A Survey on Software Estimation in the Norwegian Industry

Kjetil Moløkken-Østvold[1, 2], Magne Jørgensen[1], Sinan S. Tanilkan[2],
Hans Gallis[1, 2], Anette C. Lien[1], and Siw E. Hove[1].

[1] Simula Research Laboratory, P.O.Box 134, 1325 Lysaker, Norway
{kjetilmo, magnej, hansga, anetteli, siweh}@simula.no

[2] Department of Informatics, P.O.Box 1080 Blindern, 0316 Oslo, Norway
sinant@ifi.uio.no

Abstract

This paper provides an overview of the estimation methods that software companies apply to estimate their projects, why those methods are chosen, and how accurate they are. In order to improve estimation accuracy, such knowledge is essential. We conducted an in-depth survey, where information was collected through structured interviews with senior managers from 18 different companies and project managers of 52 different projects. We analyzed information about estimation approach, effort estimation accuracy and bias, schedule estimation accuracy and bias, delivered functionality and other estimation related information. Our results suggest, for example, that average effort overruns are 41%, that the estimation performance has not changed much the last 10-20 years, that expert estimation is the dominating estimation method, that estimation accuracy is not much impacted by use of formal estimation models, and that software managers tend to believe that the estimation accuracy of their company is better than it actually is.

1. Introduction

Over the past two decades, several research surveys have focused on software project effort and schedule estimation [1-10]. This is important, since an unbiased insight is essential in order to help the industry to make more accurate estimates. Essential data to obtain is, e.g., how estimates are made, what factors motivate the choice of estimation methods and the current level of estimation accuracy.

This paper attempts to provide an assessment of the current situation related to software effort estimation in Norway. Most of the previous surveys were conducted in the eighties and early nineties. Types of hardware, programming languages, business relations, development processes, clients and software companies have changed a lot since then. Further, the surveys were conducted in larger countries than Norway, such as the United States and the United Kingdom.

Section 2 provides an account of the previous research that motivated our research questions, which are presented in Section 3. Section 4 describes the methods used to collect and analyze the data presented in Section 5. Section 6 provides a discussion, which is summarized in Section 7.

2. Previous Surveys

The surveys conducted over the past 20 years have had varying areas of interest; some of them were conducted at company level, others at project level. As presented in a recent review [11], those areas that have attracted most attention are how companies estimate projects, how important effort estimation is perceived and what their current level of accuracy is. In this paper, we include only previous research that has been subject to peer review, or in which the research method is, at least partially, described. For that reason, the frequently quoted CHAOS Report published by the Standish Group is excluded. We discuss the validity problems of that report in [12].

2.1. Frequency and magnitude of effort and schedule overruns

The topic of estimation accuracy has been surveyed by Jenkins [3], Phan [8], Bergeron and St-Arnaud [1], Heemstra and Kusters [13], Lederer and Prasad [14, 15] and Sauer and Cuthbertson [10]. These studies have addressed either the frequency of overruns, or the average estimation accuracy, or both.

A summary of some of the results from these surveys is displayed in Table 1. A blank space in the table indicates that this information was not reported in the survey.

Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)
1530-1435/04 $ 20.00 IEEE

Table 1: Previous surveys on estimation accuracy

Study (first author)	Jenkins [3]	Phan [8]	Heemstra [2]	Lederer [14]	Bergeron [1]	Sauer [10]
Year of study's first publication	**1984**	**1988**	**1989**	**1991**	**1992**	**2003**
Cost overrun	34% (median)	33% (mean)			33% (mean)	18% (mean)
Project used more than estimated effort	61%		70%	63%		59%
Project used less than estimated effort	10%			14%		15%
Schedule overrun	22% (mean)					23% (mean)
Project completed after schedule	65%		80%			35%
Project completed before schedule	4%					3%

These surveys indicate that most projects (60-80%) are completed over estimated effort and/or schedule. The magnitude of effort overruns reported in most of the surveys [1, 3, 8] is between 30 and 40%. This has also been supported by several case studies, e.g. [16, 17].

Sauer and Cuthbert found a lesser magnitude of overruns [10], but this may have been affected by the self-selecting sample in that survey. Bergeron and St-Arnaud [1] found that 58% of the projects surveyed had cost overruns of more than 20%. Phan [8] reports on overruns at an organizational level that is not directly comparable with the other surveys. He found that cost overruns occurred always in 4%, usually in 37%, sometimes in 42%, rarely in 12%, and, never in 4% of the 191 organizations surveyed. Similarly, schedule overruns occurred always in 1%, usually in 31%, sometimes in 50%, rarely in 15%, and, never in 2% of the organizations.

2.1. Choice of estimation method

In our review [11], we found five different surveys that addressed the choice of estimation method [1, 2, 5, 18, 19]. Due to the different metrics used, and the different categorizations, it is difficult to compare the surveys. A further complicating factor is that the respondents may have interpreted pre-defined categories differently. We have grouped the estimation methods into three main categories: expert judgment-based methods, model-based methods and "other". Model-based methods include formal estimation models such as COCOMO, Use-Case-based estimation, FPA-metrics or other algorithm driven methods. In the category of "other" there are methods that are not "pure" estimation methods, e.g., capacity-related and price-to-win-based methods, and methods than can be used in combination with other models (top-down and bottom-up). An overview is presented in Table 2. A blank space in the table indicates that this alternative was not an option in the survey.

It is essential to note that the projects surveyed may have been estimated by a combination of two or more different methods, e.g. model and expert-based. The extent to which such combination methods were used was not reported in any of the surveys. We were therefore unable to draw conclusions about the possible beneficial effect of combining methods in the review.

'Price-to-win' is listed as an estimation method, but we believe that most managers would not report this as an estimation method even if "price-to-win" had an impact on their estimates. This may be the case because managers are not aware of the effect of client expectations on effort estimates [20]. They may feel that it is not an estimation method, or may believe that they should not be affected by the "price-to-win".

Lederer and Prasad [18] applied a different approach from the surveys displayed in table 2. Instead of asking what methods were used, they asked how estimates were influenced. Respondents rated alternatives on a five-point Likert scale (min=1, max=5). They had eight different categories, and the alternative "Comparison to similar, past projects based on personal memory" scored highest (average 3,77), while the alternative "A complex statistical formula (such as multiple regression, differential equations, etc.)" scored lowest (average 1.49).

Table 2: Previous surveys on estimation methods

Study (first author)		McAulay [5]	Heemstra [2]	Wydenbach [19]	Bergeron [1]
Year of study's first publication		**1987**	**1989**	**1995**	**1992**
Estimation Methods		**Percentage used (more than one method possible)**			**Importance (1-4, 4=most important)**
Expert based	Expert consultation		26%	86%	1.8
	Intuition and experience	85%	62%		3.3
	Analogy		61%	65%	2.5
Model based	Software Cost Models	13%	14%	26%	1.3
Other	Price-to-win		8%	16%	1.2
	Capacity related		21%	11%	1.6
	Top-down			13%	1.4
	Bottom-up			51%	2.4
	Other	12%	9%	0%	1.5

2.2. How important is estimation accuracy perceived?

The only survey that directly addressed the importance of estimation accuracy was conducted by Lederer and Prasad [15]. On a five-point Likert scale (min=1, max=5), the average importance rating reported by managers was 4.17. Although that survey found that managers perceived estimation as important; this does not entail that projects are estimated as a matter of course.

Related to this finding are surveys on the proportions of software organizations that estimate costs. Heemstra [2] reports that 65% and McAulay [5] report that 95% of the organizations estimate projects as a rule. Wydenbach and Paynter [19] report that 88% of the companies estimate at least half of their projects. The survey by Lederer and Prasad studied the estimation percentage at project level [15]. They found that 87% of the companies' large projects were estimated.

Moores and Edwards [6] found that 91% of the responding managers answered 'yes' to the question 'do you see estimation as a problem?', while only 9% answered 'no'. An interesting finding is that the managers reported that the accepted level of estimation accuracy was typically +/- 20%.

A finding that supports the belief that estimation is perceived as a problem was reported by Addison and Vallabh [21]. They found that the risk factor viewed as most problematic by software professionals was 'unrealistic schedules and efforts'.

In combination, these surveys indicate awareness that estimation is both a problem and an important activity. There are, however, many other important software development activities, e.g., contract negotiations with clients. An appropriate approach to reveal how important the companies regard effort estimation accuracy in practice is to investigate the organizations' actual focus on improving estimation processes and the effort expended on achieving accurate estimates. We have found no such studies.

3. Research Questions

In our survey, we wanted to investigate the same questions as in previous surveys, with some additions. While other surveys have focused on either company or project level, we wanted to focus on both, in order to compare responses.

Our topics of interest are divided into several research questions.

RQ1: What is the frequency and magnitude of effort estimation overruns?

RQ2: What is the frequency and magnitude of schedule estimation overruns?

RQ3: Does project size affect effort estimation accuracy or bias?

Choice of estimation method has been related to the study of effort estimation accuracy. There have been research results "in favor" of different estimation methods (e.g., expert-based vs. model-based), but according to a recent review, such evidence is inconclusive [22]. In our survey, we also wanted to investigate the potential benefits of combining different estimation approaches. Such combination has been advocated in other studies [1, 23].

RQ4: To what extent are different estimation methods (expert-based, model-based or combinations) used in the industry?

RQ5: Does choice of estimation method affect effort estimation accuracy or bias?

Many of the previous surveys, e.g. [13, 18] have been aimed at in-house development in large corporations that deal in insurance, banking, manufacturing etc. In our survey, we wanted to address the situation in consultancy companies that develop tailored solutions for clients, as well as in-house development in other software and/or telecommunications companies.

RQ6: Are there differences in estimation accuracy or bias between companies that develop projects internally and those that develop for clients?

An important aspect of effort estimation is how senior managers perceive the situation. If they are unaware of any problems, or if they believe that the problem is within an acceptable range, such as +/- 20% [6], it is unlikely that measures will be taken. We also wanted to investigate the possible reasons for selecting a particular method, as well as investigating how important effort estimation is perceived.

RQ7: How do senior managers perceive the company's level of estimation skill?

RQ8: On what basis is an estimation method selected?

RQ9: How important does the companies perceive estimation as being, in comparison with other aspects of development?

4. Method

The survey was conducted in Norway from February to November 2003.

4.1. The participating companies

In order to ensure a representative sample, stratified random sampling [24] from the population of Norwegian software development companies was used. The companies were categorized into different strata based on revenue and number of employees. This was based on different Norwegian sources, e.g. [25]. This is a reasonable approach, since we were going to investigate a limited number of companies and wanted to ensure that we had companies that represented different types of organization, such as software houses developing products for the mass market, contractors who develop for clients and the internal development departments of large companies. We also wanted companies of different sizes, both small (<10 employees), medium (between 25 and 100 employees) and large (>100 employees).

Each organization was contacted by phone and the study presented to them. If they agreed to participate, they were given time to prepare before they were visited by our researchers.

The unit of investigation was either the entire company or a specific department (the latter in the case of very large organizations with more than 1000 employees). We will, however use the term company for our unit of research in this paper. The companies that participated had between 10 and 750 employees, with an average of 141. Five of the companies developed projects to be used in-house, while two developed products for sale to the mass market. Out of the eleven companies who developed solutions for clients, nine had mainly private clients, while two had mainly public clients.

It is essential to note that the market situation in Norway at the time of the survey (2003) was very competitive, and many companies sustained losses. Senior managers, however, said that they coped with the situation by lowering their hourly rates (for those developing for clients) instead of underestimating projects.

About half of the companies' development was new projects (57.1%), while the rest was maintenance/re-engineering (38.5%) or combination projects (4.4%). The percentage of employees who owned shares or stock options in their own company ranged from zero to 100 percent, with an average of 43.5 %. For those companies that used contracts, either with clients or internally, most relied on fixed price (52.5%), while 22.5% were paid per hour and 25% were based a combination of these alternatives.

For the specific projects, the mean effort was 3124.5 man-hours, while the median was 1175 man-hours.

4.2. Data collection and analysis

In all companies we interviewed one or two senior manager(s) at company or, in the case of large corporations, department level. These managers were asked general questions about such matters as staff size, type of assignments, project resolution, estimation method and assessment of estimation accuracy.

We also interviewed the project managers (one or two for each project) of 52 different projects. These projects were selected by the companies themselves. To avoid selection bias, the companies were asked to submit their most recently completed projects. The only criteria were that they had to contain a workload of at least 100 man-hours. This was done in order to ensure that no small change requests and other one-developer projects were submitted. This is in line with previous surveys, in which "trivial tasks routinely handled without effort estimation" were also filtered out [18].

We collected data via personal interviews, which yields data of high quality and ensures that ambiguities are resolved [24]. This was especially important in our survey, since there may be variations in the use of, and interpretations of, terms related to estimation methods. It also allows the respondents to add valuable information that it is not possible to include when completing a predefined questionnaire. Another point in favor of this approach is that our personal involvement indicates a seriousness of intent to the participants, and this may increase the likelihood of obtaining serious contributions from them. The main limitation of the approach is that it is time-consuming and hence prevented us from investigating as many companies and projects as would be possible by using mailed questionnaires. Out of the previous surveys, presented in Section 2 and the recent review [11], only one used personal interview as an approach [3].

Each interview lasted between 30 and 70 minutes. All researchers signed a confidentiality agreement at each company. The respondents were informed that their responses were anonymous, and that no feedback about the respondents' answers was to be reported to outsiders or to company managers.

All interviews were taped. Following data collection, results from the questionnaires and tapes were entered into databases and processed by independent personnel who had no stake in the research. This is especially important, because it ensures that there are no biases regarding how possibly ambiguous data, such as the estimation approaches, are classified.

In cases where the participants reported that they followed a company-defined estimation approach, we asked them to provide thorough descriptions. After all of the interviews had been conducted, the independent analyst listened to the tapes and categorized the customized methods according to the predefined categories presented in previous sections (model/expert/combination).

In order to assess estimation accuracy, both related to effort and schedule, the Balanced Relative Error (BRE) [26, 27] was used. It is calculated as:

$$BRE = \frac{|x - y|}{\min(x, y)},$$

x = actual and y = estimated. (1)

The BRE is different from the more common MRE (Magnitude of Relative Error) measure [28]. MRE is calculated as:

$$MRE = \frac{|x - y|}{x},$$

x = actual and y = estimated value. (2)

Even though MRE is the most widely used measure of estimation accuracy [29], one must be aware that it has unfortunate properties [26, 30]. The main concern for our case is the fact that underestimated and overestimated projects are weighted unevenly. The BRE, as its name indicates, is a more balanced measure.

However, it is important to note that the difference between using MRE and BRE does not have a large impact on statistical tests based on medians, which are mainly used in this paper.

Estimation bias is calculated using a similar equation. The only difference is that the absolute value is not used in order to reveal an eventual direction of inaccuracy.

$$BREbias = \frac{(x - y)}{\min(x, y)},$$

x = actual and y = estimated value. (3)

In previous surveys in which estimation accuracy has been calculated, other measures have been used. Bergeron and St-Arnaud [1], used a formula recommended by Conte et al.[28], which is calculated as:

$$Accuracy = \frac{|x - y|}{y},$$

x = actual and y = estimated value. (4)

They argue that this measure is more meaningful, since profit or loss should be calculated on the basis of expected cost by most project managers.

The wording in the paper reporting Phan's survey, e.g., "Software projects, on average, overrun planned costs by 33%" [31], implies a use of that accuracy measure. It is also possible that it is used in the survey conducted by Jenkins, which describe vague terms, such as "under-run and over-run", with no indications of how these were calculated [3].

4.3. What is an estimate?

Even though all previous surveys on software estimation have differences in method, and study different aspects of the process, they tend, with some

Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)
1530-1435/04 $ 20.00 IEEE

exceptions [1], to treat a software estimate as a single fixed value. During the course of our research, however, we have noticed that software projects often have several estimates [32]. This aspect, and how this poses challenges to estimation models, has also been addressed by Edwards and Moores [33].

First of all, the estimate often changes over the course of a project, depending on the *stage* at which the estimate is made. For example, a project can have an early estimate, based on vague requirements, a planning estimate based on a detailed requirement specification, and one (or more) re-estimates during the course of development. These estimates, all for the total effort of the project, may or may not be entirely different in magnitude.

Second, a very important factor is who the estimate is for, or communicated to. A single project may have two estimates at the planning stage: one that is used internally in the project team, e.g., by project managers, and another that is used for clients, whether they are internal or external.

The problem becomes even clearer when we find papers that describe bidding strategies, such as "price to win" as an estimation method! This was been done in several textbooks and research papers.

For this reason, it is obvious that one project can operate with, two, three or even more different estimates during development. This may often be unproblematic for developers and project managers who differentiate between these measures, but pose a significant challenge to scientific researchers.

When conducting research on the accuracy of estimates of software projects, it is necessary to differentiate between different types of estimates. What estimate(s) to use depends on the focus of research. If the goal is to investigate the estimation accuracy of professionals in a company, as it is in this paper, it is meaningful to use the *most likely* estimates at the *planning stage*, instead of, for example, early estimates communicated to clients. The latter may be affected by factors that have nothing to do with estimation skill, such as market competition.

In our survey, we collected data on all available estimates for each project. Some had only one, while others had as many as six different estimates. Nonetheless, this is not problematic when taken into account at the time of data analysis.

5. Results

The results are structured according to the research questions presented in Section 3. We include both responses from senior managers at company level and the responses from the project managers, based on specific projects. The responses from the senior managers are often at a general level, and reflect attitudes, beliefs and prioritizations. The data from specific projects, on the other hand, are based on recorded data submitted by the project managers. An overview of the project data is displayed in Appendix I.

The required data was available in 44 out of the 52 projects we analyzed. The other projects were discarded because they involved fewer than 100 man-hours (1), or most estimation and development work had been done by outside consultants (2), or the project managers had not kept accurate track of estimated and/or actual effort (5).

Of the 44 projects whose inclusion in the analysis was meaningful, two (5%) was aborted without being completed, and five (11%) were completed on time, on estimated effort and with required functionality. The rest (84%) were challenged with respect to effort overruns, schedule overruns, functionality, or a combination of these (see Appendix I).

5.1. What is the frequency and magnitude of effort estimation overruns? (RQ1)

Out of the 42 projects that were completed, 32 (76%) had effort overruns. Two projects (5%) ended up on target, while eight (19%) used less effort than indicated.

The estimation bias (BREbias) ranged from -0.88 (estimated effort 88% higher than actual effort) to 1.91 (actual effort almost three times the estimated effort). The distribution of project effort estimation bias is shown in Figure 1.

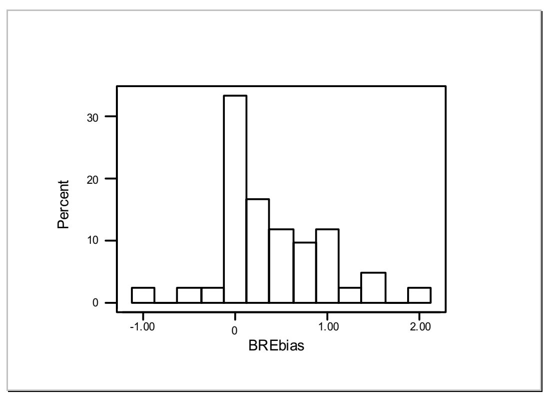

Figure 1: Distribution of effort estimation BREbias

The average effort BREbias was 0.41, while the median effort BREbias was 0.21. This corresponds to a mean cost overrun of 41%, while the median cost overrun was 21%.

5.2. What is the frequency and magnitude of schedule estimation overruns? (RQ2)

Regarding schedule, 26 projects (62%) had overruns, while 15 (36%) were completed on schedule, and one (2%) before schedule. Mean schedule overrun was 25% (BREbias 0.25), while the median overrun was 9% (BREbias 0.09).

5.3. Does project size affect effort estimation accuracy or bias? (RQ3)

To see whether project size could affect effort estimation accuracy, we divided the 42 projects into halves, based on actual effort. Mean BREbias for the 21 largest and the 21 smallest projects were 0.52 and 0.30, respectively. An Anderson-Darling test reveals that the samples are not normally distributed, but since the samples sizes are equal (both 21) and the variance is similar [34], a statistical t-test was used to examine the difference in mean accuracy values between large and small projects. The results for accuracy (BRE) and bias (BREbias) are displayed in Table 3.

Table 3: Effort estimation accuracy and bias by project size.

	Large	Small	p-value
BRE	0.52	0.46	0.34
BREbias	0.52	0.30	0.11

As indicated in the table, the large projects may be more prone to under-estimation (higher BREbias).

5.4. To what extent are different estimation methods used in the industry? (RQ4)

When asked about the kinds of estimation approaches that were used in the projects, 13 of the companies answered that they relied 100% on expert estimation. Three of the companies used a combination of expert judgment and estimation model 100% of the time, while two companies sometimes used expert estimation and sometimes a combination of expert and model. Out of the five companies that used a combination of model and expert, four of the companies used use-case based estimation models [35-37].

Of the 44 projects we analyzed (including two that were aborted), 37 (84%) of the managers reported that they relied entirely on expert estimation, while seven of the managers reported that they used a combination (16%) of expert and model estimation. Out of these seven, six stated that they relied on a use case-based estimation method tailored to their company’s types of projects and historical data. The last project used a combination of expert judgment and a company-defined estimation model.

5.5. Does choice of estimation method affect estimation accuracy or bias? (RQ5)

An Anderson-Darling test of the samples when divided by estimation method excludes normality. Since the sample sizes are unequal and the variances are unequal a non-parametric Kruskal-Wallis test on effort estimation medians is used [34]. The analysis on differences in median effort estimation accuracy (BRE) and bias (BREbias) grouped by choice of estimation method is displayed in Table 4.

Table 4: Effort estimation accuracy and bias by estimation method.

	Expert	Combination	p-value
BRE	0.30	0.35	0.84
BREbias	0.22	0.17	0.45

Based on our data, we find no indication that the use of estimation models in combination with experts leads to more accurate or unbiased estimates compared with expert estimates alone.

5.6. Are there differences in estimation accuracy or bias between companies that develop projects internally and those that develop for clients? (RQ6)

Out of the 42 projects that were completed, 11 were developed in-house (internal), while 31 were developed on contract for a client (external).

The samples do not follow a normal distribution, and there are differences in sample size and variance, as in the last subsection. The results of a Kruskal-Wallis test on differences in median accuracy values of internal and external development are displayed in Table 5.

Table 5: Effort estimation accuracy and bias by client.

	Internal	External	p-value
BRE	0.07	0.42	0.10
BREbias	0.05	0.35	0.07

This indicates that projects that are developed internally may be more accurate and less biased related to effort estimation than externally developed projects.

5.7. How do senior managers perceive the company's level of estimation skill? (RQ7)

In order to get an understanding of the managers' view of their companies' estimation accuracy, we asked them to provide us with their own assessment of the company's mean effort estimation accuracy. Responses ranged from 10% overestimation to 50% underestimation, with an average for all companies of 15.9% underestimation.

The managers were also asked about the outcome of their projects. They were asked to categorized projects conducted during the past year into three different categories: (i) "success" – delivered on schedule and on effort, with the required functionality, (ii) "challenged" – failure to meet either schedule, effort or functionality requirements, or (iii) "aborted" – projects that were either stopped or underwent major revision.

The average response was 45% for "success", 49% for "challenged" and 6 % for "aborted". Note that these calculations are not adjusted for company size or number of projects.

There was consequently a difference between the estimation accuracy observed in our study and the beliefs of the managers. This difference is analyzed and discussed in Section 6.

5.8. On what basis is an estimation method selected? (RQ8)

In order to identify a possible rationale for the use of a particular estimation approach, the managers were asked to rate several possible reasons for choosing an estimation method. Each reason had to be given a rating from one to five, where five was the maximum. A summary of the average ratings is given in Table 6.

Table 6: Reasons for choice of estimation methods

Reason for choosing method	*Rating (1-5)*
Estimator has had success with method	4.2
Consultant advice	2.7
Thorough testing	2.7
Structured analysis	2.1
Lectures at universities/colleges/courses	1.8
Review of other companies experiences	1.7
Market popularity	1.2

Out of all possible reasons, only one scored above average (3) on the ratings. This was that the estimator had had previous success with the method. A total of thirteen companies gave this reason the highest possible rating (5). The managers were also given the opportunity to describe important, non-predefined reasons for selecting an estimation method. Among the reasons provided were "using function point methods has been shown to be inefficient" and "the method used is good for persuading senior management to approve a project"!

5.9. How important does the organization perceive estimation as being, in comparison with other aspects of development? (RQ9)

When asked about how important effort estimation was viewed in comparison with other development aspects, the managers provided free text responses. Of the eighteen companies, fourteen answered that estimation was very important, extremely important or most important.

6. Discussion

The surveys available for comparison with our data are mainly between ten and twenty years old. Despite this, and other differences, our results are similar to those of the previous surveys. This suggests that the estimation methods and performance of software companies has not changed much since the 1980s. A summary of our results is displayed in Table 7.

Table 7: Summary of Results

Cost overrun	41% (mean), 21% (med)
Project used more than estimated effort	76%
Project used less than estimated effort	19%
Schedule overrun	25% (mean), 9% (med)
Project completed after schedule	62%
Project completed before schedule	2%

The frequency and magnitude of effort and schedule overruns in our survey (RQ1 and RQ2) were similar to those reported in previous surveys (see Section 2). We also observed that there may be a tendency that larger projects are more prone to underestimation than smaller projects (RQ3).

Our observations are similar to the results of the other surveys with respect to choice of estimation method (RQ4). In our 44 projects, 16% relied on an estimation model in combination with expert estimation, while 84% relied entirely on expert estimation. If we take into consideration that cost models are often complicated to use, and have not proved to be superior to expert-based methods [22], this lack of use of formal estimation models is not

Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)
1530-1435/04 $ 20.00 IEEE

surprising. Those of our companies who used models relied on use-case based models in combination with expert judgment, i.e., none of the projects relied on estimation models as the only estimation method. We were unable to provide conclusive results as to whether use of a combination of expert and model estimation was beneficial (RQ5), but this may have been due to the small sample of combination projects.

It seems as if there are differences between those who develop in-house, and those who develop for clients, regarding effort estimation accuracy and bias (RQ6). A possible reason for this observation is that in-house developers have closer proximity to the customer and more stable system properties, such as requirements, platform and implementation language.

The managers believed that there were almost as many "successful" (45%) as "challenged" (49%) projects, and only 6% aborted projects (RQ7). What we observed, on the other hand, was that only 11% were completed on schedule, on effort and with full functionality, while as much as 84% were "challenged". A potential reason for this is that a project that overruns its effort or schedule estimate by a few hours or days will not be interpreted as challenged by managers. If we define projects with a 10% overrun as "successful", the percentage of successful projects rises to 25%, while the percentage of challenged projects falls to 70%. If the success range is increased to include 25% overruns, the percentage of successful projects is 43%, while the percentage of challenged projects is 52%. Moores and Edwards [6] have previously described that most managers are comfortable with a level of accuracy around +/- 20%. The managers also believed that the magnitude of effort overruns was less (15.9%) than what we observed (mean 41% and median 21%).

Regarding choice of estimation method, the senior managers were quite aware of how the projects were estimated. However, it did not seem that they were concerned with analyzing the performance of their methods (RQ8). When asked about reasons for choosing a method, the only alternative with an above average rating was "the estimator has had success with the method". On the other hand, the managers were very clear when responding on the importance of effort estimation in comparison to other topics (RQ9), since 14 out of 18 companies rated the topic to be very important, extremely important or most important.

A possible connection between the three last observations may be that even if the managers perceive the topic of estimation to be important (RQ9), a lack of structured analysis of the situation (RQ8) may mislead them into believing that their estimates are more accurate than they actually are (RQ7).

6.1. Threats to validity

The most obvious threat to internal validity is the sample size, which is small when applying statistical inference-based analysis methods. The small sample size is a results of our labor-intensive data collection method based on interviews and the belief that it was more important to get high quality, in-depth information about a few projects instead of lower quality information about more projects, applying a questionnaire-based study method. Personal interviews helped to clarify numerous ambiguities, and to discard projects that did not retain accurate information on important aspects, such as estimated effort.

Most projects claimed to have met requirements related to functionality, but it is important to note that this was based on the managers own responses, since we did not have access to the end-users.

When evaluating the generalizability of the results one must consider that this is a survey of Norwegian companies. Norwegian companies are on average smaller (both in number of employees and revenue) and complete smaller projects than companies in countries such as the United States, the United Kingdom and Canada. There may also be business cultural issues that reduce the generalizability of the results.

Our results are, however, similar to those reported by the previous surveys presented in Section 2. This is an indication that the sample is probably not biased in any particular direction.

An important factor in general when measuring estimation skill in companies is that we can only analyze projects that were approved by managers or that won contracts. Such projects may be selected because they have optimistic estimates. At the same time, realistically estimated projects may be turned down because they seemed too expensive. This may be a "winners curse" that affects estimation accuracy in completed projects [38] which is reflected in surveys such as this.

7. Summary

The observations in our survey were similar to those reported by other researchers. It seems that choice of estimation method and level of estimation accuracy and bias are stable, being independent of year, technology and location. A possible reason for this observation is that alternative estimation approaches have failed to provide evidence that their use increase estimation accuracy.

It also appears that the focus on analysis of estimation accuracy is low in the Norwegian software industry. This may lead to an assessment of estimation skill that is misleading, which in turn may hamper improvement and education.

Further research will be to monitor the development in Norway over time, to see how different process improvement efforts or technological changes may affect estimation performance. We would also like to expand the survey, with replications in other countries.

Acknowledgements

This research was funded by the Research Council of Norway under the project INCO. Thanks to Chris Wright for valuable comments.

References

1. Bergeron, F. and J.-Y. St-Arnaud, *Estimation of Information Systems Development Efforts: A Pilot Study*. Information & Management, 1992. **22**: p. 239-254.
2. Heemstra, F.J., *Software cost estimation.* Information and Software Technology, 1992. **34**(10): p. 627-639.
3. Jenkins, A.M., J.D. Naumann, and J.C. Wetherbe, *Empirical Investigation of Systems Development Practices and Results.* Information & Management, 1984. **7**: p. 73-82.
4. Lederer, A.L., et al., *Information System Cost Estimating: A Management Perspective.* MIS Quarterly, 1990. **14**(2): p. 159-178.
5. McAulay, K. *Information Systems Development and the Changing Role of MIS in the Organisation*. in *First New Zealand MIS Management Conference*. 1987. Wellington.
6. Moores, T.T. and J.S. Edwards, *Could Large UK Corporations and Computing Companies Use Software Cost Estimating Tools? - A Survey.* European Journal of Information Systems, 1992. **1**(5): p. 311-319.
7. Paynter, J. *Project estimation using screenflow engineering*. in *International Conference on Software Engineering: Education and Practice*. 1996. Dunedin, New Zealand.
8. Phan, D., *Information Systems Project Management: an Integrated Resource Planning Perspective Model*, in *Department of Management and Information Systems*. 1990, Arizona: Tucson.
9. Purvis, M.K., S.G. MacDonell, and J. Westland, *Software Metrics in New Zealand: Recent Trends.* New Zealand Journal of Computing, 1994. **5**(2): p. 13-21.
10. Sauer, C. and C. Cuthbertson, *The State of IT Project Management in the UK 2002-2003*. 2003, Templeton College, University of Oxford.
11. Moløkken-Østvold, K. and M. Jørgensen. *A Review of Surveys on Software Effort Estimation*. in *2003 ACM-IEEE International Symposium on Empirical Software Engineering (ISESE 2003)*. 2003. Frascati, Monte Porzio Catone (RM), ITALY: IEEE.
12. Jørgensen, M. and K. Moløkken-Østvold, *How Large Are Software Cost Overruns? Critical Comments on the Standish Group's CHAOS Reports.* Submitted to Communications of the ACM (Can be dowloaded at: http://www.simula.no/publication_one.php?publication_id=711), 2004.
13. Heemstra, F.J. and R.J. Kusters. *Controlling Software Development Costs: A Field Study*. in *International Conference on Organisation and Information Systems*. 1989. Bled, Yugoslavia.
14. Lederer, A.L. and J. Prasad, *Causes of Inaccurate Software Development Cost Estimates.* Journal of Systems and Software, 1995(31): p. 125-134.
15. Lederer, A.L. and J. Prasad, *Information systems software cost estimating: a current assessment.* Journal of Information Technology, 1993(8): p. 22-33.
16. Boehm, B., *Software Engineering Economics*. 1981, Englewood Cliffs, NJ: Prentice-Hall.
17. Kitchenham, B., et al., *An empirical study of maintenance and development estimation accuracy.* Journal of Systems and Software, 2002. **64**(1): p. 57-77.
18. Lederer, A.L. and J. Prasad, *Nine management guidelines for better cost estimating.* Communications of the ACM, 1992. **35**(2): p. 51-59.
19. Wydenbach, G. and J. Paynter, *Software Project Estimation: a Survey of Practices in New Zealand.* New Zealand Journal of Computing, 1995. **6**(1B): p. 317-327.
20. Jørgensen, M. and D.I.K. Sjøberg, *The impact of customer expectation on software development effort estimates.* International Journal of Project Management, 2004. **22**(4): p. 317-325.
21. Addison, T. and S. Vallabh. *Controlling Software Project Risks - an Empirical Study of Methods used by Experienced Project Managers.* in *SAICSIT 2002*. 2002. Port Elizabeth, South Africa.

22. Jørgensen, M., *A Review of Studies on Expert Estimation of Software Development Effort.* Journal of Systems and Software, 2004. **70**(1-2): p. 37-60.
23. Höst, M. and C. Wohlin. *An Experimental Study of Individual Subjective Effort Estimations and Combinations of the Estimates*. in *Proceedings the 20th International Conference onSoftware Engineering*. 1998. Kyoto, Japan.
24. Cozby, P.C., *Methods in behavioral research*. 5th ed. 1993, Mountain View: Mayfield Publishing Company.
25. HegnarOnline, *Kapital DATAs 1500 største*. 2000.
26. Jørgensen, M. and D. Sjøberg, *An effort prediction interval approach based on the empirical distribution of previous estimation accuracy.* Journal of Information and Software Technology, 2003. **45**(3): p. 123-136.
27. Miyazaki, Y., et al., *Method to estimate parameter values in software prediction models.* Information and Software Technology, 1991. **33**(3): p. 239-243.
28. Conte, S.D., H.E. Dunsmore, and V.Y. Shen, *Software Engineering Metrics and Models*. 1986, Menlo Park: Benjamin-Cummings.
29. Briand, L.C. and I. Wieczorek, *Resource modeling in software engineering*, in *Encyclopedia of Software Engineering*, J. Marciniak, Editor. 2001, Wiley.
30. Stensrud, E., et al. *An empirical validation of the relationship between the magnitude of relative error and project size*. in *Eighth IEEE Symposium on Software Metrics.* 2002: IEEE.
31. Phan, D., D. Vogel, and Nunamaker, *The Search for Perfect Project Management*, in *Computerworld*. 1988. p. 95-100.
32. Jørgensen, M., *How much does a vacation cost? or What is a software cost estimate?* ACM Software Engineering Notes, 2003. **28**(6): p. 30.
33. Edwards, J.S. and T.T. Moores, *A Conflict Between the Use of Estimating and Planning Tools in the Management of Information Systems.* European Journal of Information Systems, 1994. **3**(2): p. 139-147.
34. Cohen, J., *Statistical power analysis for the behavioral sciences*. 1969, New York: Academic Press, Inc.
35. Anda, B., E. Angelvik, and K. Ribu. *Improving Estimation Practices by Applying Use Case Models*. in *4th International Conference on Product Focused Software Process Improvement, December 9 - 11.* 2002. Rovaniemi, Finland.: Springer-Verlag,.
36. Anda, B., et al. *Estimating Software Development Effort Based on Use Cases - Experiences from Industry*. in *4th International Conference on the Unified Modeling Language (UML2001), October 1-5, 2001.* 2001. Toronto, Canada.: Springer-Verlag.
37. Anda, B. *Comparing Effort Estimates Based on Use Case Points with Expert Estimates*. in *Empirical Assessment in Software Engineering (EASE 2002). April 8-10, 2002.* 2002. Keele, UK.
38. Gilley, O.W., G.V. Karels, and e. al., *Uncertainty, experience and the 'winners curse' in OCS lease bidding.* Management science, 1986. **32**(6): p. 673-682.

Appendix I: Survey data

Nr	Client	Method	Estimate (hrs)	Actual	BREbias	Estimate (days)	Actual	BREbias	Functionality
1	External	Expert	330.0	319.0	-0.03	235	235	0.00	100%
2	Internal	Expert	560.0	1000.0	0.79	84	140	0.67	110%
3	Internal	Expert	300.0	600.0	1.00	156	156	0.00	100%
4	Internal	Expert	700.0	1400.0	1.00	182	224	0.23	100%
5	External	Expert	4227.0	5170.0	0.22	98	98	0.00	110%
6	External	Expert	1077.0	1150.0	0.07	42	49	0.17	110%
7	External	Expert	4000.0	9000.0	1.25	293	335	0.14	110%
8	Internal	Expert	1500.0	1512.0	0.01	70	70	0.00	100%
9	Internal	Expert	1125.0	1000.0	-0.13	106	106	0.00	100%
10	Internal	Expert	1249.0	1242.0	-0.01	153	214	0.40	98%
11	Internal	Expert	1410.0	955.0	-0.48	74	64	-0.16	98%
12	External	Comb.	12000.0	14000.0	0.17	619	640	0.03	95%
13	External	Expert	1249.0	1242.0	-0.01	92	106	0.15	100%
14	External	Expert	487.5	562.5	0.15	103	106	0.03	110%
15	External	Expert	640.0	1085.0	0.70	54	84	0.56	90%
16	Internal	Expert	3937.5	4012.5	0.02	138	152	0.10	95%
17	Internal	Expert	750.0	1200.0	0.60	117	457	2.91	100%
18	External	Expert	533.5	466.5	-0.14	97	104	0.07	100%
19	External	Expert	570.0	907.0	0.59	109	116	0.06	112%
20	External	Expert	292.0	342.0	0.17	113	113	0.00	105%
21	External	Expert	914.0	1903.0	1.08	196	217	0.11	120%
22	External	Expert	400.0	432.0	0.08	42	56	0.33	110%
23	External	Expert	705.0	1000.0	0.42	60	70	0.17	105%
24	External	Expert	2265.0	2732.0	0.21	210	245	0.17	120%
25	External	Expert	1932.0	5631.0	1.91	220	281	0.28	120%
26	External	Expert	2340.0	3454.0	0.48	140	245	0.75	150%
27	Internal	Expert	650.0	696.0	0.07	49	49	0.00	100%
28	Internal	Expert	27241.0	28645.0	0.05	296	336	0.14	90%
29	Internal	Expert	7520.0	8063.0	0.07	395	395	0.00	99%
30	Internal	Expert	6728.0	n/a	n/a	151	n/a	n/a	n/a
31	Internal	Expert	5450.0	8910.0	0.63	212	304	0.43	110%
32	Internal	Expert	90.0	180.0	1.00	21	56	1.67	100%
33	Internal	Expert	2720.0	n/a	n/a	152	n/a	n/a	n/a
34	External	Comb	145.0	195.5	0.35	79	84	0.06	100%.
35	External	Comb	190.0	101.0	-0.88	60	85	0.42	100%
36	External	Expert	593.5	593.5	0.00	152	152	0.00	100%
37	External	Comb	506.0	506.0	0.00	70	70	0.00	100%
38	External	Comb	3784.0	3746.0	-0.01	266	266	0.00	100%
39	External	Expert	1030.0	1335.0	0.30	122	122	0.00	115%
40	External	Expert	2170.0	3831.0	0.77	54	54	0.00	100%
41	External	Comb	3086.0	7844.0	1.54	183	183	0.00	100%
42	External	Comb	1982.0	3140.0	0.58	153	153	0.00	145%
43	External	Expert	133.5	261.0	0.96	273	334	0.22	100%
44	External	Expert	340.0	866.5	1.55	72	91	0.26	125%