **35807**.
[11] Khatibi V and Jawawi D N A 2010 Software Cost estimation methods : a review *J. Emerg. Trends Comput. Inf. Sci.* **2** pp 21–9.

Acknowledgments

Our thanks to Informatics Department of Universitas Islam Indonesia for allowing us to use facilities, academic atmosphere, and funding support in this study.