

Are software cost-estimation models accurate?

Citation for published version (APA):

Kusters, R. J., Genuchten, van, M. J. I. M., & Heemstra, F. J. (1990). Are software cost-estimation models accurate? *Information and Software Technology*, 32(3), 187-190.

Document status and date:

Published: 01/01/1990

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Are software cost-estimation models accurate?

R J Kusters, M J I M van Genuchten* and F J Heemstra

The use of a model is one way to estimate a software development project. The paper describes an experiment in which a number of automated versions of estimating models were tested. During the experiment, experienced project leaders were asked to make a number of estimates for a project. This related to a project that had actually been carried out. On the basis of the differences found between the estimates and reality, it is concluded that no proof is given that the models can be used for estimating projects at an early stage of system development. Therefore, only limited confidence should be placed in estimates that are obtained with a model only.

software development, cost estimation, cost-estimation models

The use of a model is one way to estimate a software development project. Dozens of software cost-estimation models have been developed in the last 10 years and today many are on sale. Well known examples of estimation models include function-point analysis, COCOMO, Price, and Estimacs. The evaluation of a number of automated versions of estimating models is the subject of a study carried out by the 'Management of Software Development Projects' research group of Eindhoven University of Technology for the ISA-TMS department of Philips.

An important part of this study was an experiment in which 14 project leaders made a number of estimates using two estimation models. The goal was to evaluate two selected models in a semi-realistic situation. The experiment and its results are described in this paper.

SELECTION OF ESTIMATION MODELS

Dozens of estimation models are currently available on the market. Four models were selected on the basis of the following criteria:

- An automated version of the model must be available, up-to-date, and supported by a commercial supplier.

Philips International, Corporate O&E, Building VO-P, PO Box 218, 5600 MD Eindhoven, The Netherlands.

*Department BISA, Pav. D-3, Faculty of Industrial Engineering, University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands.

Paper submitted: 23 August 1989.

Revised version received: 10 October 1989.

- The model must be based on projects where information systems have been developed.
- The model must not use lines of code as input variables. An important requirement of the study is that the models must be applicable at an early stage of information system development. In the authors' opinion number of lines of code cannot be estimated accurately at this stage.

The preliminary selection provided four models for theoretical evaluation: Before You Leap (BYL)¹, Estimacs², SPQR³, and BIS/Estimator⁴. First of all, a list of requirements was defined in cooperation with representatives of possible, future users of the models. The authors think that the requirements should be met by a cost-estimation package if it is to provide useful support. Examples of the 20 or so requirements are:

- the completeness of the output
- the possibility of calibrating the model to its environment
- the use of information that becomes available during the development project
- the accuracy of the models
- the amount to which the known cost-drivers are taken into account
- the user friendliness of the models.

Needless to say, the importance of each requirement will vary according to each situation. A distinction was made, therefore, between mandatory and other requirements, and a weighing factor was allotted to each requirement. An extensive report on the theoretical evaluation is given elsewhere⁵. Here only the results are given. The BYL and Estimacs package achieved a satisfactory score and met all the mandatory requirements. SPQR and BIS/Estimator did not achieve a satisfactory score and did not meet all the mandatory requirements. SPQR scored unsatisfactorily as regards the mandatory requirement of calibration, as did BIS/Estimator for early applicability.

While a lot of criteria were tested in the theoretical study, clearly requirements like the accuracy and acceptance of the model by the possible, future users cannot be tested theoretically. Thus the second part of the study was undertaken: an experiment that involved 14 experienced project leaders.

OBJECTIVES

The goal of the experiment was to evaluate some aspects of the models in a semi-realistic situation. The experiments focused on three objectives. The interest was in the accuracy of the models and to determine whether these and similar models will be accepted in practice. Furthermore, could the number of lines of code be used at an early stage of development as a good indicator of the size of the product to be developed. It had to be ensured that it was right to use this as a selection criterion for the models to be examined (see the previous section).

Summarized, the objectives of the experiment were:

- to determine the accuracy of the estimate using models in a semi-realistic situation
- to determine whether these and similar models will be accepted by project leaders
- to determine whether the number of lines of code can be used at an early stage of development as a good indicator of the size of the product to be developed.

EXPERIMENTAL DESIGN

During the experiment, experienced project leaders were asked to make a number of estimates for a project. This related to a project that had actually been carried out. In this project a bonus system was developed for a sales organization. The project was described as if it was a starting project. The description consisted of three pages of text on the organizational environment, the functional specifications, and the goal of the project. Fourteen diagrams were added to this description, which included high-level dataflow diagrams, a diagram of the universe of discourse, the existing systems context (both hardware and software), and some use—create diagrams.

The preferable test set-up would have been one in which two different groups used only one package while another group acted as a control group. The size of the various groups would depend on the size of the variance to be expected. As this expected variance would be great, it follows that the size of the group would also have to be relatively large if reliable results were to be obtained. In this respect a total of 50 participating project leaders could be envisaged. Involving the necessary numbers of project leaders would lead to costs that would be out of all proportion to the importance of the study.

The first estimate of the effort and lead-time was made on the basis of the project leaders' knowledge and experience. From now on, this estimate shall be referred to as the manual estimate. Next, two estimates were made using the models selected. These estimates shall be called the model estimates. In conclusion, a final estimate was made on the basis of the project leaders' knowledge and experience together with the model estimates. Each estimate was evaluated directly using a questionnaire, and the experiment ended with a discussion session. The experiment was carried out with project leaders from a number of departments. Fourteen project leaders took part.

Table 1. Some results of experiment. Lead-time is given in months, Effort in man-months

| Variable | Mean (M) | Standard deviation |
|-------------------|----------|--------------------|
| Effort: | | |
| Manual estimate | 28.4 | 18.3 |
| BYL estimate | 27.7 | 14.0 |
| Estimacs estimate | 48.5 | 13.9 |
| Final estimate | 27.7 | 12.8 |
| Lead-time: | | |
| Manual estimate | 11.2 | 3.7 |
| BYL estimate | 8.5 | 2.4 |
| Final estimate | 12.1 | 3.4 |

RESULTS

The results of the experiment are described below. First, the results are presented. Next, the quality of the case used is evaluated. Finally, all the objectives of the experiment are considered in succession.

Results of estimates

Here the results of the experiment are presented. As has been seen from the description of the experiment, the 14 project leaders were asked to make four estimates for the 'bonus system' project. The results, i.e., the estimated effort and lead-time, of the four estimates (manual, BYL, Estimacs, and final estimate) are shown in Table 1.

As said before, the project has actually been carried out. The real effort and lead-time were 8 man-months and 6 months, respectively.

The questions that related to the models were also answered by the people who actually developed the system. Put into the models, this yielded the following results:

- effort with BYL: 18 man-months
- lead-time with BYL: 7.5 months
- effort with Estimacs: 54.4 man-months

The difference between the model estimates and reality is remarkable. In view of the system developers' familiarity with the development environment and their complete knowledge of the project, better model estimates would have been expected here. Furthermore, the model estimates of the system developers come close to the average model estimates obtained during the experiment.

Evaluation of case used

Before the results can be developed further, it is first necessary to see whether the case used is of sufficient quality. The participating project leaders were asked several questions on this subject. Asked whether the description gave more or less information than they are used to when making an estimate, 10 of the 14 project leaders said that the description given offered more information than they were used to in their everyday practice.

Asked about the subjects on which they would like to

have more information available, extra information about existing systems was mentioned five times, more information about the organization four times, and more extensive information about the required output of the software to be developed four times. During the concluding discussion the subject of the quality of the case presented was also dealt with. The general opinion was that the case gave more information than usual. Based on these answers, it is concluded that the description of the case was of sufficient quality to be useful in the experiment.

Objectives considered individually

As already mentioned, there were three objectives: to determine the accuracy of the models, the acceptance by possible, future users, and the usefulness of lines of code as an indicator. The results are discussed on the basis of these objectives, using both the quantitative (the statistical material obtained) and qualitative results (the answers to the open questions and the discussion results).

Accuracy

An estimation model must be expected to be accurate, in other words, that the mean and variance of the estimation errors obtained by using the model is small. In the experiment, however, the models were not calibrated⁷ with respect to either the environment in which the project was actually carried out or the environment in which the experiment was performed. The direct comparison of the mean estimate and reality is therefore not enough to judge the accuracy of the models.

In evaluating the accuracy of a model in the chosen experimental design the variance of the observations can be considered. The participating project leaders had a similar background. The spread in the model estimates point to the strength or weakness of the models. To be able to judge whether the variance is large or small, the variance of the manual estimate was taken as a reference point. The first conclusion is that the model estimates have not been shown to be poorer than the manual estimates. Looking at the figures, the variances in the model estimates are admittedly not statistically significant, but they are nevertheless lower than those of the corresponding manual estimates (see Table 1). A second conclusion can be drawn on the basis of the remarkable difference between the average estimation results for the BYL and Estimacs models. There is a difference of almost a factor of two, while the variances do not differ much from each other. This again underlines the need for calibration.

Acceptance

Knowing the scepticism of software developers towards cost models, it is important to find out whether they will accept a model as an estimation tool. The project leaders were therefore asked the following questions for the two evaluated models:

- Would you use this model in practice?

Table 2. Overview of answers to questions about acceptance

| Question | Answer | | |
|------------------------------------|--------|----|---------|
| | Yes | No | Missing |
| Would you use BYL? | 6 | 8 | 0 |
| Would you use Estimacs? | 7 | 5 | 2 |
| Would you use one of these models? | 11 | 2 | 1 |

- If one or more of these models were available to you would you use it or them for estimating software projects?

The answers to these questions are summarized in Table 2.

The view that the present method of drawing up an estimate was inadequate was virtually unanimous among the project leaders. Even though the quality of the present models was not great, they still considered it advisable to use them as a tool. In the project leaders' opinion, the greatest advantage attainable with such models at present was the possibility of using them as a means of communication or as a kind of check-list: 'The models draw your attention to a number of aspects which you would otherwise have overlooked'. Another advantage was the possibility of ascertaining the sensitivity of the cost-determining factors.

Volume

The question asked was whether the number of lines of code could be used at an early stage of system development as a measure for the volume of the system to be developed. The project leaders were asked to estimate the number of lines of source code of the software product to be developed. Function-point analysis — another method for determining the volume of a product — was used as a reference in the analysis. Both BYL and Estimacs models gave an estimate for the volume of the product, expressed in function points. The Ansari-Bradley-Freund test⁶ was used for the comparison:

H_0 : the relative variance of the volume, estimated in function points, is equal to that of the volume estimated in lines of code.

H_1 : the relative variance of the volume, estimated in function points, is smaller than that of the volume estimated in lines of code.

The estimated function points by both Estimacs and BYL were used for the test. In both cases, the zero hypothesis was rejected ($\alpha = 0.05$). The statistical material clearly shows, therefore, that lines of code is a poorer estimator for the volume of a product at an early stage of development than an available alternative, namely function points. This conclusion was further confirmed by the fact that only seven project leaders regarded themselves as capable of giving such an estimate of the volume in lines of code and that also during the discussion it emerged that the project leaders had absolutely no confidence in this measure.

CONCLUSIONS

The BYL and Estimacs models were evaluated in the experiment. The conclusions of the experiment were based on quantitative results and the opinions of the project leaders concerned. On the basis of the differences found between the estimates and reality, it is concluded that it has not been shown that the selected models can be used for estimating projects at an early stage of system development. This conclusion is strengthened by the fact that over half of the project leaders stated that the project description given offered more information than they were used to in their everyday practice.

All in all, the participants were not wildly enthusiastic about these packages, but they were nevertheless felt to be useful. If a model is used as a tool it will, in their opinion, mainly be valuable as a check-list and as a means of communication. On the basis of the striking difference between the average estimation results of the BYL and Estimacs models, simply using a model without adapting it to the environment in which it is used will not lead to accurate results. Calibration is essential⁷.

Now return to the title of this paper. The answer to the question stated is that in this study it has not been shown that the selected models are accurate and can be used for estimating projects. Other studies⁸⁻¹⁰ yielded similar results. Despite the flood of publications on software cost-estimation models, the authors are not aware of any empirical study that shows the ability of software cost-estimation models to predict effort and lead-time of projects accurately. Therefore, they believe that only limited confidence should be put in estimates that are obtained with a model only.

REFERENCES

- 1 **Gordon Group** *Before you leap, users guide* Gordon Group (1986)
- 2 **Computer Associates** *CA-Estimacs user guide, release 5.0* Computer Associates (July 1986)
- 3 *SPQR user manual* (1987)
- 4 **BIS** *BIS Estimator user manual, version 4.4* BIS Applied Systems Ltd, London, UK (1987)
- 5 **Heemstra, F J, Genuchten, M J I M and Kusters, R J** 'Selection of software cost estimation packages' *Research report no 36* University of Eindhoven, Eindhoven, The Netherlands (1989)
- 6 **Hollander, M and Wolfe, D A** *Non parametric statistics* John Wiley, New York, NY, USA (1973)
- 7 **Cuelenaere, A M E, van Genuchten, M J I M and Heemstra, F J** 'Calibrating a software cost estimation model: why and how' *Inf. Soft. Technol.* Vol 29 No 10 (December 1987) pp 558-567
- 8 **Abdel-Hamid, T K and Madnick, S E** 'On the portability of quantitative software estimation models' *Inf. Manage.* Vol 13 (1987) pp 1-10
- 9 **Kemerer, C F** 'An empirical validation of software cost models' *Commun. ACM* Vol 30 No 5 (May 1987)
- 10 **Mohanty, S N** 'Software cost estimation: present and future' *Soft. Pract. Exper.* Vol 11 (1981)

BIBLIOGRAPHY

- Boehm, B W** *Software engineering economics* Prentice Hall, Englewood Cliffs, NJ, USA (1981)
- Jones, T C** *Programming productivity* McGraw-Hill, New York, NY, USA (1986)