Understanding Reasons for Errors in Software Effort Estimates

Magne Jørgensen, Kjetil Moløkken
Simula Research Laboratory
{magne.jorgensen, kjetilmo}@simula.no

Abstract: This study is a first step towards better processes of understanding why errors occur in software effort estimation. Within one software development company, we collected information about estimation errors through: (1) Interviews with estimation responsible employees in different roles, (2) Estimation experience reports from 68 completed projects, and, (3) Statistical analysis of relations between characteristics of the 68 completed projects and estimation error. We found that the role of the respondents, the data collection approach, and the type of analysis had an important impact on the reasons for estimation error that were given. We found, for example, a strong tendency to perceive factors outside the respondents' own control as important reasons for *inaccurate* estimates. Reasons given for *accurate* estimates, on the other hand, typically cited factors that were within the respondents' own control, and were determined by the estimators' skill or experience. This bias in types of reason means that the collection only of project managers' viewpoints will not yield balanced models of reasons for estimation error. Unfortunately, previous studies on reasons for estimation error have tended to collect information from project managers only. We recommend that software companies combine estimation error information from in-depth interviews with stakeholders in all relevant roles, estimation experience reports, and results from statistical analyses of project characteristics.

1 Introduction

In [1] we found, through a review of surveys on software estimation, that the average effort overrun of software projects seems to be in the range 30-40%[1], i.e., the average estimation error of software projects is high. In order to reduce the estimation errors, we need to have means of understanding *why* estimation errors occur. Important questions for that purpose are: Should the software organization base their collection of error information on interviews, project reviews, or, statistical analyses of project characteristics? What is the relation between the types of reasons for error provided and the data collection approach? Previous studies on reasons for estimation error (see Section 4) have been based mainly on questionnaires to project managers and statistical analysis of project characteristics. Does this bias our understanding of why estimation errors occur? This paper aims at answering these questions and is based on a study conducted within one medium-large Norwegian software development organization. Our goal is to identify how different roles, information collection approaches and analysis techniques may supplement each other and lead to better, and more comprehensive, models why estimation errors occur, i.e., our focus is *not* the identification of the most important reasons given for estimation error in the studied organization, but the *processes* by which a better understanding of why estimation errors occur may be gained.

One reason for our focus on the *process* of coming to understand why estimation errors occur, rather than on the estimation errors themselves, is that we believe that companies should attempt to understand why estimation errors occur in their own particular context, and that it may

[1] We are aware of the study by Standish Group which reports an average 89% overrun. That study, however, seems to have methodological weaknesses that lead to over-representation of projects with inaccurate estimates. See our discussion in [1].

be difficult to learn much from general studies on estimation errors in other companies. For example, assume that we ask project managers A, B, and C in three companies about the most important reason for estimation errors. Project managers A and B cite "overlooked tasks" and project manager C "immature users" as the most important reasons. To conclude from this finding that "overlooked tasks" is more important than "immature users" may not be a very useful conclusion. It may be that company A and B have no "immature users". In that case, a better interpretation of the responses is that "overlooked tasks" is perceived to be the most important reason for estimation error in contexts without "immature users"; otherwise, "immature users" is the most important reason. In our opinion, the problems of collecting important organizational context makes the transfer of error reason results between companies difficult.

The remaining part of the paper is organized as follows: Section 2 describes the design of the study, including limitations and challenges. Section 3 provides the results of the study. Section 4 evaluates the validity of our results through an examination of results from other studies. Section 5 concludes and describes further work.

2 Design of Study

2.1 Types of Explanatory Factor for Estimation Error

What does it mean to believe that an event or characteristic, e.g., a major unexpected change in requirements during project execution, is a major causal factor for software development effort overruns in software projects? The answer to this question is not trivial. Potential interpretations of something (X) being a causal factor for effort overruns are, for example:

- There is a *direct* causal link between X and the overrun, i.e., X is a *direct cause* of overrun.
- X leads to events that, in turn, lead to overruns, i.e., X is an *indirect cause* of overruns. If the events leading to overrun started with X, we may call X the *root cause* or the *trigger cause*.
- The events actually leading to overrun would have been harmless if X had not been present, i.e., X is an important *contributory cause*, or *necessary condition* for the overruns.
- The overrun increases when X is present, i.e., X is a *deterministic* cause.
- The presence of X increases the probability of overrun, i.e., X is a *probabilistic* cause
- Mainly the large overruns were caused by X, i.e., X is mainly a *large overruns* cause.
- The main contributor to high *average overrun* is X, i.e., X is an *average-overruns* cause.

We decided not to impose the notion of 'cause' on the respondents. Instead, we asked the responding software professionals why they thought estimation errors occur, and let them use their own judgment as to what would constitute a reason for estimation error, where 'reason' is, for our purposes, just viewed as a linguistic representation of a causal factor. This lack of definition of 'reason' served the main purpose of this paper, which is to investigate how the roles of the respondents, the data collection and analysis approach affect the types of identified factors for estimation error. More discussions on types, definitions and interpretations of cause can be found in [2].

For the purpose of this study we decided to focus on direct reasons, indirect reasons, and, contributory reasons. Our interpretation of these types of reasons is illustrated below:

- *Direct reason*. A reason is categorized as a direct reason if the estimation error is explained by an immediate reason for estimation error. For example, "unexpected problems with the testing tool" is a reason that may immediately lead to estimation error.
- *Indirect reason*. A reason is categorized as an indirect reason if the estimation error is explained by reasons not directly connected to estimation error. For example, "lack of project manager culture" may lead to "insufficient effort on project planning", which in turn may lead to "overlooked tasks" and estimation error. "Lack of project manager culture" and "insufficient effort on project planning" are both indirect reasons of different distance to the direct reason "overlooked tasks". This category also covers more complex mental models of reasons, such as multiple-reasons with joint effect on estimation error.
- *Contributory reasons*. A reason is categorized as a contributory reason if the reason is better described as a necessary condition of estimation error than a direct or indirect reason. For example, assume that "overlooked tasks" is considered to be the direct reason of estimation overrun in a project. A contributory reason of the estimation overrun could then be "too low a buffer for dealing with unexpected events (contingency buffer)", i.e., "overlooked tasks" would not have led to effort overrun, had the contingency buffer been larger.

We focus on these three categories of reason because the categories enable a separation between simple, complex, and context-dependent types of reasoning models. We believe that there may be important relationships to be identified based on these categories. For example, Brickman, Ryan, and Wortman [3] found when studying car accidents, that the drivers dominantly reported direct reasons, while external observers reported more indirect reasons. In other words, who you ask may not only lead to differences in observations and viewpoints, but also in types of reasons given.

2.2 Knowledge of Explanatory Factors for Estimation Errors

A software project includes a high number of factors that may affect, and/or correlate with, the estimation error. Many of these factors are difficult to observe and measure, and the relation between the factors may be very hard to extract without systematic observations or controlled experiments. In this network of interconnected factors leading to estimation error, the task of identifying the importance of one specific factor is obviously a very difficult task. The study described in [4] illustrates the difficulty of identifying reasons for estimation error. That study reports that one important reason for cost overruns was, according to the projects' experience reports, incomplete requirement specifications. Surprisingly, when we compared the requirement specification information from the projects' experience reports with the cost estimation precision data there were indications of the opposite! More often, high estimation accuracy was connected with a *lack* of precise specifications. Incomplete requirement specification may easily lead to more effort than estimated on meetings with the client to clarify needs, but it may also lead to high flexibility in functionality and quality of the delivered solution, e.g., the project may achieve high estimation accuracy due to functionality and quality-reductions that were possible because of the incompleteness of the specification. It may, therefore, be difficult for a project member to identify the total effect of incomplete requirements on estimation accuracy. We should consequently interpret the factors cited by the software professionals as 'perceived factors' and not automatically accept them as 'actual factors'.

2.3 Data Collection

Company: The company that provided the data is a medium-large (about 100 employees) Norwegian software development company that produces web-portals, e-commerce solutions and content management systems for their clients. The main work process is based on the waterfall development model and contains six phases: strategy and concept, specification, development, test, implementation and evaluation. There was some use of evolutionary-incremental development models. Most projects were "multi-disciplinary", i.e., they involved professionals in the role of "graphic designer", "user interaction designer", "project manager" and "programmer". The company had not implemented a formal estimation process and the actual estimation processes, consequently, varied within the organization. The dominant estimation approach was bottom-up, i.e., work breakdown structure-based. The project manager was usually responsible for the project's estimate. Most projects were small and there were many different clients, whose experience level regarding software projects varied from those who requested their first web-system to companies that had based their daily operations on software systems for many years. Most of the estimates were completed as part of the project planning process.

Limitations related to selection of company: The studied company had mainly small projects, informal estimation and development processes, and, dominantly, immature clients. This means that other types of company may give different reasons for estimation error. The *types* of reason provided and the impact from data collection and analysis approach to the types of reason may, however, be more robust towards size of projects, formality of processes and maturity of clients, i.e., we believe that our results may be valid for other types of software companies, as well.

Data collection approaches: To examine the impact of the role of respondent, data collection approach, and analysis technique we decided to collect reasons for estimation error based on three approaches in the same organization, i.e., (1) through general *interviews* with eight employees responsible for the estimates, (2) through 68 *project estimation experience reports*, and, (3) through statistical analysis of associations between project characteristics and estimation error of the same 68 projects as in (2). None of the interviewed employees provided project experience reports. The 68 project experience reports were completed by 29 different employees in the roles of project manager or developer. The data collection lasted about one year.

Limitations of the data collection: The interviews and the projects do not describe exactly the same estimation situations, because they were collected at different times and because we excluded projects with planned duration of more than four months from the logging. Both limitations were the result of practical concerns. The permission to log information about projects came as a result of the analysis of the interviews, and, we initially intended that the logging would not last more than a few months, because projects longer than four months would be difficult to complete within our study. There were, however, no large changes in estimation or development process in the period between the interviews and the project logging, and the company had very few large projects. Nevertheless, the limitation may have had an impact on the difference in the reasons for estimation error provided in the interviews, and in the project data. We do, however, not believe that this limitation has an important impact on *how* estimation error reasons are described, i.e., *what types* of reasons people give to explain estimation errors. As stated earlier, our goal is not to examine the reasons for estimation errors in the studied company, but to analyze the impact of the role, the data collection approach, and, the analysis approach on the types of error reasons given.

Interviews: One of the authors of this paper interviewed the following eight management personnel responsible for estimates:

- The manager of the technical personnel (M-Tech)
- The manager of the human-computer-interaction personnel (M-HCI)
- The manager of the graphic design personnel (M-Graph)
- The most senior project manager (PM-Sen). This project manager was frequently used to review other project managers' estimates.
- Two project managers with technical background (PM-Tech1 and PM-Tech2)
- A project manager with human computer interaction background (PM-HCI)
- A project manager with graphic design background (PM-Graph)

Following an introduction about the purpose of the interview and general questions about the estimation process we asked the above-mentioned personnel to give reasons for both accurate and inaccurate effort estimates. No pre-defined categories of reason or templates for the answers were used. We instructed the interviewed employees to base their reasons on experience from a large set of projects, and not, for example, one or two projects with especially large overruns. Each interview lasted 1-2 hours. The interviews were meant to be the organization's first step towards the improvement of the estimation process. However, due to re-organizations and other unexpected events, the planned estimation improvement work was never continued.

Experience Reports and Statistical Analysis of Project Characteristics: Over a period of approximately a year we collected information about projects with an estimated effort of less than approximately four calendar months, i.e., we excluded the largest and the smallest projects. In total, information about 68 projects was collected. All these 68 projects provided "estimation experience reports", where reasons for accurate or inaccurate estimates were provided, together with other information about the project. The Chief Project Manager of the company was in charge of data collection. He asked the estimators to complete one questionnaire just after completing the project planning and another one after the project was completed. The completion of the questionnaires was supported by a spreadsheet-based tool that guided the project manager through a number of questions.

The information collected before a project started was as follows (with pre-defined categories):

- Company role of the estimator (Project manager, developer)
- Brief description of the project (Free text).
- The estimators' assessment of the complexity of the project (Easy, medium, complex).
- Type of contract (Payment per hour, firm price).
- The estimators' assessment of how important the project is for the client (Low/medium importance, high importance, critical).
- The priority that the client assigns to the project (Cost, quality, or time-of-delivery).
- The estimators' self-assessed level of knowledge about how to perform the project (Know something about how to solve the task, know much about how to solve the task).
- The estimators' planned participation in the completion of the project (0%, 1-50%, 51-100% of the work planned to be completed by the estimator him/herself)
- The estimators' perception of his typical accuracy when estimating similar projects (Accuracy categories from "less than 10%" to "more than 100%".)

- The estimated effort in work hours. (We found that the estimators had slightly different interpretations of 'estimated effort'. In most projects, however, 'estimated effort' seemed to be interpreted as 'planned effort', i.e., 'most likely effort' added a contingency buffer. All remaining project activities were included in the estimate, e.g., project administration, design, programming, and, test.)

After the project was completed the estimators provided an estimation experience report in terms of:

- The actual effort in work hours.
- Comments on the actual use of effort.
- Descriptions of unexpected problems during the execution of the project.
- Reasons for high or low estimation accuracy

All project characteristics were included in the statistical analyses of estimation error. The estimation experience report was based on all information collected immediately after the completion of the project, in particular the responses when we asked for "reasons for high or low estimation accuracy".

2.4 Measures

We apply two common measures of estimation error in this study. One measure is of the mean magnitude of relative error (mean MRE) and the other is of the mean relative error (mean RE). The mean relative error shows the mean over-run or the bias of the estimates, e.g., a high RE means that there is a strong tendency to under-estimation.

Mean MRE (Magnitude of Relative Error) is measured as:

$\frac{1}{n}\sum_i \frac{|Act_i - Est_i|}{Act_i}$, where Act_i is the actual effort on project *i*, Est_i is the estimated effort for project *i,* and *n* is the total number of projects.

Mean RE (Relative Error) is measures as:

$$\frac{1}{n}\sum_i \frac{(Act_i - Est_i)}{Act_i}$$

3 Results

3.1 Interviews

Table 1 describes the most important reasons (or, in some cases, reasoning models) for estimation error as perceived by each interviewed subject according to our categorization of, and notation for, reasons (direct reason →, indirect reason →→, or contributory reason ↓→). We have translated and condensed the most important reasons provided by each subject without, as far as we are aware, changing the intended opinion of the respondents.

Table 1: Interview-based Reasons for Estimation Error

Subject	Reasons
M-Tech *(Manager of the software developers)*	No systematic feedback to enable learning (→→). Insufficient time on estimation and planning (→→), leads to overlooked tasks (→).
M-HCI *(Manager of the HCI personnel)*	Lack of processes enabling learning from experience (→→). Insufficient focus on HCI in the estimation process (→→). Lack of client realism in HCI-requirements (→→).
M-Graph *(Manager of the graphical designer personell)*	Project managers are not skilled in planning multi-disciplinary projects (→→), which leads to insufficient focus on graphic design in the estimation process (→→), and inefficient allocation and use of graphic design resources (→). No systematic feedback to enable learning (→→).
PM-Sen *(Senior project manager with extensive experience from project bidding and planning)*	Insufficient focus on the project manager role (→→), leads to insufficient training and feedback (→→). Insufficient standardization of planning and development processes (→→). The experience database of previous projects is not used (→→).
PM-Tech1 *(Project manager with technical background)*	Clients unable to deliver a good requirement specification (→→), leads to unplanned re-work (→). Lack of requirement change control processes (→→). Insufficient time spent on estimation and planning (→→).
PM-Tech2 *(Project manager with technical background)*	Projects are frequently different from earlier projects (→→), leads to lack of relevant experience when estimating (→), because of lack of checklists (↓→) and experience database (↓→).
PM-HCI *(Project manager with HCI background)*	HCI is involved too late (→→), which leads to unrealistic expectations by clients (→→), and unplanned activities (→). Project manager has insufficient knowledge about HCI (→→).
PM-Graph[2] *(Project manager with graphic designer background)*	Insufficient focus on graphic design in the estimation process (→→). No systematic feedback to enable learning (→→).

There are several interesting observations that can be derived from the interviews summarized in Table 1:

- Although there are common patterns in the responses, e.g., the need for better learning opportunities, the role of the respondent seems to have a strong bearing on the type of reasons provided. For example, there seems to be a pattern that general managers (M-Tech, M-HCI, M-Graph) more frequently provide more general reasons for estimation error than the project managers (P-Sen, P-Tech, P-HCI, P-Graph). In addition, there

[2] *Important comment from PM-Graph*: "We seldom have large overruns on graphic design activities. A graphical design can be completed in 5 hours or 5 months, dependent on how much money the client is willing to spend to ensure a good design, for example, on iterations and user tests".

seems to be a tendency *not* to criticize work connected to one's own role. In other words, the factors cited for estimation error have a tendency to be outside the control of the respondent. This pattern of not criticizing factors controlled by oneself is further supported by findings reported in Section 3.2 and Section 4, and by studies in other domains. For example, Tan and Lipe [5] found, in a business management context, that low estimation accuracy was explained as due to uncontrollable external factors.

- Only one of the respondents provided reasons that were described as contributory reasons (↓→), i.e., important enablers of estimation error outside the main chain of reasons leading to estimation error. We obviously need more observations to evaluate whether this is typical or not, but there may be a need for explicit focus on contributory reasons when these are important.
- Frequently, the steps in the chain from an indirect reason (→→) to the estimation error were not well explained. For example, PM-Sen claimed that "insufficient standardization of planning and development processes" is a reason for estimation error. More standardization is, however, no "silver bullet" to improve estimation accuracy. Its impact on estimation accuracy depends, amongst other things, on properties of the standards applied, and, the organization's ability to establish processes that enable learning from experience with standardized processes. To really understand the provided reasons for estimation error, we may have to push the respondents for more comprehensive structures of reasons, where all important steps and all non-obvious contributory reasons are included.
- All of the reasons were described deterministically. This suggests, but does not prove, that the models cited by the respondents to explain estimation overruns are deterministic and not probabilistic. Hammond [6] suggests that the ability to understand relationships in terms of probabilities instead of purely deterministic connections is important for correct learning in situations with high uncertainty, such as effort estimation of software projects. For example, instead of the deterministically described reason for estimation errors: "Clients unable to deliver a good requirement specification" (M-Tech1), a probabilistic description of the same reality may be: "When clients are not able to deliver a good requirement specification, there is sometimes a high uncertainty in the use of development effort and, consequently, a high probability of inaccurate estimates." This ability to think about reasons in probabilistic terms can, according to Brehmer [7], hardly be derived from experience alone, but must be taught. We have included a more comprehensive discussion about the importance of probability-based reasoning models when learning from software estimation experience in [8].
- It was frequently unclear whether the respondents described reasons for the largest overruns or the typical overruns, i.e., the scope of the reasons were not described.

Interviews may enable the description of complex models of reasons. However, our interviews suggest that models of more than 2-3 steps, with contributory reasons, with probabilistic relationships, and of well-defined scope, are not provided by software professionals when simply asked for "reasons for estimation error", i.e., there may be a need for more structure in the process of elicitation of comprehensive models of estimation error.

3.2 Project Experience Reports

Based on repeated readings of the 68 experience reports, we developed a classification of estimation errors. This classification process was based on joining reasons that we perceived to be closest to each others until a "reasonable number" (15 in our case) of categories had been identified. Table 2 (reasons for inaccurate estimates) and Table 3 (reasons for accurate estimates) summarize the reasons based on our classification. The estimators themselves decided whether they considered the estimate to be accurate or inaccurate, i.e., whether they should report reasons for accurate or inaccurate estimates. This decision depended, amongst others, on the previous estimation accuracy experience and the complexity of the estimation work. For each set of projects belonging to a particular category of reason we calculated the mean MRE, the mean RE, and the proportion of "over median large projects". The median effort of a project in the company was 45 work-hours, and all projects with estimated effort more than 45 work-hours were hence classified as "over median large". One project may have mentioned more than one reason for estimation inaccuracy or accuracy. In fact, there were a few project experience reports that described reasons for both accuracy and inaccuracy. In every such case there were good reasons for including both types of reason. For example, Project 7 reported reasons for both inaccuracy, i.e., "unexpected change requests", and accuracy, i.e., "large contingency buffer". The total estimation over-run of Project 7 was 11%. The project manager's explanation for including both reasons was that the unexpected change requests did lead to more work than planned, i.e., to inaccuracy, but the large contingency buffer saved the project's estimate and led to an overall acceptable level of accuracy.

The mean MRE of all tasks was 28% and the mean RE was 8%. The relatively low bias towards underestimation (indicated by the low, positive RE value) is probably caused by the small size of most of the projects, see for example [9] for similar results.

Table 2: Experience Report-based Reasons for Inaccurate Estimates

Id.	Reason	Reported in Project	Mean MRE	Mean RE	Over Median Large Projects
1	Unexpected events and overlooked tasks (→)	5, 8, 11, 12, 21, 25, 26, 30, 35, 43, 47, 49, 50, 51, 52, 58, 60, 62, 63, 64, 65, and, 66	0.32	0.31	45%
2	Change requests from clients or "functionality creep" (→)	5, 7, 9, 14, 15, 16, 17, 18, 22, 23, 31, 48, 55, 61, and, 67	0.30	0.30	47%
3	Simpler task or more skilled developer than expected (→)	34, 36, 55, 57, and, 68	0.69	-0.69	20%
4	Resource allocation problem (→→)	8, 28, and, 47	0.30	0.30	33%
5	Poor requirement	44, 45, and, 54	0.31	0.27	67%

	specification or problems with communication with the client (→→)				
6	Too little effort on estimation work (→→)	63	0.70	0.70	100%
7	High priority on quality, cost accuracy not of high importance (→→)	30	0.38	0.38	0%
8	More reuse than expected from other projects (→)	4	0.63	-0.63	100%

Table 3: Experience Report-based Reasons for Accurate Estimates

Id.	Reason	Reported in Project	Mean MRE	Mean RE	Over Median Large Projects
1	Inclusion of a large buffer to deal with unexpected events and/or changes in specification (→)	7, 9, 13, 16, 25, 29, 32, 45, 55, 59	0.18	-0.12	50%
2	Simple project (→)	1, 2, 6, 13, 19, 20, 24, 27. 53, 56	0.13	-0.05	60%
3	Experience from a similar project (→→)	1, 4, 10, 12, 19, 28, 41, 51, 54	0.16	-0.01	44%
4	Good knowledge of how to solve the requirement specification (→→)	3, 9, 14, 21, 33, 37, 46	0.14	-0.02	71%
5	A high degree of flexibility in how to implement the requirement specification (→)	3, 16, and, 42	0.17	-0.17	67%
6	Much time was spent on estimation work (→→)	9, and, 29	0.16	-0.16	100%
7	Good cost control (→→)	42	0.26	-0.26	100%

Interesting observations that can be derived from Tables 2 and 3 include:

- Most reasons were direct reasons. There were relatively few projects that described indirect reasons and none that described contributory reasons. The actual reasons provided in the estimation experience reports were only to some extent overlapping the reasons provided in the interviews. In general, it seems as if interview-based reasons focus more on process and learning issues, while estimation experience reports-based reasons focus more on specific events and specific project or estimator characteristics. Similar to in the interviews, the respondents had a tendency to report reasons *outside* their own control as reasons for estimation error. For example, "unexpected events and overlooked tasks" typically referred to events and tasks outside the control of the project. Interestingly, the respondents reported reasons *within* their own control or their own skill and experience, e.g., "inclusion of a large buffer" as factors contributing to accurate estimates.

- Frequently, the estimators regarded over-estimated projects as having accurate estimates, except when the over-estimation was very high. For example, the estimator of Project 42 perceived that he had delivered an accurate estimate, although he over-estimated the project by 26%. His explanation was that the good estimation performance was a result of: "... *very strict project management, tight control of time reporting and competent developers.*" To some extent, the project manager may be correct in his interpretation of accuracy. He could easily relax the control of the project, spend more effort, and consequently improve the estimation accuracy. From his point of view, it would have been unfair to perceive his estimate as inaccurate. Regardless of the correctness of his viewpoint, it means that when we ask project managers to provide reasons for estimation error we mainly get reasons for under-estimation and very high over-estimations, not medium-large over-estimations. It may be important to be aware of this when it is important to understand reasons for medium-high over-estimation, e.g., for companies in bidding situations losing contracts if the bids are unnecessarily high.
- There were no clear connections between the frequency of a reason and the mean accuracy of the reasons, i.e., we should not conclude from the fact that a factor is frequently mentioned that it also leads to the largest overruns.
- There were no clear patterns relating reasons for estimation error to the size (larger or smaller than the median project) of the project. A possible conclusion is that the size of project does not affect the reasons for estimation error very much within the limited variation of project sizes studied in this company.
- There were reasons we would expect to be reported but that nobody provided. One such reason is the "political estimation games" described, for example, in [10, 11]. For example, a clients expects to pay a certain price for the software and the estimator is under strong pressure to reduce the initial "too high" estimate to ensure that the project can be started. We found no descriptions of such "political games" as reasons for estimation error. However, from informal meetings and lunch-discussions with some of the software developers we know that unrealistic client expectations of low cost may well have been an important factor affecting estimation overrun in some of the projects. This means that some reasons may not be mentioned because they are sensitive, or perhaps because the estimators feel uncomfortable about, for example, admitting that they sometimes succumb to pressure from clients.

Review of project-specific reasons for estimation errors, such as in the estimation experience reports studied in this study, seems to stimulate the description of direct reasons. It may, therefore, be necessary to actively stimulate the provision of indirect and contributory reasons to get a broader picture and stimulate so-called double-loop learning [12], i.e., learning that includes better understanding of the core factors that affect estimation error. In addition, to understand how "political games" affect estimation overruns, we may need structures that provide incentives for the giving of sensitive reasons.

3.3 Statistical Analysis

Earlier [13] we applied a subset of the dataset applied in this paper, i.e., the 49 earliest out of the current set of 68 projects, to develop a regression model for the prediction of estimation error. The regression models of MRE (absolute estimation error) and RE (relative estimation error) were developed by applying stepwise regression with backwards elimination and an alpha-value of 0.1 to remove variables. The variables, i.e., all the project characteristics described in Section 2.3, were coded as binary variables. A full description of the coding and its rationales are provided in [13]. The resulting regression models were the following:

```
MRE = 0,14 + 0,13 Company Role + 0,13 Participation + 0,13 Client Priority,
    (p=0.03)      (p=0.08)             (p=0.07)             (p=0.09)

RE = 0,12 - 0,29 Company Role + 0,27 Previous Accuracy
   (p=0.05)      (p=0.004)           (p=0.01)
```

The variables included in the proposed models were defined as follows:

- Company Role: The project was estimated by a software developer = 1, The project was estimated by a project manager = 0.
- Participation: The estimator estimated the work of others = 1; The estimator participated in the estimated project = 0.
- Client Priority: The client prioritized time-to-delivery = 1; The client had other project priorities than time-to-delivery, i.e., cost or quality = 0.
- Previous Accuracy: The estimator believed that he/she had estimated similar tasks with an average error of 20% or more = 1; less than 20% error = 0.

The adjusted R^2-values were low, i.e., 11% for the MRE-model and 21% for the RE-model. This indicates that the models did only explain small proportions of the variances of mean estimation errors.

A re-analysis of the project data, including the new projects, i.e., with 68 projects instead of 49, led to the following regression models:

```
MRE = 0,14 + 0,13 Company Role + 0,14 Participation + 0,14 Client Priority,
    (p=0.01)      (p=0.07)             (p=0.04)             (p=0.05)

RE = 0,10 - 0,22 Company Role + 0,23 Previous Accuracy
   (p=0.08)      (p=0.02)           (p=0.03)
```

The adjusted R^2- value was the same as before (11%) for the MRE-model, while it decreased for the RE-model (from 21% to 11%). As can be seen, the same variables were significant in the updated, as well as in the original models, with almost the same regression coefficients and similar p-values. This suggests that the regression-based relationships, in particular the MRE-model, are robust towards extensions of the dataset.

Table 4 shows the relations between estimation error and other project variables through a description of mean MRE and RE for all variable categories.

Table 4: Project Characteristics in Relation to Estimation Error (MRE and RE)

Variable	**Categories (# observations)**	**Mean MRE**	**Mean RE**
Role of Estimator	Developer (22)	0.33	-0.05
	Project manager (46)	0.25	0.14
Complexity	Easy (27)	0.25	0.11
	Medium (29)	0.31	0.05
	Difficult (12)	0.26	0.09
Type of payment	Fixed price (38)	0.30	0.10
	Per hour (30)	0.25	0.06
Importance	Low/medium importance (20)	0.29	0.03
	Very important (40)	0.24	0.09
	Critical (8)	0.47	0.18
Priority	Cost (18)	0.22	0.01
	Quality (29)	0.24	0.11
	Time-to-delivery (21)	0.38	0.09
Knowledge on how to solve the project	Much knowledge (45)	0.29	0.04
	Some knowledge (23)	0.26	0.16
Estimation error of similar projects	0-10% error (21)	0.29	0.05
	11-20% error (26)	0.29	0.05
	21-30% error (7)	0.19	0.04
	31-50% error (6)	0.37	0.37
	51-75% error (1)	0.20	0.20
	> 76% error (2)	0.47	0.47
	information not provided (5)	0.17	-0.10
Participation of estimator in project	No participation (28)	0.31	0.17
	1-50% of total work (23)	0.21	-0.04
	51-100% (17)	0.32	0.06
Estimated size of project	Estimated effort < medium effort (33)	0.25	0.10
	Estimated effort >= medium effort (35)	0.31	0.06

As stated earlier, the main goal of our study is *not* to analyze the reasons for estimation error, but to compare the difference in types and models of "reasons for estimation error" found when applying different approaches to data collection and analysis. Examining the regression model and Table 4 we find that the statistical analysis supports and extends the understanding of the reasons found by interviews and estimation experience reports. For example:

- The statistical analyses, i.e., the regression models and the differences in mean MRE and RE, point to the importance of project managers being the estimators, and, the interviews point to the importance of skilled project managers. This may be viewed as two descriptions of similar factors, i.e., that skills typically possessed by project managers, and not so much by software developers, are important for accurate estimates.

- The statistical analyses point to higher uncertainty when there is no participation in the project by the estimator, and the interviews suggest that HCI and graphic design work are not properly understood by many project estimators. Both descriptions support the hypothesis that stronger involvement in the project work leads to more accurate estimates.
- The statistical analyses point to the higher uncertainty of projects with a focus on time-to-delivery, and the interviews and the experience reports focus on the lack of good requirement specifications or frequency of unplanned changes. Combining these two information sources, we may state the preliminary hypothesis that the explanation for lower estimation accuracy in situations with priority on time-to-delivery is that too short a time is taken for the development of proper requirement specifications. This, in turn, may lead to unplanned changes and effort overruns. Here we see that a diversity of information sources may support the building of more comprehensive theories than single information sources.

4 Evaluation of the Validity of the Results

Most published studies on reasons for software development estimation errors are based on questionnaires. The design and results of these questionnaire-based studies are briefly described in Table 5. In addition, a few studies on estimation errors have been based on statistical analyses. Table 6 summarizes these studies. Notice that the main purpose of this summary of previous studies is *not* to compare the reasons found in previous studies with the reasons found in our study, but to evaluate the validity of our findings regarding the impact of roles, data collection approach and analysis technique on the types of reason provided.

Table 5: Questionnaire-Based Studies on Reasons for Software Estimation Error

Study	Population	Study Design	Results
Phan et al. [14]	Software professionals (80% of them were project managers or developers) in 191 organizations.	Four pre-defined categories: Long duration, over-optimism, poor analysis and design, and frequent changes.	The two most important reasons were “unrealistic over-optimism[3]” and “frequent changes”. The least important reason was “poor analysis and design”.
Van Genuchten [15]	Project managers responsible for the estimation of 160 activities in six development projects within one department.	Pre-defined classification of reasons for error. The six project managers marked one (or more) of these for each activity.	Most frequent reasons were “more time spent on other work than planned” and “complexity of application underestimated”.

[3] It is unclear how to interpret “unrealistic over-optimism” as a *reason* for estimation overruns. To some extent, “unrealistic over-optimism“ is the same thing as “effort overrun”.

Lederer and Prasad [16]	Estimation responsible (mainly project managers and developers) personnel in 112 organizations.	Pre-defined list of reasons where general importance for estimation error was marked with a value from 1 to 5.	Most important reasons were "frequent requests for changes by users", "users lack of understanding of their own requirements", and "overlooked tasks".
Standish Group - 1994[4]	"IT executive managers" (mainly project managers?) from 365 organizations.	Pre-defined classification of reasons.	The three most important reasons for estimation over-runs were "lack of user input", "incomplete requirements and specifications", and, "changing requirements and specifications".
Subramanian and Breslawski [17]	Project managers in different companies representing 45 projects.	Reasons classified by the authors based on responses from the project managers.	Most important reasons were "requirement change/addition/deletion", "programmer or team member experience, turnover", and, "design changes, scope, complexity".

There are clear similarities in the results in the above studies. Most studies seem to focus on direct reasons and reasons related to the clients and users, e.g., "frequent changes", "frequent requests for changes by users", "changing requirements and specifications", and, "requirement change/addition/deletion". The reported reasons are similar to those we found in our study when requesting project estimation experience reports from the project managers, but do not report many of the indirect and contributory factors that were cited in the interviews by employees in different roles.

This focus on user-related issues, and lack of focus on indirect reasons, supports our findings concerning the impact of the respondents' role and the data collection approach. The studies in Table 5 have a predominance of project managers as respondents, and simple questionnaires as the approach to data collection. Similarly to the results of our study, project managers focus on factors outside their own control when providing reasons for estimation errors and simple questionnaires do not lead to the provision of indirect reasons and comprehensive reasoning models. This is not mainly due to the pre-defined reasons in many of the studies, since most of the pre-defined reasons were indirect reasons. In other words, the results reported in Table 5 do only represent, in the main, the project manager's perspective and direct reasons for estimation error.

[4] www.standishgroup.com/sample_research/chaos_1994_1.php. There are commercially available updates of the 1994 report available.

Table 6: Statistical Analyses of Factors Associated with Estimation Error

Study	Population	Design of Study	Results
Lederer and Prasad [18]	Same dataset as in [16], see Table 5.	Analysis of statistical significance of differences.	Significant ($p<0.05$) findings were that the estimation error increased: (1) with use of estimation tools, (2) when not estimating own work, (3) when there was no revision of estimates by the management, (4) when there was no independent evaluation of development process, (5) when there was no formal process of cost control, (6) when there was no evaluation of estimation accuracy to assess the managers, the estimators or developers.
Standish Group, Chaos Report, 1994	"IT executive managers" (mainly project managers?) from 365 organizations.	Analysis of difference in mean effort overrun of different types of project.	Estimation error increases with increased size.
Gray et al. [19]	Information about the development of 77 modules of a large health-care system.	Several types of statistical analysis associations between estimation error category and module properties.	The analyses showed, amongst other things, that over-estimation was connected with changes on small modules, development of screens and modules accessing one or less data tables, while under-estimation was connected with changes on large modules, development of reports, and models accessing more than one data table.

The statistical analysis-based results described in Table 6 are different from the questionnaire-based reasons found in Table 5. The difference in reasons is probably not only a result of difference in the variables collected and analyzed, but may also have been caused by differences in the method of analysis. For example, while the statistical analyses suggest that the size of the project is an important indicator of estimation error, the interviews, estimation experience reports and questionnaires did not mention size of project as a factor affecting estimation error. This may mean that there are relationships that are easier to examine through statistical analysis than through "on the job" experience. In other words, the studies in Tables 6 and 7 support our result that experience-based reasons and associations found through statistical analyses supplement, rather than replace, each other.

Table 7 summarizes findings from two interesting studies on reasons for estimation error in manufacturing and construction projects that provides further support for our findings that the role of the respondents may have a strong impact on the type of reasons provided for estimation errors.

Table 7: Two Manufacturing and Construction Projects

Study	Population	Design of Study	Results
Thambain and Wilemon [20]	304 participants in project management workshops and seminars.	Pre-defined categories on reasons for cost and time over-run. Use of questionnaires.	The general managers perceived that "insufficient front-end planning", and, "unrealistic project plans" were the most important reasons, while the project managers believed that "client/management changes" and "technical complexity" were the most important reasons.
Chan and Kumaraswamy [21]	147 organizations involved in construction projects in the role of clients, consultants, or contractors.	Pre-defined categories of reasons for delays in building and civil engineering projects. Use of questionnaires.	In building projects the clients believed that "poor site management and supervision", and, "inadequate managerial skill" were the two most important reasons for delays. The consultants also believed that "poor site management and supervision" was the most important reason, but included "unforeseen ground conditions" as the second most important reason. The contractors, i.e., the organizations responsible for the delay, believed that the delays were mainly caused by "delays in design information", and, "long waiting time for approval of drawings".

These two studies demonstrate, perhaps even more clearly than the software studies, the importance of the respondents' role when providing reasons for errors and failures. There is no reason to believe that we would receive different results when including, for example, the user, the client or independent observers in studies of software development estimation error.

5 Conclusion

It matters whom you ask and how you collect reasons for estimation error, and there are clear patterns with respect to types of reason for estimation errors, dependent on respondents' role, data collection approach, and approach to data analyses. Interestingly, we did not identify contradictory reasons for estimation errors when applying different information sources, data collection and analysis approaches on the estimation error data collected within one company. Instead, we found that the different information sources, data collection approaches, and, techniques supported and supplemented each other.

Potentially useful observations from our comparison of interviews, experience reports and statistical analyses include the following:

- Identification of indirect reasons (enabling double-loop learning) was much more frequent in general interviews than in project specific estimation experience reports, i.e., to get comprehensive reasoning models a company may need interviews with senior personnel with a general focus on reasons for estimation error and not only project specific estimation experience reports and questionnaires.
- The identified reasons for estimation *inaccuracy* were connected to factors not controlled by the respondent, while reasons for estimation *accuracy* were connected to factors within the control of the respondents or related to the respondents' skill or experience. For example, reasons for estimation error provided by the project manager/estimator led to an emphasis on client-related issues, while the interview with the managers of the project managers focused on the need to improve the project managers' skills and processes.
- The use of statistical analyses improved the interpretation and validity of subjective project experience reports and experience-based reasons stated in interviews. This was somewhat unexpected, in light of the low explanatory power and the associative nature of the regression models, (regression models are based on co-variation and not necessarily on cause-effect relationships.)

Based on our findings and further observations, we intend to establish guidelines and frameworks to ease the extraction of better and more comprehensive models of factors affecting estimation error. This, in turn, should lead to improved estimation processes and better estimation accuracy. One promising framework, developed for the analysis of errors made in the practice of medical science, is described in [22]. That approach is based on pre-defined levels of reasons, where level 1 reasons describe factors that directly influence the behavior of the individual practitioners, level 2 reasons affect the team-based performance, level 3 reasons relate to the management or organizational level, and, level 4 reasons are on a governmental or national level. Other candidate frameworks are Post Mortem Analyses, Ishikawa (fishbone)-diagrams and Root Cause Analysis.

Acknowledgment: Thanks to Stein Grimstad at Simula Research Laboratory and Dr. Chris Wright at Cambridge Language Consultants who provided valuable comments and suggestions to this paper.

References

1. Moløkken, K. and M. Jørgensen. *A Review of Surveys on Software Effort Estimation*. in *IEEE International Symposium on Empirical Software Engineering (ISESE 2003)*. 2003. Rome, Italy: p. 223-230.
2. Pearl, J., *Causality: Models, reasoning, and inference*. 2000, Cambridge: Cambridge University Press.
3. Brickman, P., K. Ryan, and C.B. Wortman, *Causal chains: attribution of responsibility as a function of immediate and prior causes.* Journal of Personality and Social Psychology, 1975. **32**: p. 1060-1067.

4. Jørgensen, M., L. Moen, and N. Løvstad. *Combining Quantitative Software Development Cost Estimation Precision Data with Qualitative Data from Project Experience Reports at Ericsson Design Center in Norway*. in *Proceedings of the conference on empirical assessment in software engineering*. 2002. Keele, England: Keele University.
5. Tan, H.-T. and M.G. Lipe, *Outcome effects: the impact of decision process and outcome controllability.* Journal of Behavioral Decision Making, 1997. **10**(4): p. 315-325.
6. Hammond, K.R., *Human judgement and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* 1996, New York: Oxford University Press.
7. Brehmer, B., *In one word: Not from experience.* Acta Psychologica, 1980. **45**: p. 223-241.
8. Jørgensen, M., *A Review of Studies on Expert Estimation of Software Development Effort.* To appear in: Journal of Systems and Software, 2004.
9. Kitchenham, B., et al., *An empirical study of maintenance and development estimation accuracy.* Journal of Systems and Software, 2002. **64**(1): p. 57-77.
10. Flyvebjerg, B., B.S. Holm, and S. Buhl, *Underestimating costs in public works projects.* Journal of theAmerican Planning Association, 2002. **68**(3): p. 279-295.
11. Lederer, A.L. and J. Prasad, *The validation of a political model of information systems development cost estimating.* A-ACM Transactions on Computer Personnel, 1991. **13**(2): p. 47-57.
12. Argyris, C., *On Organizational Learning*. 1993: Cambridge, MA: Blackwell.
13. Jørgensen, M., *Regression Models of Software Development Effort Estimation Accuracy and Bias.* To appear in: Empirical Software Engineering, 2004.
14. Phan, D., D. Vogel, and J. Nunamaker, *The search for perfect project management*, in *Computerworld*. 1988. p. 95-100.
15. van Genuchten, M., *Why is software late? An empirical study of reasons for delay in software development.* IEEE Transactions on Software Engineering, 1991. **17**(6): p. 582-590.
16. Lederer, A.L. and J. Prasad, *Causes of inaccurate software development cost estimates.* Journal of Systems and Software, 1995. **31**(2): p. 125-134.
17. Subramanian, G.H. and S. Breslawski, *An empirical analysis of software effort estimate alterations.* Journal of Systems and Software, 1995. **31**(2): p. 135-141.
18. Lederer, A.L. and J. Prasad, *Information systems software cost estimating: a current assessment.* Journal of Information Technology, 1993. **8**(1): p. 22-33.
19. Gray, A., S. MacDonnell, and M. Shepperd. *Factors systematically associated with errors in subjective estimates of software development effort: the stability of expert judgment*. in *Sixth International Software Metrics Symposium*. 1999: IEEE Comput. Soc, Los Alamitos, CA, USA: p. 216-227.
20. Thambain, H.J. and D.L. Wilemon, *Criteria for controlling projects according to plan.* Project Management Journal, 1986(June): p. 75-81.
21. Chan, D.W.M. and M.M. Kumaraswamy, *A comparative study of causes of time overruns in Hong Kong construction projects.* International Journal of Project Management, 1997. **15**(1): p. 55-63.
22. Johnson, C., *The causes of human error in medicine.* Cognition, Technology & Work, 2002. **4**: p. 65-70.