ESTIMATION ACCURACY IN LARGE IS PROGRAMS - INSIGHTS FROM A DESCRIPTIVE CASE STUDY

Complete Research
Kaiser, Marcus, Senacor Technologies AG, Munich, Germany, marcus.kaiser@senacor.com
Ullrich, Christian, FIM Research Center, Augsburg, Germany, christian.ullrich@fim-rc.de

Abstract

Information systems (IS) projects are famous for experiencing severe cost overruns, which amongst others are often caused by inaccurate ex-ante cost estimations. Against this background, this article presents a descriptive case study located in an IS transformation program at a major German financial services provider. In this case study, a multi-stage cost estimation process, which was applied to 79 IS projects, is described and the estimation accuracy of the cost estimations of all IS projects is determined using different estimation accuracy measures: Estimating Quality Factor, Forecast Error, and Mean Absolute Percentage Error. Depending on the concrete estimation accuracy measure used for the evaluation, the overall estimation quality of the program turns out to be evaluated as good or at least average – which seems to be contrary to most studies in scientific literature. However, the results further reveal that the estimation accuracy also depends on the estimation accuracy measure chosen for the evaluation. These differing judgements are discussed from a management perspective.

Keywords: IT Project Management, IS Projects, Cost Overruns, Estimation Accuracy.
1 Introduction

Failing IS projects have drawn attention throughout their history. The success or failure of IS projects are usually measured along three criteria: scope, schedule, and budget (Jurison 1999; Sauer et al. 2007). The relevance of these criteria is demonstrated in a lot of famous examples of failed IS projects reported in literature (Goldfinch 2007; Krigsman 2008; McFarlan 1981; Newcombe 1998; Oz 1994). Many authors examined the reasons for project failure in general (Zimmerer and Yasin 1998) and IS projects in particular (El Emam and Koru 2008; Nelson 2007; Whittaker 1999). Cerpa and Verner (2009) e.g. analysed 70 failed IS projects and concluded that more than 80% were underestimated.

Kahneman and Tversky (1979) discovered that decision makers tend to underestimate costs, time to completion, and risks, whereas the same people usually overestimate benefits. Especially in the context of IS projects, several studies revealed that project managers tend to overestimate their own capabilities (Atkinson et al. 2006; Budzier and Flyvbjerg 2011). Though there are numerous methods for estimating parameters of IS projects, Eveleens and Verhoef (2009) conclude that these kinds of bias lead to wrong project evaluations. Due to biases inherent in the estimates, the authors state that quantifying the estimation accuracy “provides crucial information for executives to help steer their organization“ (Eveleens and Verhoef 2009, p. 935). This is confirmed by Jørgensen (2007), who states that estimation accuracy is still a major topic of interest within IS research.

Important approaches that attempt to determine estimation accuracy emerged already several decades ago (see e.g. Boehm 1981 or DeMarco 1982). But given the huge body of literature about cost overruns of IS projects, there are only very few studies that analyse the estimation accuracy inherent in IS projects. Therefore, only little information about estimation accuracy is available so far. Even worse, since there is only little information available, the interpretation of the resulting values for the estimation accuracy is difficult (see e.g. Eveleens and Verhoef 2009; Little 2005).

Against this background, our paper presents a descriptive case study of a large IS transformation program consisting of more than 90 IS projects at a large German financial service provider (FSP). Within this program, the costs of these IS projects were estimated by experts at different stages of the projects. We were able to gather the cost estimations for 79 IS projects that were estimated according to the same estimation process. Therefore, this data allows for the evaluation of the estimation accuracy. In order to do so, we determined the estimation accuracy for every IS project (each having multiple cost estimates) using different approaches, since this allows for a comparison of the different approaches. To guide the research throughout the paper, the following research questions are posed:

*How was the estimation accuracy within the IS transformation program of the FSP?*

*How suitable are current methods for determining the accuracy of cost estimations for IS projects?*

To answer this question, this paper is organised as follows: The next section provides the theoretical background and therefore cost estimation methods as well as methods for the determination of estimation accuracy are introduced. Further, the research methodology is described. Section 3 presents the case study: After some general information about the IS transformation program, the processes for cost estimations are described. After that, the estimation accuracy is determined using different estimation accuracy measures. Further, the findings of the case study as well as its implications and limitations are discussed. The last section sums up and concludes the paper.

2 Research Background and Research Methodology

First, this section sums up the literature concerning cost estimations methods and measures for determining estimation accuracy. Then, the research methodology, namely a case study, is presented.
In order to derive conclusions about the estimation accuracy of the forecasted costs, it is important to understand how the costs were estimated. There are several different types of cost estimation methods (for a systematic review, the interested reader is referred to Jørgensen and Shepperd (2007)). In the following, we briefly describe expert-based judgements in order to provide the theoretical background for the cost estimation methods used in the IS transformation program.

Expert-based judgements rather rely on past experience of experts. Mukhopadhyay et al. (1992) find – in congruence with Bergeron and St-Arnaud (1992) – that experts perform best when predicting the costs for an IS project. However, as Molokken and Jørgensen (2003) state, expert based estimation techniques can also lead to skewed results due to unclear questions or – as mentioned in the introduction – too optimistic estimates due to biases. There are also several documented applications of expert-based estimations within scientific literature. Within these studies students are often considered as experts and asked for their estimations (see e.g. Höst and Wohlin 1997; Johnson et al. 2000; Ohlsson et al. 1998; Prechelt and Unger 2001). Jørgensen (2004) is the only article known to the authors that evaluates experts’ cost estimations by determining budget overruns in a business environment. He applies cost estimation methods to IS projects in a real-life business environment and reports average budget overruns of 19% and 45% (depending on the chosen estimation method). However, these figures were not used in real-life projects and thus don’t have any management purpose. In addition, the estimated costs – and hence project sizes – are relatively low in all studies. Hence, these studies are hardly comparable to the given research setting, which will be described in greater detail later on.

According to Dysert (2006, p. 1), estimation accuracy is the “degree to which a measurement or calculation varies to its actual value; thus estimate accuracy is an indication of the degree to which the final cost outcome of a project may vary from the single point value used as the estimated cost for the project”. This approach was formalised for instance by Hyndman and Koehler (2006) and denoted as Forecast Error ($FE$). The $FE$ is therefore defined as

$$FE = Act - Est$$

with $Act$ standing for the actual realised costs and $Est$ for the estimated and thus forecasted values. If the $FE$ amounts to 0, the estimation quality can be interpreted as very good, whereas deviations in either direction represent lower estimation quality. A major weakness of the measurement of absolute deviations is that the results do not contain any information about the size of the projects. Therefore it is difficult to compare the results. In order to address this shortcoming, relative instead of absolute figures are often used.

One popular measure for the relative determination of estimation accuracy is the Mean Absolute Percentage Error ($MAPE$) (Hyndman and Koehler 2006). Despite the term ‘absolute’, the $MAPE$ determines the relative deviation of the cost estimations from the actual realised costs. Literature on software cost estimations labels this figure (Mean) Magnitude of Relative Error ($(M)MRE$; cf. Conte et al. 1985; Mukhopadhyay et al. 1992; Vicinanza et al. 1991). The $MAPE$ is defined as follows:

$$MAPE = (Act - Est) / Act$$

The $MAPE$ is criticised for returning distorted results if the $Act$ value is close to zero (especially between 0 and 1). As Hyndman and Koehler (2006) further state, $MAPE$ puts a heavier penalty on positive deviations than on negative deviations, although the deviation is the same in absolute values. Consequently, we will slightly adapt $MAPE$ and introduce the so called $MAPE(Est)$ as a similar measure, which is defined as follows:

$$MAPE(Est) = (Act - Est) / Est$$

It puts the absolute deviation in relation to the estimate instead of the actual value. Further, it has the advantage that, contrary to $MAPE$, in most cases, positive and negative deviations are treated equally. Additionally, this was the measure requested by management in the case study described later.
But despite measuring absolute or relative deviations there are also several methods that were explicitly developed to determine estimation accuracy. Two famous approaches emerged several decades ago, the Cone of Uncertainty introduced by Boehm (1981) and the Estimating Quality Factor (EQF) introduced by DeMarco (1982). The former tool is used to measure deviations between forecasts and realised values by plotting ratios over the duration of the projects, whereas the latter quantifies the quality of forecasts of the estimators by quantifying the deviations between the forecasts and the realised values (Eveleens and Verhoef 2009). According to Eveleens and Verhoef (2009, p. 946), the EQF enables the assessment of „the quality of individual forecasts, but more importantly, the quality of the process of IT forecasting“. Furthermore, the EQF requires estimation data of several points in time during a project and – in contrast to the Cone of Uncertainty – aggregates this information to one key figure. Since the aggregation of estimation accuracy to one key figure has several advantages like e.g. a better interpretation or the comparability, we prefer the EQF to the Cone of Uncertainty and will describe this measure in greater detail in the following.

The EQF compares the actual realised value to different estimates that were made over time. Each estimated value is further weighted with the time the estimate was valid, whereas the realised value is weighted with the total time of the project. Therefore, according to Eveleens and Verhoef (2009), the EQF can be expressed as:

$$EQF = \frac{\text{Area under the actual value}}{\text{Area between forecast and actual value}}$$

According to Shelley (1993), the EQF further has the advantages that it encourages re-estimations and can be applied to a broad variety of IS projects. However, as he further states, it is important to note that – like all methods which rely on actual costs accrued – the insights generated by analyzing the EQF will be of little use to currently ongoing projects, as the EQF is an “after-the-fact measure” (Shelley 1993, p. 76) and therefore its insights are primarily useful for future projects.

Unfortunately there is no precise interpretation of the value obtained via the EQF. But if one aggregates single EQFs to an overall EQF, it is useful to rather use the median than the average value. This is due to the fact that outliers have a stronger influence on the average that on the median value (Eveleens and Verhoef 2009). DeMarco (1982) reported a median value of 3.8 for the software industry. Compared to that, Little (2005) analysed 100 software projects and concluded that “approximately 10 percent of (…) [the] projects had EQFs lower than 2.8, half the projects had EQFs less than the median of 4.8, and 90 percent of the projects had EQFs lower than 11.7”. Eveleens and Verhoef (2009) analyse estimation accuracy for a FSP using the EQF. They report a median EQF value of 8.5. Despite all these results, there is no objective interpretation of EQF values.

Summing up, to the best of the authors’ knowledge, there is only little real world data provided within scientific literature that allows for an interpretation of estimation accuracy. Furthermore, most of the existing studies base their determination of the estimation quality only on single accuracy measures (except for the article written by Eveleens and Verhoef (2009)). Therefore, the aim of this paper is to add knowledge and data to this stream of research by providing a detailed descriptive case study of a large IS transformation program of a German FSP. Within this case study, the estimation accuracy of the costs of a total of 79 IS projects is determined using different estimation accuracy methods.

Despite the huge interest in cost estimation accuracy of IS projects, there is no clear understanding of how to interpret the derived estimation accuracy figures. Literature agrees that given a limited knowledge base as outlined so far, a case study is an appropriate research methodology (Benbasat et al. 1987; Eisenhardt 1989; Dubé and Paré 2003; Yin 2009). Another reason for selecting a case study research strategy is the fact that the IS projects under consideration are contemporary phenomena in a real-life context (Yin 2009, p. 8).

Therefore a descriptive case study was chosen as research methodology in order to describe the accuracy of cost estimations for IS projects. The projects were located within an IS transformation
program at a German FSP. As the cost estimation rules were already established within the program and therefore the behaviour of the participants could not be controlled, an in-depth case research approach was chosen (Yin 2009). This approach allowed the researchers to study the phenomenon in its natural setting. Due to this positivistic character of the context, a descriptive case study seems adequate; it constitutes the next iteration (Eisenhardt 1989, p. 546) towards a theory on measuring and interpreting the quality of cost estimations for IS projects.

During the whole duration of the program (from winter 2008 until autumn 2012), the data on the cost estimations of the IS projects was centrally collected in the program management office (PMO). The PMO was the organizational unit that was responsible for gathering, storing, and publishing the cost estimations for all IS projects in the different phases. Within the first two years, the data was kept within an excel spreadsheet; at a later point during the program, the excel spreadsheet was amended by a database. Regarding the cost figures, there were two data sources: On the one hand, the costs for human resources were computed automatically within a time-capturing tool (via multiplying the captured hours by an hour rate). However, this way the time capturing tool did not include other cost than human resources, as for instance fix prices, software, hardware, etc. That is why, on the other hand, the costs figures for the projects were delivered monthly from the accounting department via an excel spreadsheet; its figures were exported from SAP accounting systems. As the data from the accounting department were more complete and the FSP included them in its books, they were used for the data analysis. The information on the dates of the cost estimations was taken from different sources, depending on the particular phase. The dates for the baseline budget derived from presentations to the management board during the beginning of the program, in which the budget was formally approved; the date for the all other phases were taken from the official (weekly) project status reports which included the current milestones for a project specifically. Each milestone was a particular date and the end of each of the phases matched with a milestone. The presentations to the management board as well as the status reports were also centrally prepared and published by PMO.

The data on both, the cost estimations and the cost figures for the particular IS projects, were centrally stored in an excel spreadsheet and was made available to the researchers in this form. As a consequence, the data was comfortable to analyse and apart from normalisation (for reasons of confidentiality), no codification was needed.

The process of estimating costs (incl. activities, responsibilities, point in time) was defined in the ‘governance guidelines’: The purpose of this document was to establish processes and structures particularly for the IS transformation program. Thus, the process of estimating costs was standardised for all IS projects. PMO was responsible for setting up and updating the governance guidelines.

3 Determining Estimation Accuracy at a German FSP

As already mentioned, the case study was conducted within an IS transformation program at a large German FSP lasting from 2008 to 2012. The planned budgets amounted to a total of more than 100 Mio. EUR. The main goal of this program was to consolidate the scattered IS landscape while simultaneously reducing process costs and enhancing key functionalities. In the following, we first discuss how the project portfolio was set up and how the projects were structured in order to provide the reader with a clearer understanding of the background. Second, we describe how and by whom the costs of the IS projects were estimated. Third, we determine the accuracy of the cost estimations.

3.1 Project Setting

Before any IS project was planned or even conducted, all business domains that were involved in the transformation first had to define so-called management requirements. Each of these requirements defined a particular scope, which comprised the transformation or creation of an existing or new information system. At this early stage of the program, the corresponding scope description was not
very detailed. Nevertheless, each management requirement was labelled with associated costs estimated by experts (details on the expert-based estimations are given in section 3.2). This way, the management board was able to decide on whether or not to implement these management requirements already at the very beginning of the IS transformation program. Next, the management requirements that were to be realised were grouped into several releases according to their priority and temporal interdependencies. The releases succeeded each other (with partly overlapping time periods).

Second, based on the management requirements, the IS projects were defined within the first phase of a release: the work it took to establish a solution to a management requirement was usually organised within one project; sometimes, several management requirements were combined into one IS project.

Third, within IS projects, the project work was done by two major groups: The IS demand teams on the one side and the IS supply teams on the other side. The IS demand teams were responsible for the business side of the IS projects in terms of defining the requirements as well as designing and executing test cases. The IS supply teams were each responsible for a particular application and had to make sure that these applications are adapted/designed in a way that they meet the requirements.

From a timeline perspective, the IS projects were organised along successive phases: During the first phase of the actual conduction of the project (‘scope agreement’), the IS demand teams broke down the management requirements into more detailed project requirements in close collaboration with the demanding business domains. In the next phase of the project (‘specification’), both IS demand teams and IS supply teams created a specification document which denoted a precise description of the business requirements. Again, this was done on the granularity level of the project requirements. Thereafter, the scope of the project was considered to be stable and the implementation work began. However, even after the phase ‘specification’, cost estimations were changing due to changes within the requirements. Such changes are managed through change requests.

During the phase ‘delivery’, the software was implemented and tested. The phase ended with the milestone ‘Go live’. After the ‘Go live’ there was a six-week period of ‘post-go live support’, after which the project was over (‘project end’).

During the overall IS transformation program, 6 releases and 97 IS projects had been completed. It can be said that the scope that was implemented possessed good quality. Furthermore, any time overruns materialised in cost overruns, which is why we focus on the cost estimation in the following.

### 3.2 Cost Estimation

After outlining the organisation of the IS projects, the cost estimation process is described in the following. The description is split into two parts: First, we describe the generic cost estimation structure, i.e. the actual cost estimating approach, according to which the experts estimated the costs in each of the projects’ phases. Second, we describe important aspects of the cost estimations that differ within different phases.

All cost estimations were based on the requirements (i.e. either management or project requirements). For each requirement, it was estimated how many internal and external person days (PD) were necessary for both the IS demand teams and the IS supply teams. The estimated PDs always included the total effort for an IS project in terms of the following activities: analysis, design, specification, implementation, test, post roll-out support and hand-over to maintenance. Internal staff was priced with a daily rate of 0 EUR, since the budget controlling was focused on additional accounts payable caused by the IS projects and internal staff members therefore did not cause any (additional) costs. In contrast, for each group of externals a daily rate was estimated by the experts and afterwards multiplied with the corresponding PDs. Non person-related investments (e.g. hardware, software) were also estimated in each phase as a monetary sum, but account only for a relatively small portion of the whole program budget.
The cost estimations in all phases were done according to the structure depicted in Table 1:

<table>
<thead>
<tr>
<th>Requirement (Person Days)</th>
<th>IS Demand Team A</th>
<th>IS Demand Team B</th>
<th>...</th>
<th>IS Supply Team X</th>
<th>IS Supply Team Y</th>
<th>...</th>
<th>Software</th>
<th>Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirement (Person Days)</td>
<td>40</td>
<td>50</td>
<td>20</td>
<td>210</td>
<td>320</td>
<td>12</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>Daily Rate (EUR)</td>
<td>500</td>
<td>0</td>
<td>400</td>
<td>0</td>
<td>600</td>
<td>0</td>
<td>700</td>
<td>0</td>
</tr>
<tr>
<td>Budget Requirement (EUR)</td>
<td>20,000</td>
<td>0</td>
<td>8,000</td>
<td>0</td>
<td>192,000</td>
<td>0</td>
<td>31,500</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Generic Cost Estimation Structure (Example)

As mentioned earlier, the costs were not only estimated in the beginning, but also during several phases of each IS project. The reason for this (as documented in the governance guidelines of the IS transformation program) was the assumption, that the more precise the requirements were specified in later phases, the more accurate the needed budgets could be estimated.

Over the project phases there were differences concerning the estimation base (How granular was the requirements description broken down?), and the estimator (Who were the experts to estimate?). The main characteristics of the cost estimation process are depicted in Table 2.

<table>
<thead>
<tr>
<th>Project Phase</th>
<th>Program Planning</th>
<th>Scope Agreement</th>
<th>Specification</th>
<th>Delivery</th>
<th>Project End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Stage</td>
<td>Initial estimation</td>
<td>Scope-based estimation</td>
<td>Specification-based estimation</td>
<td>Specification-based estimation plus change requests</td>
<td>Costs determination</td>
</tr>
<tr>
<td>Estimation base</td>
<td>Management Requirement</td>
<td>Project Requirement</td>
<td>Project Requirement</td>
<td>Project Requirement</td>
<td>Project</td>
</tr>
<tr>
<td>Estimator</td>
<td>Program Management</td>
<td>IS Demand Team Project Leads, IS Supply Team Leads</td>
<td>IS Supply Team Leads</td>
<td>IS Demand Team Project Leads, IS Supply Team Leads</td>
<td>Accounting</td>
</tr>
</tbody>
</table>

Table 2. Different Characteristics of the Estimation Process in different Project Phases

**Phase I: Program Planning**: At the beginning of the IS transformation program an initial estimation for the whole program was conducted by three experts, which were members of the program management. During a period of four weeks, the three experts sat together in several meetings that lasted three to six hours. The goal was to agree on the estimations in terms of PDs for each management requirement and the daily rate per team accordingly. There was no formal structure for the meetings besides an excel file representing the estimation structure as given in Table 1. The experts used a task-based approach by asking “What must be done to meet this requirement?” and assigning a number of estimated person days to the discussed task. Another approach was to ask for the number of people that are necessary to fulfil a requirement and thus derive the necessary number of PDs. It should also be mentioned that – during the meetings – the three experts also contacted other experts (for instance via phone) for further information on the required scope or on necessary effort. The resulting estimations were gathered in the excel file mentioned earlier and proposed to the management board for approval and used as budgets afterwards.

**Phase II: Scope Agreement**: The next cost estimation took place after the IS projects had been defined. Compared to the initial estimation, two characteristics changed in this phase: First, the estimations were now done based on a more detailed description of the scope (so called ‘scope agreement’). The IS demand teams had about 2 month time for (a) breaking down the scope of the
management requirements, which should be realised through the particular project, into several project requirements and (b) providing a more detailed description of the associated scope. Consequently, the costs were then estimated for each project requirement. Second, the experts changed: The IS demand team project lead was responsible for the IS demand team estimations, whereas the head of each IS supply team estimated the efforts for their applications. Both roles could be considered as more experienced in their particular subject than the original three experts. These ‘scope-based estimations’ were collected according to the structure of Table 1.

**Phase III: Specification:** During the ‘specification’ phase (lasting 3 months), the IS demand teams detailed the project requirements further. However, no additional breakdown of the project requirements took place. Whereas the budget of the IS demand teams remained the same as after the scope-based estimation, the IS supply teams were allowed to adapt their estimated costs based on the new insights generated through the more detailed project requirements. Again, the structure of the estimations followed Table 1.

**Phase IV: Delivery:** After the ‘specification’ phase, the estimated costs were supposed to be stable for each IS project. From that time on, costs – and thus budgets – could only be adapted via change requests. Change Requests were the means to deal with changing requirements during the project work on the one hand and wrong planning assumptions on the other hand. As a consequence, change requests contained estimations that followed the same level of detail and structure as in previous phases. Whereas the IS demand teams had to submit change requests for changing their cost estimations after the phase ‘scope agreement’, the IS supply teams had to do this after the phase ‘specification’. Each change request had to be approved by program management.

**Phase V: Project End:** The final data series are actual realised costs for all IS demand and supply teams per project. As a consequence, there are five cost values for each project; four cost estimation values (by experts in different phases of the projects) and the resulting actual costs. They serve as benchmarks to which the different costs estimates provided by the experts are compared in section 4.3.

As can be seen, all data used for this case study were officially recorded; hence, they are documentary in its nature. Consequently, the authors do not consider further data triangulation necessary in order to answer the research question in an objective way.

### 3.3 Cost Estimation Accuracy

In this section, the accuracy of the cost estimations is analysed using the measures introduced in section 2 (i.e. $EQF$, $FE$, and $MAPE(Est)$) and the results are discussed. Thus the cost estimates of the different project phases described in section 3.2 were used (we aggregated the estimates from the IS demand and IS supply teams) and the estimation accuracy measures were computed via Excel. First, we provide some examples for the computation of the three metrics and discuss their interrelation. Second the overall estimation accuracy of the whole transformation program is analysed from a top down perspective. Third, the estimation accuracy of the 79 single IS projects of the transformation program is determined. The corresponding results are depicted in Figure 1, where all IS projects are ordered by descending $EQF$ (Figure 1a). Further, the $FE$ and the $MAPE(Est)$ were calculated for each IS project, whereas the order of the projects within the Figures 1b/1c stays the same as in Figure 1a.

In order to illustrate the computation of the metrics we use project 19 as an example. Starting with the $EQF$, we note that for each project there were four time periods (lasting $t_1 = 81d$, $t_2 = 98d$, $t_3 = 248d$, $t_4 = 42d$) leading to five estimations $e_1 = 1.887.546$ EUR, $e_2 = 2.069.734$ EUR, $e_3 = 2.477.480$ EUR, $e_4 = 2.437.378$ EUR and actuals of $a_5 = 2.360.425$ EUR. The $EQF$ for project 19 was then calculated as:

$$
EQF_{19} = \frac{(t_1 + t_2 + t_3 + t_4)a_5}{t_1abs(e_1 - a_5) + t_2abs(e_2 - a_5) + t_3abs(e_3 - a_5) + t_4abs(e_4 - a_5)}
$$
The $FE$ for project 19 was calculated as $a_5 - e_1 = 472.879$ EUR, $MAPE(Est)$ as $(a_5 - e_1) / e_1 = 25.05\%$. It can be seen that all formulas include the difference between the actual accrued costs $a_5$ and the initial estimation $e_1$. The $FE$ consists of this difference only, where the $MAPE(Est)$ puts this difference into perspective by relating it to the initial estimation. Both perspectives were relevant in the IS transformation program: When addressing over-/underspending of budgets (especially between start and end of the projects), the $FE$ was relevant in order to decide on (re-)allocating budgets, as these were decisions on amounts of money. In contrast, $MAPE(Est)$ was used for discussions on estimation quality itself, as a deviation of 50,000 EUR was considered as relative large for a project with an initial estimation of 100,000 EUR, whereas the same deviation was hardly discussed for projects with an initial estimation of 1,000,000 EUR. Both, $FE$ and $MAPE(Est)$, can have negative values which was also a valuable information for management in terms of over-/underspending of budgets. The $EQF$ also relies on the difference between actuals and initial estimation, but apart from this, differs in several aspects. First, it uses the absolute values of the differences (a not uncommon perspective of deviations), i.e. the (for management purpose often valuable) information on over-/underspending of budgets is lost. Second, the $EQF$ does not only incorporate the difference between initial estimation and actuals, but also between all other estimations and the actuals and weighs these differences with the duration how long these estimations were valid. This means, that the $EQF$ can provide the same direction of judgement as $FE$ and $MAPE(Est)$, but – as will be illustrated later – it can also judge quite differently. The example above also illustrates how the estimations vary across the different phases: Whereas the estimations rise from the initial until the specification-based estimation, they drop afterwards again towards the actuals. However, the actuals are still clearly higher than the initial estimation.

Concerning the results, it can be stated – from a top down perspective – that for the overall IS transformation program the average $EQF$ value is 8.09, whereas the median only accumulates to 4.03. According to the discussion about the interpretation of the $EQF$ (cf. section 2), we can conclude that the overall estimation accuracy based on the $EQF$ is better than the results of the IS industry (DeMarco 1982), but worse to the result of the FSP (Eveleens and Verhoef 2009).

Regarding the $FE$, all deviations amount to an overall deviation of 26.66 million EUR (in this calculation, positive and negative deviations compensate each other, as on the level of the complete project portfolio of the program, budgets not spent for project A could be spent for project B). Comparing this deviation to the 129.5 million EUR initial cost estimates leads to the third indicator, i.e. a $MAPE(Est)$ of 16.54\%. Unfortunately, literature doesn’t provide defined thresholds that allow for an objective interpretation of the value determined by using the $MAPE(Est)$: For instance, Whitaker (1993) denotes a $MAPE(Est)$ value greater 30\% as ‘project failure’ and values greater 50\% as ‘serious budget overruns’. According to Eveleens and Verhoef (2010), the Standish group calls projects with $MAPE(Est)$ values greater 0\% ‘challenged’; Compared to the more arbitrary thresholds 30\% and 50\%, 0\% comes closest to an objective threshold, as it separates over- from underestimation. Hence, using this scale, the overall program can be considered as “challenged”, but it is far from “failure”. Comparing the $MAPE(Est)$ of 16.54\% of the IS transformation program to the $MAPE(Est)$ values of several studies summed up by Eveleens and Verhoef (2009, p. 983), we can state that the value of the IS transformation program is amongst the lowest (minimum 13\%) of these values. Consequently, based on $MAPE(Est)$, the estimation accuracy can be considered as good. It can be seen that $EQF$ and $MAPE(Est)$ lead to somewhat differing judgements on the estimation accuracy of the IS transformation program. Another fact is noteworthy in this context: The program management added a contingency buffer to the cost estimations during the phase ‘initial estimation’:

$$EQF_{19} = \frac{(81 + 98 + 248 + 42) \times 2.360,425}{81 \times 427.879 + 98 \times 290.691 + 248 \times 117.055 + 42 \times 76.953} = 11.18$$
Figure 1. Estimation Accuracy Measures of all 79 IS projects.
This amounted to 25% for the first three releases and 15% for the forth and fifth release. I.e., program management already anticipated some inaccuracy in the initial estimations. Taking into account this contingency buffer, there was no budget overrun on the level of the whole IS transformation program.

Taking a look at the single IS projects, we observe \( EQF \) values ranging from 55.49 (project 1) to 0.29 (project 79). As DeMarco (1981) considers an \( EQF \) value higher than 4.00 as a figure for good estimation accuracy, we require a bit more and put an \( EQF \) value of 5.00 as the corresponding threshold. Thus, we can state that 34 of the 79 IS projects show good estimation accuracy (“well estimated”), whereas the remaining 45 IS projects can be considered as “poorly estimated”. The “poorly estimated” IS projects caused 81.20% of all budget overruns.

It is important to note that we were not able to find any connection between the project size (measured through costs) and the corresponding accuracy estimate. Further, the analysis of the \( FE \) values of the IS projects goes in line with the results obtained by the \( EQF \): The higher the \( EQF \), the lower the \( FE \) tends to be. The same holds true for the analysis of the \( MAPE(Est) \). As Figure 1c shows, the \( MAPE(Est) \) tends to increase with decreasing \( EQF \). This is shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Well estimated IS projects</th>
<th>Poorly estimated IS projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median ( FE )</td>
<td>17,979 EUR</td>
<td>220,333 EUR</td>
</tr>
<tr>
<td>Median ( MAPE(Est) )</td>
<td>2.81%</td>
<td>50.77%</td>
</tr>
</tbody>
</table>

Table 3. \( FE \) and \( MAPE(Est) \) of all IS projects

Given the information of the different estimation accuracy measures on the same IS projects as shown above, we will now discuss the results. One difference that attracts the reader’s attention is that the \( EQF \) leads to positive values, whereas the \( FE \) and the \( MAPE(Est) \) result in positive and negative values. As Figure 1 reveals there is no connection between the value of the \( EQF \) and the fact whether \( FE \) or \( MAPE(Est) \) are positive or negative. This is due to the definition of the \( EQF \), which takes positive and negative deviations equally into account, since a deviation from the estimated costs in either direction is considered as not desirable and represents bad estimation accuracy. This two-side perspective on deviations represents the management’s ex-ante view on estimation accuracy when deciding which IS projects to include in the portfolio, because deviations in both directions lead to suboptimal decisions: If projects are underestimated at this stage, more effort will be needed during the project, which leads to higher costs. In contrast, if projects are overestimated, too much budget is allocated towards them. Consequently, other projects might be denied due to missing budget. However, given the fact that the projects were successful in terms of scope as mentioned earlier, lower-than-expected costs are more preferable than higher costs in that real business environment. This especially holds true during the projects, when the “saved” costs might be urgently needed in other projects. Hence, \( FE \), \( MAPE(Est) \), and other measures which distinguish between positive and negative deviations are suitable for both project controlling and evaluating estimation accuracy, whereas \( EQF \) and similar measures rather focus on evaluating the estimation accuracy. Against the background of the case study, the authors can report that the \( EQF \) was not used for reporting during the IS transformation program; cost estimations were – upon request from management – put in relation to the estimations of earlier phases in terms of absolute (\( FE \)) or relative (\( MAPE(Est) \)) values.

Despite the major correspondence between \( EQF \), \( FE \), and \( MAPE(Est) \) as reported in Table 3, there are also some projects which are judged quite differently according to the three measures:

Considering project 5 e.g., it can be stated that although the overall estimation accuracy can be considered as rather good (\( EQF = 26.32 \)), the \( FE \) value was -525.485 EUR and the \( MAPE(Est) \) amounted to -9.07%. As mentioned earlier, the amount “saved” might have helped to add additional resources to other projects which were underestimated. In any case, at least 20 out of the 79 projects
could have been financed out of the saved amount of money based on their ‘initial estimation’; it seems discussable whether the judgement of the \( EQF \) as good is adequate for such a misallocation of budget and resources. **Project 19** is another project which is judged quite differently according to the three measures; whereas its \( EQF \) of 11.18 seems to be a good estimate (referring to a \( EQF \) threshold value of 5.00), the “initial estimation” was overrun by +472.879 EUR (\( FE \)), corresponding to a \( MAPE(Est) \) of +25.05%. Taking into account that the additionally needed budget could not be spent otherwise, resources were blocked for other tasks. Further, since the \( MAPE(Est) \) value is close to a project failure (according to the 30% threshold by Whitaker (1993)), an evaluation as ‘good’ might seem inadequate. However, the high \( EQF \) can be explained as follows: Indeed, the “initial estimation” is clearly differing from the accrued costs, but this deviation was detected already early in the project and the initial estimation was replaced by a more accurate one. As this second and more accurate estimation was valid for a longer time period, it got a higher weight and thus outweighed the more inaccurate initial estimation. Similar considerations can be made for the projects **28, 31, and 33**: All three projects have an \( EQF \) greater 5.00 and therefore possess a good estimation quality, but they all have a \( MAPE(Est) \) of greater +30.00% and would be therefore considered as failure by Whitaker (1993). Further, two of them have a \( FE \) of more than +1,000,000 EUR.

If we analyse IS projects that were well estimated according to \( MAPE(Est) \), we reveal the results depicted in Table 4. It seems that \( MAPE(Est) \) values around 15.00% come along with an \( EQF \) of below 5.00. Assuming that 5.00 divides good from bad estimations, a \( MAPE(Est) \) value 15.00% might also be a suitable value to distinguish a good estimation accuracy from a bad one.

<table>
<thead>
<tr>
<th>( MAPE(Est) ) Threshold</th>
<th>Number of IS projects (( EQF &gt; 5 ))</th>
<th>Number of IS projects (( EQF &lt; 5 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 12.00%</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>≤ 15.00%</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>≤ 20.00%</td>
<td>33</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table 4. Number of IS projects related to different thresholds of \( MAPE(Est) \)**

Besides comparing the judgments expressed via the three different measures, we also examined whether there was any correlation between the estimation accuracies within the different releases. Table 5 depicts the results of the metrics for the five releases.

<table>
<thead>
<tr>
<th>Release</th>
<th>( EQF )</th>
<th>( FE )</th>
<th>( MAPE(Est) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Average</td>
<td>Median</td>
</tr>
<tr>
<td>1</td>
<td>6.67</td>
<td>4.68</td>
<td>62,912</td>
</tr>
<tr>
<td>2</td>
<td>8.10</td>
<td>9.42</td>
<td>-767</td>
</tr>
<tr>
<td>3</td>
<td>3.62</td>
<td>5.06</td>
<td>24,310</td>
</tr>
<tr>
<td>4</td>
<td>4.55</td>
<td>6.85</td>
<td>82,429</td>
</tr>
<tr>
<td>5</td>
<td>2.57</td>
<td>4.27</td>
<td>821,844</td>
</tr>
</tbody>
</table>

**Table 5. Median and Average for \( EQF \), \( FE \), and \( MAPE(Est) \) of project in different releases**

It can be seen that the different metrics only agree on the best (release 2) and the worst (release 5) estimation accuracy. Releases 1, 3, and 4 are ranked in between, but the order differs depending upon which measurement is chosen. One can also note that estimations did not get better from release to release, as one could have expected: The estimation accuracy of the last release is the worst. Analysing the reasons for this is subject to further research.
Summing up, we can state the following answers to our research question:

(a) The estimation accuracy of the IS transformation program can be considered to be average. The different measures lay within already documented ranges that partially were considered as acceptable. However, it is interesting to see that the contingency buffer (initiated by the program management at the beginning of the program) almost exactly covered the cost overruns. Therefore, a $\text{MAPE(Est)}$ of 16% seemed to be acceptable and was even anticipated by the IT executives of the FSP.

(b) We found that estimation accuracy measures judge cost overruns differently. As we elaborated above, the use of multiple measures provides valuable information about the real estimation accuracy. Especially the $\text{EQF}$ has major drawbacks when the measure should be used for management purposes: First, its results are not normalised and the values are hard to interpret. Second, the information of under-/overestimation is lost. On the contrary, the $\text{EQF}$ is the only metric that takes into account more than one estimation.

### 3.4 Implications and Limitations

The following implications for researchers’ future work can be derived: First, it seems that there is room for improvement regarding measures for estimation accuracy in IS projects. $\text{FE}$ and $\text{MAPE(Est)}$ are practically useful measures, but they only take into account a single estimation and the accrued cost. In contrast, $\text{EQF}$ aggregates several cost estimations made at different points in time. Researchers might keep this advantage of the $\text{EQF}$ while developing it further in order to (a) align its results with the ones of other measures (b) incorporate requirements like normalisation, interpretability, or aggregation in order to make it more comprehensible for management purposes. Practitioners might benefit from the described process of cost estimations that provided good results. Moreover, the discussion of the different estimation accuracy measures might guide the selection of a suitable measure in real-life business environment. The first limitation consists of the fact that we only studied IS projects of one particular company. Although the data set can be considered very rich, an application of similar studies at different companies is desirable. Further, we only conducted a descriptive case study by answering the research question “How good or bad was the estimation accuracy?”. Thus, the study does not answer the question “Why was the estimation accuracy good or bad?”. Answering this question is also subject for further research.

### 4 Conclusion

Against the background of a large number of reports on budget overruns of IS projects, we described the cost estimation process and the cost estimation accuracy of 79 IS projects in this case study. Therefore, the estimation accuracy measures $\text{Estimating Quality Factor}$, $\text{Forecast Error}$, and $\text{Mean Absolute Percentage Error}$ were applied. Whereas for some IS projects serious budget overruns could be identified, the overall estimation accuracy of all projects in terms of the $\text{MAPE(Est)}$ was amongst the lowest reported in literature. In addition, the results of different measures of cost estimation accuracy were compared to each other. Thereby it became apparent that the estimation accuracy measures used in this paper judge accuracy differently: Whereas the measure $\text{EQF}$ takes into account all estimations made during the project as well as the duration how long these estimates were valid, $\text{FE}$ and $\text{MAPE(Est)}$ only focus on a single estimation and the accrued costs. Further, the $\text{EQF}$ doesn’t account for the direction of the deviations.

Given the results obtained in this case study, we agree with Eveleens and Verhoef (2009) and believe that information about the estimation accuracy – and thus about the quality of the cost estimation process – is crucial for IT executives. However, improving the quality of the cost estimations is only one step towards lower budget overruns of IS projects. Another element are measures which are taken to make the costs meet the estimations during the duration of the projects. The authors are currently working on a theory explaining the effect of budget control on budget overruns of IS projects.
References


