Kristian Sandaa

What Causes Schedule Overrun in Complex Engineering Projects?

Master’s thesis in Mechanical and Industrial Engineering
Supervisor: Nora Johanne Klungseth
Co-supervisor: Nils Olsson
June 2021

Master’s thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

Kristian Sandaa

What Causes Schedule Overrun in Complex Engineering Projects?

Master's thesis in Mechanical and Industrial Engineering
Supervisor: Nora Johanne Klungseth
Co-supervisor: Nils Olsson
June 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

Abstract

This study is a deep dive into the estimated work hours needed in complex engineering projects. Through exclusive access to detailed financial records and early phase estimates, this study analyses the estimation accuracy of one of Europe's leading technology- and industry companies. Based on a sample of 7 recent projects, carefully selected for their unchanged scope during their lifetime, the estimations are categorized and analyzed in detail. A literature review is also performed across multiple industries to establish comparable baselines. The analyzed projects totals almost 80 000 work hours and were found to have an average overrun of 46.21%, with 5 out of 7 projects underestimating their needed time. The projects are evaluated in 6 main categories following the company's work-breakdown structure, as well as many sub-categories. Engineering alone represents over 50% of all project hours and is on average underestimated with 55.69%, while project administration had the best accuracy with only 24.25% overrun on average. Out of the different engineering categories, software engineering was found to have the least accurate estimates, with each activity estimate being about 460 hours off target on average - more than 2.5 times that of the second-worst category.

When looking into the lowest levels of the estimates, it is discovered that only 43% of activity estimates are actually underestimated and that the size of the average overestimation is almost twice as big as the underestimation, with the average activity estimate hitting 92 hours below target. In addition, it is found that a total of 43.6% of all project hours are accounted to activities that the project team did not initially consider, indicating that more meticulous planning in the early phases would likely help avoid these overruns. When comparing to other studies, we find the average cost overrun of the company is comparable to the European standard within similar industries.

Sammendrag

Denne studien er et dypdykk i estimerte arbeidstimer til komplekse ingeniørprosjekter. Gjennom eksklusiv tilgang til detaljerte økonomiske arkiver og estimater fra tidligfase, analyserer denne studien estimeringsnøyaktigheten til et av Europas ledende teknologi- og industriselskaper. Basert på et utvalg av 7 prosjekter ferdigstilt de siste 8 årene, nøye valgt for deres uendrede scope, er estimatene kategorisert og analysert i detalj. En litteraturgjennomgang blir også utført på tvers av flere bransjer for å finne sammenlignbar data. De analyserte prosjektene har sammen nesten 80 000 arbeidstimer og har en gjennomsnittlig overskridelse på 46,21%, med 5 av 7 prosjekter som undervurderte den nødvendige tiden. Prosjektene evalueres i 6 hovedkategorier etter selskapets WBS, samt mange underkategorier. Ingeniørarbeid alene representerer over 50% av alle prosjekt timer og er i gjennomsnitt undervurdert med 55,69%, mens prosjektadministrasjonen har den beste nøyaktigheten med bare 24,25% overskridelse i gjennomsnitt. Fra de forskjellige ingeniørkategoriene ble det funnet at software hadde de minst nøyaktige estimatene, med hvert aktivitetsanslag i gjennomsnitt ca. 460 timer fra estimatet, som er mer enn 2,5 ganger den nest verste kategorien.

Når man ser på de laveste nivåene av estimatene, oppdages det at bare 43% av aktivitetsestimatene faktisk er undervurdert, og at størrelsen på den gjennomsnittlige overestimeringen er nesten dobbelt så stor som undervurderingen, med et gjennomsnittlig aktivitetsestimat som treffer 92 timer under estimatet. I tillegg er det funnet at totalt 43,6 % av alle prosjekttimene er regnskapsført på aktiviteter som prosjektgruppen i utgangspunktet ikke planla for, noe som indikerer at mer grundig planlegging i de tidlige fasene sannsynligvis vil bidra til å unngå disse overskridelsene. Sammenliknet med andre studier, finner vi at den gjennomsnittlige kostnadsoverskridelsen til selskapet er sammenlignbar med den europeiske standarden innen lignende bransjer.

Table of Contents

List of Figures

List of Tables

1 Introduction

Cost is traditionally seen as one-third of the iron triangle of time, cost, and quality, factors on which a project's success is measured (Pollack, Helm, and Adler 2018; Atkinson 1999). Even though modern organizations and literature have started realizing that the success of a project is more complex than that (Jha and Iyer 2007; Pollack, Helm, and Adler 2018; Toor and Ogunlana 2010), cost remains an important parameter to measure and control. In order for a project manager to successfully hit the cost target, it is naturally critical that the estimations from the early phase are correctly done and that all parts of the project are accounted for, with research suggesting that as much as 70-80% of total product cost is determined in the early phases of product design (Rehman and D.Guenov 1998; Ou-Yang and Lin 1997), and a significant portion of said cost goes to work-hours. Due to the inherent uncertainty of any project undertaken, this estimation task is often far from trivial. This study seeks to gain insight into the accuracy of the time estimations done for the technologically complex projects undertaken by one of the largest technology- and industry companies in Europe and how the uncertainty is distributed throughout the project organizations.

Although a significant amount of work has been put into research in measuring the estimation accuracy of a large number of different project types, separating them on size in terms of both budget and time, complexity, type of industry, the method used for estimation in the early phase and many more, it appears as little attention has been given to how the different parallel parts of a complex project organization perform. This report is a follow-up study on the cost estimation accuracy study (Sandaa 2020) (Unpublished Specialization Report) on the anonymized company Techstick. The previous report delivered an overview of the general performance of the estimations and how they were created, further detailed in Section 2.1. Still, it offered little insight into the source of the observed results. It became apparent that further investigation was necessary into the underlying factors in the observed results and where the observed over- and underruns had their roots. As the investigated organization runs their projects in a similar way as many other companies appear to do, with all project departments estimating their work effort and cost the same way within the same project, the question arose whether or not some departments would benefit more from this than others.

1.1 Problem statement

This report aims to increase the resolution of the estimation measurements by looking into the most detailed parts of the breakdown structure of the different parts of the project organization. By separating all the different parts of the investigated projects into their belonging departments, such as software, hardware, electrical work, production, management, and more, and compares their individual results to the others, the goal is to identify the most uncertain parts of the projects undertaken. As the prestudy indicated that the work-hours estimated in early phase could be a good place to gain further insight, this will be the primary focus instead of the fixed costs. By comparing the data with similar research done by others, the objective is to establish a benchmark for expected results. The overall goal is to better our insight into the most uncertain parts of technologically complex projects and to what extent the project organization can predict uncertainty elements that affect the results.

1.2 Research Questions

1. **What parts of the estimated work experience the most deviation from the actual work done?**

   By looking into each and every estimated work package, down to the lowest level in the project estimation, and comparing it to the actual hours used at project end, one can attempt to use statistical methods to determine what parts of the projects deviate furthest from the estimates.

2. **How does the estimation accuracy compare to other similar industry analyses?**

   By gathering similar data from research done in a wide variety of other areas and comparing their results with Techstick's overall and sub-performance, it will be attempted to establish multiple industry benchmarks for reference.

3. **What are probable explanations for the observed results?**

   Discussing the data against internal structures and results uncovered from the interviews in the prestudy (Sandaa 2020) and observations made by other researchers will hopefully further increase the understanding of the results.

1.3 Background Information

Given that the prestudy to this report in an unpublished report, an introduction to the company in question and their structure will be given next. TechStick is a very large technology company situated in Europe. They employ a 5-figured amount of people and have yearly revenues in the size of billions of dollars. TechStick is a technology and industry company that typically undertakes highly advanced projects requiring technology for a multitude of different areas. The company is organized into several departments sorted by expertise and area of responsibility, e.g., mechanical, electrical, software, and so on (this will be further elaborated on in Section 3.3). Multiple of these departments exist across the company, sorted under the different product families in the company. This research will focus on the performance of the projects undertaken by one of these product families. When a new project within a particular product type is initiated, a project team of people from all necessary departments is put together to fill all roles required for the project. Employees typically work on 2 or 3 different ongoing projects at a time, with varying amounts of time allocated to each. Each delivery project is often also used for development to further advance the technology of their product families and therefore naturally inherits uncertainty regarding development, testing, and debugging time.

1.4 Structure of Report

Following the introduction is the literature review, Section 2. This report consists of a three-part literature review that elaborates on the theory and relevant research needed for this analysis. The first part is a very brief overview of the prestudy, covering the method followed and the most important results. The second part reviews basic estimation theory as well as the history and workings of the Work-Breakdown-Structure, a central part of this report. The third and final part presents a comprehensive overview of similar research done in other studies. Section 3 is a thorough explanation of the methodology utilized in this report, including how the data is gathered and processed, privacy, formulas, the method for literature review, and more. Section 4 presents the data collected and the empirical results found in the analysis, as well as an overview of the results gathered in part three of the literature review. Section 5 is the final part of this report and will consist of a structured discussion of the three research questions selected for this report.

2 Literature

This chapter will primarily be a review of relevant literature needed to correctly answer the chosen research questions. As this work is closely related and dependent on the work done and results found in the prestudy, the first section will summarize the method followed, and the results found there. Thereafter, the most relevant literature on estimation techniques and practices will be discussed before a more comprehensive literature review of relevant, similar studies will be presented. The focus will be on gathering an overview of how different industries performers with regards to their cost estimates, in an attempt to gain comparable data to the findings in this study. In addition to this, some insight into the theories for the observed results will be presented.

2.1 Previous work

This thesis is a follow-up study on the unpublished specialization report (Sandaa 2020), where the project cost estimation done by the anonymized company TechStick was analyzed. Due to the fact that the report is unavailable for the public and the relevance to this thesis, a summary of the report and the most important findings will be presented in this section.

The study utilized a combination of qualitative and quantitative methods to investigate how TechStick does their cost estimations ahead of projects and how they performed. In order to answer this, the study worked to answer the following research questions:

1. What methods are currently utilized as their estimation methods, and what is the intended purpose of the implementation?
2. How accurate are the cost estimations done in the early phases of projects?
3. What could be the possible observed reasons for the performance of the estimates?

Through meticulous study of the internal project documentation and interviews with the respective project managers, a suitable date set of projects were selected for further analysis. After ensuring that the included projects had all the required information available in the documentation, sustained no significant changes to the scope after initiation, and satisfied the remaining selection criterias, the project could be included in the portfolio. This naturally limited the amount of projects that could be included but was deemed necessary to ensure the reliability of the results. A total of 10 delivery projects were included in the final data set. In the empirical analysis, the projects were measured based on three data points each; an early estimate done prior to contract signing, a new and more thorough estimate done by the project team at project initiation, and the final results after project completion.

In addition to quantitative analysis, the author reviewed the internal procedures described in the company's own documentation and performed a series of interviews with different employees to further complement the written findings. A total of 12 interviews were conducted with project engineers, project managers, tender offer managers, and technical marketing support members. They were all asked questions regarding the procedure they followed when making an estimate and their own experiences and roles in the process. The results were finally compared with the empirical findings as well as the findings from the most relevant literature. Figure 1 illustrates the identified project process in the early phases, as well as at which points the data was taken (marked by white arrows). Where the prestudy gathered data from two different stages, this study will solely focus on the estimates made at the latest stage at project initiation and look further into the details of their break-down.

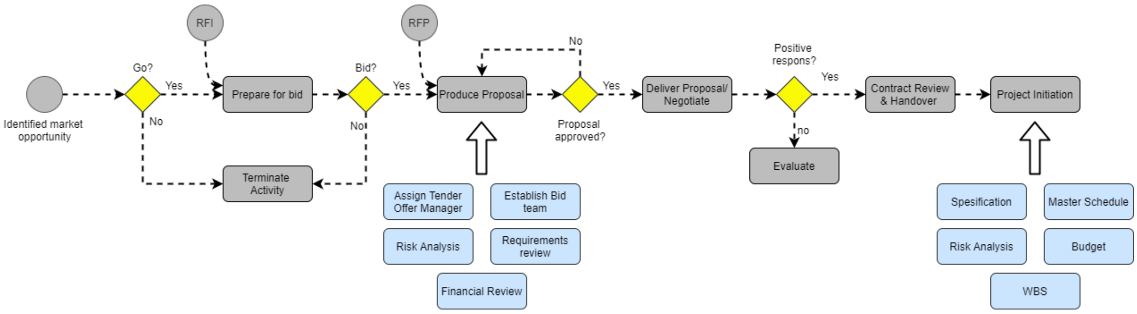

Figure 1: Simplified illustration of process up until project initiation

Source: Sandaa (2020)

The results showed the earliest estimates were typically made by a handful of individuals (or perhaps only one), while the estimates made at project initiation were a greater effort, including a more significant portion of the project organization. The projects were usually broken down into a set work-breakdown Structure (WBS) as illustrated in Figure 2, where the top layers were predefined while the most detailed, bottom layers varied from project to project.

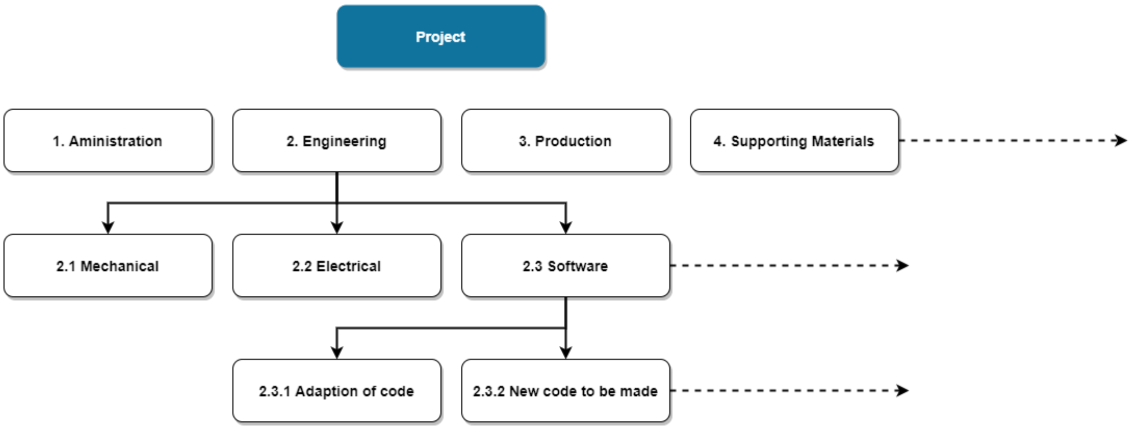

Figure 2: Conceptual illustration of structure on WBS.

Source: Sandaa (2020)

One or more representatives were typically responsible for the estimates of the workload in each department for the selected project, and detailed work package estimates were drawn up. TechStick had a range of different templates available, but which once were used, or to what extent, varied from project to project. The estimates relied heavily on the experience of individuals and their 'expert opinion' of how much time the individual tasks would require. Some preferred to use previously made estimates from similar projects as a baseline when estimating a new project, while some preferred to start from scratch with a blank template. Very few said that they actively looked at older estimates when making new ones, while many would often try to think back.

Following standard measurement for estimation accuracy as shown in Equation 1 below,

$$\text{Absolute Accuracy} = \frac{|\text{Actual Cost} - Estimate|}{Estimate} \tag{1}$$

The results showed that TechStick had an average absolute accuracy of 15% off actual in their early phase estimates and 19% for the estimates made at project initiation. The results were deemed statistically significant at P-values of 0.0017 and 0.003 respectively, when assuming mean equals zero. When looking closer at trends for over- and underestimation, the averages showed a slight tendency to underestimation of actual project cost in both of the estimates with an average of -9% for the early phase and -11% for the estimates at project initiation. These results were however found to not be statistically significant with P-values of respectively 0.13 and 0.08. Figure 3 shows the plotted accuracy of the analyzed projects for both the data points. The chosen confidence was set to 0.05. In addition to the mentioned empirical results, the actual overruns, the gross margins, and the relationship between the estimation accuracy and the project duration were investigated. Due to the nature of the selected method for data gathering, the data and results related to the estimates made at project initiation were deemed the most reliable.

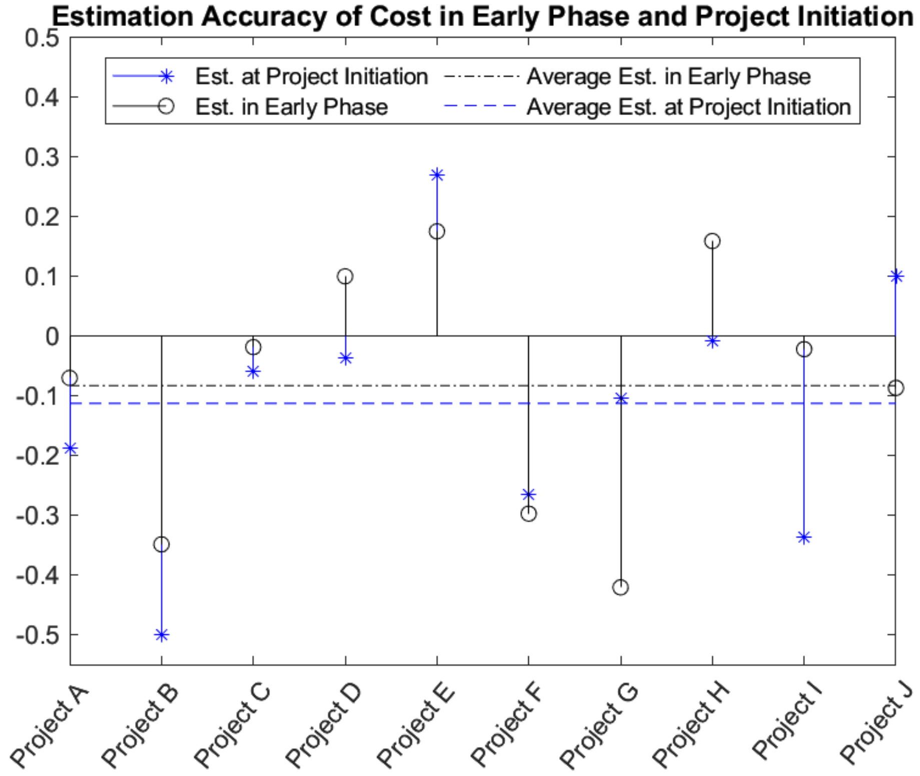

Figure 3: Plotted estimation accuracy for project initiation and early phase estimate

Source: Sandaa (2020)

2.2 Estimations and Scheduling

It is well known that the way you make your estimates and the chosen level of detail will affect the overall quality of the estimates. Being able to accurately predict project cost in advance of initiation is critical to profitability-/affordability analysis, budget allocations, and decision making in general. Despite the extensive amount of existing literature on project management, cost and effort scheduling remain an exacting task, in which a large number of projects fail to deliver precise results, as further demonstrated in Section 3.4. The reasons for this are rarely solely bad estimating in the early phases, as projects are inherently uncertain, and humans are yet to uncover the secrets of seeing into the future. Unpredicted events will almost certainly occur, however the number of events can often be dramatically reduced by doing proper due diligence in the project's early phase and hence also increasing the probable accuracy of the estimates (Kolltveit and Grønhaug 2004; Hussein and Klakegg 2014). Other common drivers for budget overruns are typically identified as late changes to specifications or quantity, poor understanding of the task resulting in rework, other outside-influencing factors, and just poor budgeting quality (Reichelt and Lyneis 1999; Jackson 2002). As this thesis takes a closer look at the accuracy and errors in the estimates, with an attempt to eliminate a large quantity of the before mentioned sources of error right off the bat, we'll in this chapter discuss the estimation literature closer.

As also discussed by P.Smirnoff and J.Hicks (2008) and P. E. D. Love et al. (2015), there is a need to further elaborate on utilized terminology within the research field before continuing. Although some authors emphasize a difference between the terms in their local research, a wast amount of research uses the terms cost overrun, cost growth, and sometimes also cost escalation synonymously. The usual differentiation is that a cost growth or escalation will be directly related to a direct and measurable change in specifications or quantity representing a change in scope, or an altered exchange- or interest rate. On the other hand, a cost overrun is harder to quantify, typically with an unknown origin, and simply adds to the final cost. This terminology distinction is however not consistent and may introduce misunderstandings as to what the respective authors are discussing. This thesis will predominantly be using the term overrun, with some mentions of cost growth when relevant, using the previously mentioned distinction. When referencing other researchers' work, it will to the best of ability, be attempted to keep the terminology consistent.

When discussing estimation techniques, there exists numerous ways this can be done. Research on the topic started gaining serious traction in the '50s as the first instance of the Critical Path Method was published by Kelly and Walker (1959) in an effort to structure and better general project planning. The method laid out a chained network of components with belonging attributes. The idea was to identify and all components of a project and sort them in their logical order along with a prediction on the required time to complete said activity. When all dependent and independent parallel trails of activities are identified and sorted, one can easily calculate the longest path and hence also the project duration. These principles have later been refined into multiple modern, well-known schedule estimation methods such as the popular Gantt charts and the successive planning by Lichtenberg (2000).

As the Critical Path Method was introduced as the first proper, deterministic method for project scheduling, the American department of Defence rapidly followed up with their own take on the Critical Path Method. Following similar principles, they introduced the idea of adding three-point estimates to each activity and fitting the points to a probability density function, consequently also introducing a stochastic technique for project scheduling (Malcom et al. 1959), often also referred to as a parametric method in the literature (Niazi et al. 2005). This was named the Program Evaluation and Review Technique, or more commonly known as PERT. The three-point estimates would represent the most pessimistic estimate, the most likely and the most optimistic for each activity, and were fitted to a beta distribution. This gave decision-makers a probability along with an expected project schedule and increased the robustness of the estimations. The beta distribution allowed for different weighting of the extreme's than of the most likely and gave estimators the opportunity to calibrate the formulas with different weightings based on experience. These properties also made this type of parametric estimate particularly useful for planning and managing a large portfolio of projects and controlling their risks. Attempts have also been made in recent years to utilize a different distribution than the beta distribution, as, e.g. by Hajdu and Bokor (2014), in order to find more suitable statistical distributions to the typically observed results.

2.2.1 Work-Breakdown Structure

The introduction of PERT and CPM was quickly also followed up by one of the more important supporting topics of this report, the Work-Breakdown Structure (WBS), as this is actively employed by TechStick and is one of their more essential tools for project planning. The concept was originally developed by the DoD in the '60s, building on their experiences with the Critical Path Method And PERT. The project structure was slowly adopted as the standard project planning tool across multiple branches and was later also adopted by the Project Management Institute (PMI) in the '80s as the go-to template for project delivery planning for civilian purposes as well.

The purpose of the Work-Breakdown Structure is to create a deliverable-oriented overview of the entire project, and every piece of work included within its scope. It creates a hierarchical vantage point to organize complex tasks into comprehensible sub-tasks in structures with logical relationships. Unlike the CPM, the WBS does not contain any timed activity-, effort- or cost dependencies and scheduling information, but is, however, the ruling project overview that those plans should be based on (Burek 2013).

The WBS breaks down a project into minor, structured elements by dividing every task into its smallest practical components. The principle is to divide the project into several larger deliverables that are within the project scope and then again break those parts down into minor sub-tasks to be completed. One may in theory, divide the tasks into as many levels of sub-tasks as needed, but it should remain within a reasonable number so that it is intuitively understandable for everyone part of the project. The largest defined deliverable represents the work packages of the given projects and will often be referred to as the top level of the WBS, or Level 1. Each successive subdivision of tasks below level 1 will follow the naming convention level 2, level 3, and so on until you reach the lowest defined tasks in each work packages. These tasks are called activities. Each work package is then assigned a number or reference code that can intuitively track and show relationships between levels and activities, which can be done in numerous different ways (Jung, A.M.ASCE, and Woo 2004). Examples of such are shown in Figure 7 and 2.

The use of a flexible numbering system that is adjustable from project to project has its advantages in its adaptability and reduced complexity by lowering the amount of activities needed in each structure to the bare essentials, making it easier to for project members to get an overview. A rigid system, on the other hand, has its advantages in its standardization, making it easier to utilize historical databases and custom systems to control cost and estimates, as well as making it easier for employees not directly involved in the project to understand the WBS (ibid.). The backside of such rigid WBS is however, that it might appear to be bloated and sometimes also contain empty Work packages, making the included numbering inconsistent.

There is no ”one correct way” to organize a WBS, but its creation should follow a set of guiding principles to ensure it's correctly constructed to serve its purpose. The first is as mentioned, that all parts of the project should be included in the WBS and that no work should be duplicated in any work package. The second rule is that the sum of every activity defined at a certain level should equal 100% of needed work to complete its parent deliverable. When deciding on how to determine the top-level Work Packages, there are multiple possible solutions available, depending on the project- type and organization, some ways might prove more optimal than others. As explained by Globerson (1994), it is shown to be favorable to have it organized in a way that allows a single organizational unit to be responsible for an activity, as opposed to a shared responsibility.

2.3 Industry Data

This section will be a review of results from studies found in the literature review that presents data relevant for comparison to our own data set and supporting literature that attempt to explain some of the observed results. The four main categories investigated are Infrastructure-, Off-Shore-, IT-, and Defence projects. An overview of the results from the literature review will be presented in 4.5 for the eager reader.

2.3.1 Infrastructure

Gao and Touran (2020) Analysed 81 US railroad projects from the last 40 years of varying size using publicly available information, found a correlation between project duration and cost overrun magnitude. The study looked into estimates at the concept phase and after risk assessment was made, and found a significant improvement in estimates after a risk assessment was done, as compared to the early phase. The study found an average overrun of 48.5% for major projects, costing more than 500 Million USD, and an average overrun of 24.8% for smaller projects. It was also identified a gap in accuracy between the sizes with regards to time as well. The study gives no root cause analysis, but points to a range of different authors also investigating US Rail projects, that determined that some of the leading causes for the cost overrun are predicted inflation, lengthy project execution delays, scope changes during implementation, and insufficient estimation.

Huo et al. (2018) Analysed 57 major infrastructure projects from Hong Kong, containing road, railway, tunnel, and bridges completed between 1985 and 2015. They found an average overrun of 58.08% for rail, 22.52% for road projects, and 35.58% for bridges and tunnels. This study did not find any correlation between project size and overrun, however the length of the projects appeared to be strongly, positively related to the percentage overrun. The average across all projects in the analysis was found to be 39.18%. The data in this study was collected from a range of different sources such as annual project accounts, internet research, news broadcasts, and archival research at the Legislative Council, Highways Department, and more, making it hard to determine what part of the early phase the estimates collected from or if there is any consistency in this.

Cantarelli et al. (2012) analyzed a set of 78 Dutch infrastructure projects similar to those in the Hong Kong study, including Road, Railway, bridges, and tunnels. The utilized data for this study was set to be the estimate available from the first year of construction compared to the final cost at the year of construction finish. They found that road- and rail projects in The Netherlands had an average overrun of 18.6% and 10.6% respectively. Fixed-link projects (bridges and tunnels) had the highest average overrun of 21.7% in this study, making the overall average about 16.5% overrun from estimates. The researchers found that both project size and duration had a correlation with the relative size of the cost overruns, however contrary to the typical results, these results showed that the smaller projects typically had a larger percentage in cost overrun. Compared to similar studies from other countries, the authors state that the observed overruns were small.

Welde et al. (2019) takes a deeper look at the accuracy of the estimates done by the Norwegian public quality assurance program and how they perform after projects were finished. The study includes a total of 85 large, public projects and analyses estimates for 70 of these projects, using final estimates made in the approval process in Parliament. The report thoroughly presents their selected loss- and evaluation functions for the analysis and the reasoning for it, evaluating both absolute accuracy and skewness for the P50 and P85 estimates used. The project categories are divided into road, railway, construction, defense, and others, with 65 of these being infrastructure and construction projects. Overall the study finds that the deviations are relatively low overall with an average of around 14% deviation for both the P50 and P85 estimates, and a slight tendency to underestimation in this category of projects. They further demonstrate that even with perfect estimation, it is not reasonable to assume zero deviation from estimates, but closer to 8-10% deviation.

Andrić et al. (2019) did a statistical analysis of 102 major infrastructure projects in Asia, including project types such as roads, railway, and energy sectors. The study focused especially on the differences between regions within Asia, as well as the differences between the sizes of projects. Looking specifically at the railway and road projects, the study found an average overrun of 21.11% and 10.47%, respectively. This study also identified a relationship between project size and overrun, where the larger projects would typically experience a larger percentage in overrun as well. The study utilized unison data collected from the same official source, containing all the project's completion reports, and is analyzed in a structured and analytical approach. The study found that cost overruns are equally common as underruns in Asia, however the magnitude of the average overruns was a lot larger than the underruns, with averages at 26.24% over and only 12.24% under. Similar to Huo et al. (2018) and Cantarelli et al. (2012) a correlation between duration and overrun was proven.

Flyvbjerg, Holm, and Buhl (2003) did a large worldwide statistical study covering a total of 258 infrastructure projects from 20 countries, using the estimate made primarily from when the decision to build was made against the final cost reported at project completion. The results showed an average of 45% cost overrun for railway projects, 34% for fixed-links, and 20% for roads, some of the highest numbers seen in studies of the kind. The researchers also found little difference between the different geographical regions in terms of the relative size of the overruns. A follow-up study by Flyvbjerg, holm, and Buhl (2004) on the same data set also showed a positive correlation between both total project cost and implementation time to the total overruns observed.

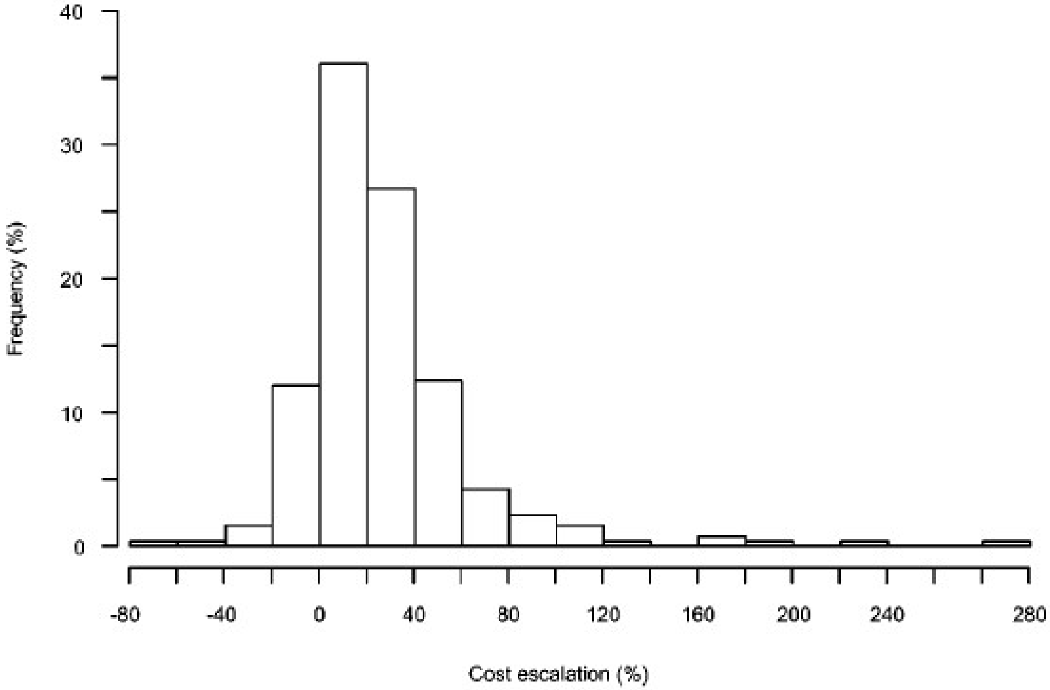

Figure 4: Distribution of cost overruns for 258 infrastructure projects.
Source: Flyvbjerg, Holm, and Buhl (2003)

The results showed that any given project had an 86% chance of experiencing an overrun, with the averages across all projects showing an overrun of about 27.6% from estimates. Figure 4 shows the distribution of the magnitudes of overrun reported. This study has in later years been criticized by P. E. Love and Ahiaga-Dagbui (2017) for selectively picking its data points to produce "news-worthy" results, along with misuse of statistical methods yielding overly confident results. This study was still chosen to be included, primarily due to Flyvbjerg, Ansar, et al. (2018) addressing most concerns in a convincing enough manner. This situation does however highlight the importance of proper methodology and its description.

Odeck (2004) did a similar study on road projects completed in Norway between 1992 and 1995. The study included a total of 620 projects of varying sizes and found an average overrun of 7.9% across all projects, but with a large standard deviation of 29.2%, indicating an inconsistency. The range of cost overrun in this report was from as low as -58.5% to 182.67%, but with longer projects showing a bigger tendency to cost overruns than the shorter projects. Similar to the research done by Cantarelli et al. (2012), this report also found that the smaller projects in terms of cost also typically had the higher cost overruns. Odeck (2017) also published a follow-up analysis as a summary of 48 other studies on infrastructure projects from across the world, summarizing the results from more than 21 000 projects. These results were similar to those observed previously, with an average of 34% overrun across the different project types, with road projects showing the lowest average overruns in the category with an average of 26.9% overrun and rail project significantly more at 36.3%, also supported by the findings of Huo et al. (2018). Here the fixed link category was integrated into the road category.

P. E. Love, Sing, et al. (2014) analyzed 58 Australian transportation infrastructure projects, including bridges, roads, tunnels, and subways, with 13.3% cost overrun on average across the projects in the portfolio. Fixed link projects including bridges, elevated highway sections, tunnel, and subways were found to have an average overrun of 15.75% while regular road projects were slightly lower at 12.49%. No relation was found between the project size in terms of cost, and the experienced cost overrun in projects.

2.3.2 Offshore

Oil and gas projects are typically recognized by their large, industrial-scale, and the amount of manufacturing needed, combined with their technical complexity. Cost overruns in major offshore projects seem to be a common occurrence in the news, possibly because of the magnitude of the values involved in the projects, but how common are cost overruns in the offshore industry? As for the causes, P. E. D. Love et al. (2015) analyzed a range of different hydrocarbon projects in search for root causes for cost overruns and highlighted factors such as project organization, complexity, degree of innovation, location, and technical challenges as major factors, among other, but pointing out that there is no unison scapegoat to be found.

The Norwegian Petroleum Directorate published (Oljedirektoratet 2020) a report covering the cost development of the 66 approved Offshore projects in Norway from 2007 to 2018. The report showed that 73% of the projects were completed within the general uncertainty margin of +-20%, 17% experienced cost overruns, and 10% were completed below the expected cost. Across all projects, the average overrun was about 8%, with only a handful of the projects accounting for most of the total overruns. Only 4 out of 66 projects had an overrun of more than 75%. The study also showed a clear improvement over time with regards to the cost development, with overruns being more frequent in the earlier projects. The study also reports that many of the projects experienced server overruns in the needed engineering hours for the development and planning phases without quantifying these numbers further.

Rui et al. (2017) did a review of 206 larger oil & gas projects, including fixed- and floating platforms, sub-sea systems, onshore wells, and production facilities. The projects were well distributed around the world with projects from Europa, North America, South America, Asia, Africa, and Oceania. The smallest represented region being South America, containing only 9% of the total included projects, and the largest being Europe with 27% of the included projects. The results showed that only about 25% of the total projects were overestimated, while the remaining projects were underestimated. 36% of the projects were completed within the +-10% range margin, and the average overrun out of all the projects was approximately 18.2%. The average overrun in these projects was found to be more severe than the underruns. The report also showed a correlation between the size of the projects and the size of the overrun, with the bigger projects having the biggest overrun in terms of percentages.

Merrow (2012) analyzed 318 oil- & gas megaprojects, defined as those budgeted to cost more than USD 1 billion, with a distributed sample of projects from all around the world, the most considerable portion being Europe. The study found that about 78% of all investigated projects experienced severe cost- and schedule overruns, with an average cost overrun of 25%, significantly more than what is reported from mega projects within other sectors from the same time. The study also found that European oil- & gas megaprojects were approximately twice as likely to be successful as similar projects from other regions, though still more likely to experience an overrun than a non-megaproject.

EY (2014) similarly investigated the performance of 365 different oil-and-gas megaprojects from around the world, looking into both cost and schedule accuracy. They found that

65% of all analyzed projects experienced severe cost overruns, and 78% of projects had schedule overruns. The average cost overrun in the analysis was found to be 53%, with Europe again having the smallest percentage of projects experiencing overruns. South America and The Middle East had the largest average overrun per project, with 102% and 68% respectively. Interestingly, the report also identified the lack of efficient and interconnected work breakdown structures as one of the main challenges identified, with those existing too often being made as a response to poor project performance instead of a preventive measure.

2.3.3 IT

A great effort has been put into estimating effort for new software development projects in the last decades (Kemerer 1987; Heemstra 1992; Qasim et al. 2017), and so the associated cost. Many studies claim that IT projects are the biggest sinner of all the project types when it comes to cost estimation accuracy. This section will be a review of studies investigating cost overrun in IT projects.

A report published in cooperation with McKinsey & Company and The BT Centre for Major Programme Management at the University of Oxford (Bloch, Blumberg, and Laartz 2012) reviewed 5,400 IT projects completed by McKinsey & Company with initial price tags exceeding $15 million. Tackling budgets, schedules, and predicted performance, the study revealed that over half the completed IT projects blow their budget massively. Across all their IT projects, the average cost overrun was found to be 45% over budget and with an average reduction in the delivered value of more than 50%. 17% of the analyzed projects were found to have an overrun of more than 200%. Flyvbjerg and Budzier (2011) dissected parts of the same dataset in another analysis but included IT projects of smaller size than $15 million. The study consisted of 1471 IT projects and revealed an average overrun of 27% across all the projects.

Another report published by Yang et al. (2008) looked further into the Chinese industry. It analyzed 112 IT projects completed within a range of different sectors, such as transportation, health, education, finance, and more. The projects had an average size of about 8000 work hours. Somewhat more conservative results were found here, with a little more than half of the projects reporting cost overruns, and 22% of the included projects experienced overruns of more than 20%. Similar findings as in previous studies were also reported, with a clear tendency to larger overruns in both schedule and cost for the larger projects. The study also found an average of 12% effort overrun in the estimated work hours, with the projects demanding more work hours showing a 16% accuracy. In contrast, the smaller projects reported an average of 7% accuracy.

Similarly, Moløkken-Østvold et al. (2004) analyzed 44 IT projects from Norway, looking into both schedule- and cost estimations done by a range of different companies and their methods used. The report showed that about 76% of the covered projects experienced schedule- or cost overruns and that the average cost overrun was about 41%. In addition, the results showed that the larger projects were more prone to underestimation than the smaller projects in the sample. Looking to the UK, Sauer and Cuthbertson (2003) did a survey covering some 1500 IT project managers and the projects they had been involved in. They found an average cost overrun of 18% in their data and that about 23% of the surveyed projects were finished over schedule. Also, this analysis showed a clear correlation between size and overrun, with the bigger projects typically experiencing both bigger and more frequent overruns.

2.3.4 Defence

Welde et al. (2019) included 13 defense projects in their study of Norwegian cost estimations in public acquisition projects. The study included a total of 85 projects, also including railroad, road projects, and buildings. Out of all the included projects, the defense acquisition projects had the lowest average overruns with over-estimations as the norm, as opposed to the remaining project categories. However, the defense projects also showed the lowest estimation accuracy in the sample i.e. they had the largest variance.

Bolten et al. (2008) did a study on cost growth on 35 large, mature US defense procurement programs in an attempt to locate the root causes of the observed increase in cost. The included projects had their reasons for cost changes categorized into four main categories, Errors, Decisions, Financial and Miscellaneous, with the two most prominent being Errors and Decisions. Faults during estimation in the early phase would be categorized as an error, while e.g. change of delivery size or project scope would be classified under the Decisions-category. The possible changes in delivery size is a factor somewhat separating the work Bolten et al. (ibid.) did from this study and many others, making the comparison a little harder, as *this* study actively avoids the projects with large scope changes to focus on the actual estimation accuracy. The reason for this being that the study showed a trend of unit/scope reduction to counter overruns as a mean to control cost, possibly indicating more severe estimation errors than what the results reflect, as also discussed by Bolten et al. (2008). The results from the study concluded that the programs had a total average cost growth of 73.5%, for helicopters and aircraft, 43.9% for missile programs, and 27.5% for the electronics programs, combined about 38.8% overrun overall. Cost estimation errors and scope changes make up the majority of the cost growth.

Another similar analysis was done by Younossi et al. (2007), building on a previous study by Arena et al. (2006), also looking into US DoD programs. The study found that the vast majority of programs experienced cost overruns, ranging from 0.8 to 2.3, and an average of 1.46, meaning a 46% overrun on average. Is was identified a clear link between the duration of the program and the overrun observed, and the majority of growth appeared to originate from the development phases of the programs in this study as well. One of the most relevant findings from this study is that it was identified a statistically significant difference between the program types, with the ones categorized as Electronics Procurement's demonstrating a significantly lower cost overrun on average, while the more complex programs categorized as aircraft and helicopters shows the largest average cost growth of the programs, with almost 50% more cost growth than the electronics programs. Of the 13 included reasons for cost growth investigated in the study, estimation errors along with changes in quantity and requirements changes were found to be one of the most prominent contributors across all the different program types. These numbers are also supported by Swank et al. (2000), showing about 40% cost overrun on average in US DoD programs competed between 1980 and 1996, using data from the same database.

A report written for US Congress by Schwartz (2014) in advance of treating a new acquisition reform also reported that US defense projects have experienced a median of 32% cost growth since 1993, with more than a billion USD spent on projects that were ultimately terminated due to cost overruns between 1996 and 2010.

Numbers reported by the UK Ministry of defense in NAO (2015) show similar trends as observed in the US, however with slightly more conservative numbers. When analyzing the 15 largest military investment programs ended in 2010-11, the auditor finds an average growth in cost of 11.4% across all the projects at their completion. Similar to what was reported by the US studies, the report also noted that many of the projects had undergone several cost-reducing changes during their lifetime, such as a reduction in units, which otherwise would have lead to more severe overruns.

One of the main factors discussed and identified for high cost growth and overruns in the defense sector is the phenomenon known as the defense specific index (DSI). The DSI is the index representing the approximate increase in cost on defense equipment as measured by annual cost increase per unit output, and is typically found to be 2-5% higher than the average consumer price index (CPI). There exists a large number of reasons for this phenomenon that total up to the accumulated totals, but one of the largest factors is technical complexity. Increasing technical complexity on systems as new generations are introduced, demands deeper technical understanding, better materials, and more specialized contractors to keep up with the development of other countries. Hove and Lillekvelland (2016) did a report on the root causes of the DSI in the Norwegian Armed Forces as well as an estimation of how large it is. They found that the DSI is steadily increasing each year and that the DSI related to advanced weapon systems and other technologies was currently about 6%, while the DSI related to activities unrelated to such projects, such as personnel, was about 2% above CPI. They concluded that if long-term budgets assumed a DSI of a conservative 3%, we could expect an 80% cost increase per unit in 20 years. Hartley (2016) also did a comprehensive analysis on the UK DSI, and found comparable results to those presented by Hove and Lillekvelland (2016). Knowing that the DSI is significantly higher than the CPI and adjusting for this is found to be important to successfully estimate the cost of defence projects. Especially so when using historical data as the baseline or for long-term planning.

3 Method

This chapter aims to give a detailed description of how the data and following results in this study were obtained, as well as give an understanding of the writing process of this report. It will contain a brief section describing how previous research was identified and its criterias for inclusion in the study for easy traceability. The main body of Section 3 will focus on the process of data gathering and project analysis. Why/why not projects were included in the study and how the gathered data was analyzed once obtained, as well as a discussion regarding the implications of the different available methods to utilize and how it could affect the results. The structure, analysis method, and defined research questions in this report follow procedures defined by J. W. Creswell and J. D. Creswell (2018).

This report presents a statistical analysis of empirical data gathered through internal project documentation from TechStick, and the results are thus mainly obtained using quantitative methods. The duration of the study was approximately 20 weeks starting January 2021, with the first half of the research being background study and some data gathering and the second half mainly analysis. The report builds heavily on the experience obtained by the author in the prestudy(Sandaa 2020), which also included comprehensive qualitative research on the projects in question, as further described in Section 2.1.

3.1 Literature Review

One of the secondary data sources of this research is the complimentary studies done by other researchers within the field of study. An array of additional research was necessary to include, to properly cover the relevant grounds in this study. The history of estimation and scheduling and the most notable methods used traditionally are important to gain an understanding of the current state within the industries and the pros and cons with their chosen way, as well as interpreting the results. The way these estimations are typically measured to gain an overview of the current state of affair is also a topic covered by the literature review, and lastly, a number of similar studies needed to be found to gain an overview of how well other industries and similar companies typically perform and what could be considered normal. The two search engines used during the study were NTNU Open and Google Scholar, and the most important criteria for inclusion was that the studies were peer-reviewed. More detailed search terms and selection rules are further described in the following sections for the different research areas. Table 1 presents an overview of most search terms used in the research for the different areas, however a large portion of included studies were identified through backwards snowballing in the references found in other major studies, a method described in further detail by Wohlin (2014). This method was instead of the classically structured literature review, primarily for its highlighted ease of use for novice researchers, see e.g. Jalali and Wohlin (2012). Initial disqualification of papers was based on title, conclusion, abstract, and in some cases a rough overview of the introduction where the abstract was ambiguous. The remaining papers were read in further detail and stored for possible inclusion at the relevant time. The different chapters were typically written simultaneously as the relevant research was done to ensure the content was accurately cited.

For the historical development of the different estimation methods used and the performance measurement methods, this study relies some on the research done for the prestudy (Sandaa 2020), but also includes several new literature to further broaden the perspective. This part of the research aimed to gain an overall understanding of the existing methods and their developments. It was not a goal in itself to include all relevant research, but simply enough to get the required overview and understanding, as well as any new developments done in recent years. The literature review done to identify similar studies sought out a much broader range of studies and had a goal of identifying as many as possible studies in the selected range of industries (see Chapter 3.4), that used the same or a similar standard for calculations as this study. This will be further elaborated in Section 3.4.

Methods and developments	Performance evaluation	Similar Studies
Cost Estimation Schedule Estimation Project Cost Scheduling Software Cost Estimation Work Breakdown Structure	Estimation Accuracy Cost Estimation Performance Cost Overrun Project Success	Defence Overrun Project Cost Overrun Offshore Project Overrun Infrastructure Cost overrun Software/IT Cost overrun

Table 1: Overview of keywords used for literature review

The vast majority of literature was quickly discarded in the selection process, with less than 10% of the reviewed publications being included in the study. Figure ?? shows approximately how many studies were discarded at what stages through the review.

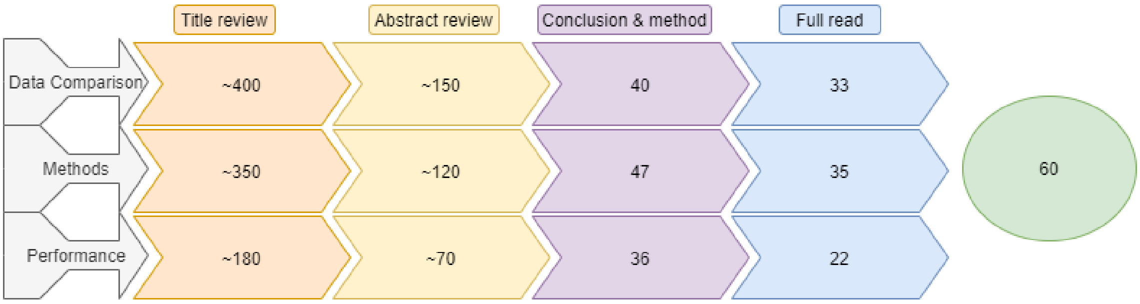

Figure 5: Literature review process

3.2 Anonymity and Data Security

The company in question and all related data have been anonymized to protect their privacy and sensitive data. The company will hereafter, and as previously, be referenced as TechStick. For the purpose of this research, it was granted exclusive access to employees, internal guidelines, and documentation archives otherwise not available to the public. The pseudonym chosen for the company was selected by combining the names of random objects observed in the room by the author. A multiple of combinations was tried before a decision was made, somewhere between 30-40, as it was actively attempted to avoid both previously and existing corporations with similar names and names that could easily be associated with the actual company in question. To sum up, it exists no affiliation between the findings of this report and any company, current or previous, with a similar name.

All data presented in this report has been anonymized by changing its associated names and scrambling the project relations. The numbers analyzed are company internal information; therefore, additional hour costs are not added, to prevent the data from being recognized by any individual customers, clients, or suppliers. No data was ever exported from the company's internal network during analysis or processing before the final results were exported to this report. For that reason are all calculations and graphics made using software available on internal servers, which for the most part will be Microsoft Excel. This was seen as necessary given the large amount of data, in order to avoid exporting such amounts of proprietary information during the research.

No numbers are rounded off or redacted in any way, as opposed to how it was done in the prestudy, Sandaa (2020). This was now found necessary due to the relative size of a lot of the numbers in question, and would introduce a source of error larger than acceptable. The specific naming of the different work packages has been altered to not bear such a distinctive naming convention, but without changing their meaning or order.

3.3 Project Analysis

What projects to include in the study and how they are selected is a topic that demands quite some attention. There are a number of reasons for why one can not simply include every single project in the portfolio, and for research purposes, it will be thoroughly explained what selection criterias were selected and why. In the pre-study, a part of the focus was set on the sales perspective of the estimations and how they performed compared to the estimations made at project initiation. This perspective gave certain constraints to the selection, such as the project had to be initiated by a tender offer to an open bidding amongst relevant participants and have a paying customer. In addition to this, the project had to be a delivery project, as opposed to for example a development project. In this study, the focus will solely be on the more thorough estimates made by the project organization at project initiation, specifically on the estimated hours. Although the perspective of the Tender Offer Managers gave value to the research, it was also uncovered that they typically lacked the common structured approach to the estimation, making it very had to do any sort of methodological comparison to those results. It was also in the author's opinion that the different estimates for the work breakdown structure at project initiation was the area in demand of the most research going forward.

Where the prestudy's data was largely based on the project results reported in the respective finalization reports, this study is getting the data directly from the financial reporting done by the project controllers during the course of the projects. This gives a much higher resolution to the data and allows to further study the details of the project development on a month-to-month basis. This further renders the need for a thorough finalization report to exist obsolete but increases the need for additional available information. In total, these changes disqualified a few of the previously included projects and released a few more projects available for analysis. The total sample size was reduced by too seven projects, mainly due to incomplete data. This report also depends on the knowledge gained from the interviews done in the prestudy, as this will be a strictly empirical analysis. Because of the exclusion of the interviews from this part of the study, the need for high-quality project documentation for any new projects is critical, for them to be included in the portfolio. Given the mentioned changes, the final selection criterias for the projects to be included in the portfolio are as follows.

- The project needs to be concluded with almost no remaining expenses.

  Given that some projects might experience expenses years after they are closed, related to warranty, spares, and such, it was decided that an unknown amount of remaining expenses totaling less than 2% in projects would be accepted. Meaning that some of the most recent projects are missing 0-2% of their expected expenses.

- No major changes to the scope of the projects were made after the project was initiated and estimated.

  In order to focus the study and increase the quality of the results, this is considered important to ensure that a larger portion of deviation found was caused by estimation error, as opposed to changes made to the actual scope of the project during its lifetime.

- A proper record of of the different estimated work packages most be obtainable form the projects.

  In order to actually evaluate the early phase estimates, an intact and unadjusted record of WBS needed to be available for the project. As the financial department keeps a meticulous record of all expenses associated with any given project, the limiting factor was often found to be the documentation done by the project team.

Compared to the evaluation criterias selected for the prestudy, it will at first glance appear to be a lower bar for inclusion in this study as there is no longer a demand that the project is a delivery project to a paying customer, and details regarding contract value are no longer needed. However, the detailed approach to this analysis has actually dramatically increased the requirements for available, good documentation, leading to an even stricter selection process. The strict selection process was seen as necessary to ensure the quality of the results. These changes have however also lead to the inclusion of internal product development projects as well, as long as they meet the other demands.

For the purpose of the research, it was given the freedom to chose suitable projects from the archive. As the documentation for each project varies greatly, the identification and elimination process consisted of a large amount of manual labor, following an approximate process as illustrated in Figure 6. The easiest of eliminations could be done within about 30 minutes of reading, while the average project would be eliminated after about two hours. Some projects were also eliminated after multiple days of work.

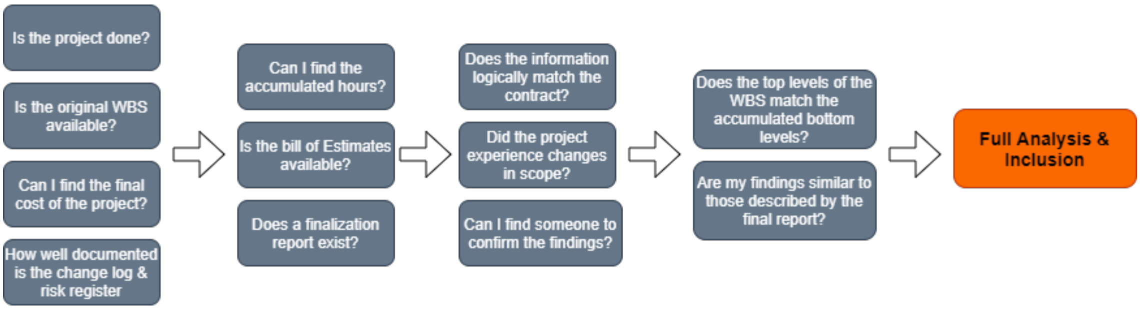

Figure 6: Project selection process

Figure 6 illustrates the approximate process followed to ensure that proper documentation was available for the projects. All the boxes most not necessarily be checked for a project to be investigated further, as information would often complement each other in different resources and could sometimes be recreated from another source. For example, could a project be partially analyzed even though the second layer of the WBS did not match its accumulated bottom layer. This would only mean that we know that some information is missing from the estimates, as opposed to it not being estimated in the first place, and that part of the data will only contribute to top-level analyses, but not conclusions regarding the detailed analysis. The projects still included from the prestudy had the advantage of information from the interviews during the selection process.

Despite the fact that it is actively avoided projects that have increased the scope of the project during its lifetime, it is hard to guarantee that this is the fact. Project Managers have limited memory regarding the finest details of their projects, especially those that are many years old. In addition to this, is it hard to guaranty that details about this are documented in every project. To mitigate the risk of such errors in the results, signs of this have been searched for in documented progress meetings, contracts, risk registers, finalization reports, and more. Some scope creep is however expected to have occurred, but is considered almost impossible to detect. This is for the same reason, something that should be accounted for in the estimates of every project and will be considered such in this report.

3.3.1 Data Collection

The data used in this report comes from a number of different internal sources, but the bulk of it is gathered from the project archive. Most of the project data is pieced together from multiple different sources to a final data set and cross-examined against at least one secondary source to confirm the magnitude of the numbers found. Such a secondary source has typically been a late finical project report or a finalization report containing information regarding the hours spent during the project.

Information about the Work-Breakdown Structure has two primary sources and contains two different relevant pieces of information gathered. The first piece of data is the original list of activities in the WBS ordered at the start of the project. This is a list with numbers and belonging activity names on which they expect to be spending time. This list was typically found in the project documentation and would contain all the bottom-level activities. The names of the bottom-level activities are mostly project-specific and would therefore vary, but their number is a good indication of what type of work it is. The names of the bottom-level activities will not be disclosed in this report, as they are often somewhat revealing of the specific project type. The top-level and second level work package is mostly consistent across projects, but have changed some over time, meaning that their names will, in some cases, be important to get an accurate comparison across projects, not just the numbers. The projects analyzed contain somewhere between 30 and 100 bottom-level activities, usually at level 3, but sometimes also containing a fourth level as well in some work packages.

Figure 7 illustrates the typical top level work package defined for the projects, with some examples of second- and third-level activities as well. Because of the typically large amount of activities included in a project, they are not all illustrated. Note that Work package three is rarely used and often integrated into WP2. WP5 and WP9 are also mostly empty for the selected projects in the portfolio.

The first WP contains overall administrating activities such as the project manager, the financial controller, configuration management, material planning, quality assurance, and more. WP2 mostly contains all engineering activities, such as the ones illustrated, testing, safety engineering, and more. WP4 is the estimated work for customer support, technical support, and more activities directly connected to the customer. WP6 is all activities related to production and preparation for this. Note that this does not include actual production hours, as this is not logged directly to the project team but rather a direct cost. WP7, Logistics Support, is preparation of user manual, training of customer, maintenance manuals, and more. WP8 is work and cost estimated for creating and handling customer warranties.

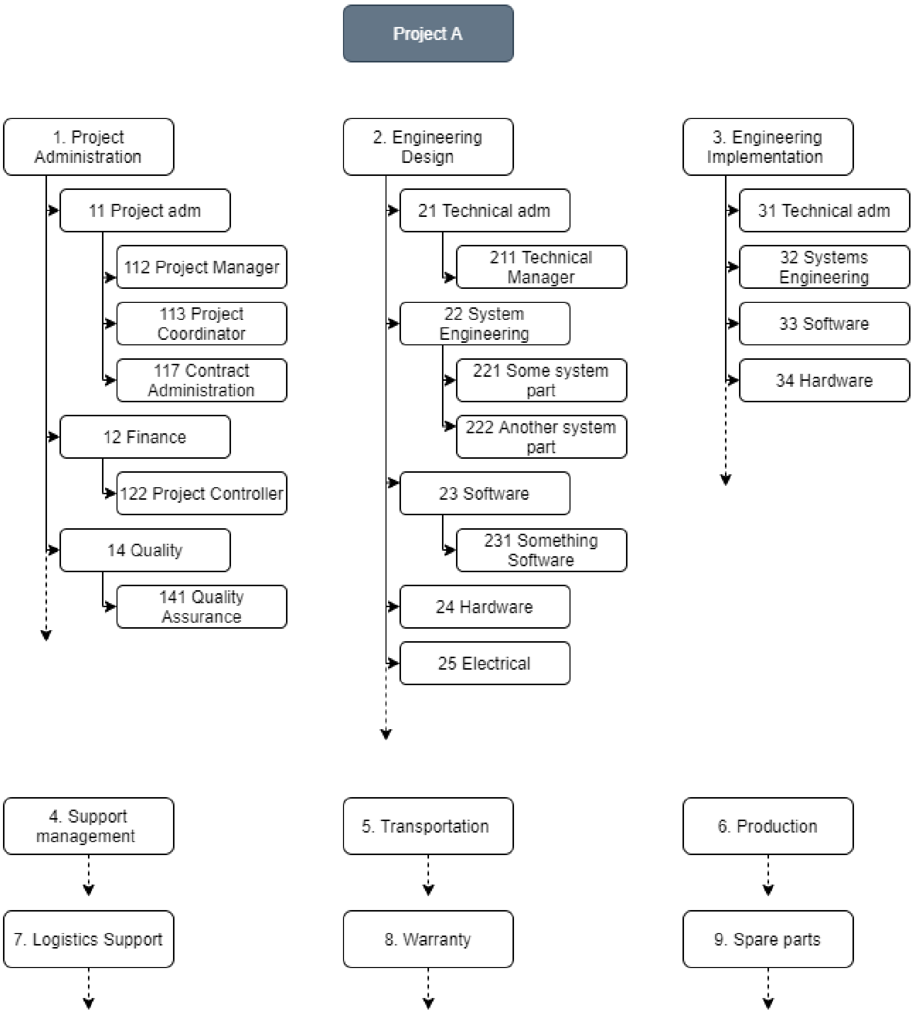

Figure 7: Overview of TechStick's WBS

The second piece of information is the actual estimates made by the different departments belonging to the different activities. The bottom-level activity estimates were the most challenging pieces of information to obtain for this analysis. The estimates were found in various locations across the archive with inconsistent naming conventions, more often than not distributed across folders belonging to different departments. The totals for the work-package estimates were however kept by the financial department and were provided upon request. The provided top-level estimates allowed for a cross-reference that would tell if all bottom-level estimates were located and accounted for. For most included projects, all the bottom-level activity estimates were located and matching the provided top-level totals. In the projects where some of the bottom level estimates are unaccounted for, the projects were not included in the detailed activity analysis for that particular WP.

The last piece of information is the actual hours spent on the different projects. This information was provided by the financial department upon request after a project was decided to be included in the analysis. This information contains an accurate record of every logged hour to every activity. The provided information in this record was most importantly activity number, activity name, department, hours used, month logged, year logged, and cost per hour for the different entries. Each included project is made up of between 15 000 and 40 000 relevant data points that are part of the analysis. Each data set took 2-5 days worth of work to locate all relevant information and combine it into useful data, depending on how scattered the information was to begin with. Another day or two was spent on the actual analysis of the data for each project after it was combined, and approximately 10 more days for the final, overall analysis.

3.3.2 Data Analysis

To obtain answers to the questions asked in this report, the projects are analyzed using a number of different data points and formulas. As the main research question asked in this part of the research is the accuracy of the estimates, the two most interesting metrics are the value of the estimated activity and the value of the accumulated time spent on that activity. Note again that this report is primarily analyzing the hours estimated and spent on projects, not the fixed cost elements or the total costs. For data regarding the average accuracy of the overall cost estimates, it is relied on the results found in Sandaa (2020).

The estimates are evaluated at their lowest level in the WBS for the recurring individual elements in the projects, and the accumulated totals in the second level. Lastly, the top-level accuracy across all the projects will be measured. It would, for example, make little sense to measure the average accuracy of activity 233 across multiple projects, as what this is could vary greatly. It is, however possible to know that every activity 23X is software engineering based on their number and naming conventions, leaving it possible to examine the accuracy of the software work estimations across the projects. However, there are also a few consistent bottom-level activities such as hours for the project manager, technical manager, project controller, and similar that remain comparable across all projects and can therefore be evaluated individually.

Each activity and work package is analyzed using a few relatively simple metrics that together give a good overview of the estimation accuracy. Each individual project will be analyzed using the same methods and formulas, and simple statistical tests will be deployed on both the individual project activities and overall results. Following the same calculation conventions for estimations accuracy as in Sandaa (2020) and in other recognized literature such as Flyvbjerg, Holm, and Buhl (2003) and Welde et al. (2019), we define any given accuracy as

$$\text{Relative Accuracy} = (-1 * \frac{\text{Actual - Estimate}}{\text{Estimate}}) * 100, \tag{2}$$

with the negative sign being an accuracy identifier used in the context of accuracy. The practical implication of this convention is that the accuracies are measured as percentages of the original estimate. Anything with a perfect correlation between the estimate and the actual cost or hours will come out as 0%, while a post that used twice that of the estimate will be -100%, while half gives 50%. This formula is excellent for evaluating tendencies to over - or underestimations, and the closer to zero it measures, the better the accuracy is. A weakness with this convention is however that the averages will not actually tell how accurate the estimations are, only if the tendency is an under- or overestimation. One could easily imagine a situation where all the estimated activities in a project came in with actual values far from the estimates, simply in opposite directions. The averages would, in this case, be pretty good with numbers around 0%, while the actual estimates in reality would be hugely inaccurate. Note that this formula does not allow for a relative accuracy measurement larger than 100%, as this would require a negative amount of actual cost. This leaves a theoretical larger available range of values for the underestimations that on need to be aware of using this convention.

For the purpose of the research, it will not be speculated on what a preferable result is in terms of over- or underestimation, meaning that any equal deviation from 0% will be considered equally bad. Hence no mathematical weighting will be added to the results during evaluation. For this reason, it will be utilized absolute values when evaluating the quality of the estimates. Using a slightly modified version of formula 2, as

$$\text{Absolute Accuracy} = \frac{|\text{Actual} - \text{Estimate}|}{\text{Estimate}} * 100 \tag{3}$$

we move every deviation from the actual to a positive number. Now the averages will give a better picture of how accurate the estimations actually are in general. In other words, how many hours away from the estimates an activity usually is. But simply looking into the percentages is not sufficient to get a proper situational understanding. Imagining an activity estimated to take 2 hours to complete, but ended up actually taking 6 hours would give the same results as an activity estimated to 2000 hours, but ending up needing 6000 hours. It is in the authors' opinion that the later example is a more severe underestimation and would naturally have more severe consequences for the project as a whole. For this reason, it comes becomes apparent that the actual hours need to be analyzed both as itself and as a part of the total project value to determine its severity.

The last data that is evaluated are the activities not estimated at the beginning of the project, but turned out to be needed. As these activities were initially unestimated, it is not possible to include them in the analysis as an underestimated percentage. It is, however possible to analyze what type of activities are typically neglected during the early phase estimation and how severe they usually are based on their numerical quantity. As mentioned earlier, it is possible to know if an appearing activity was originally unaccounted for in the estimates if the totals of the bottom level activity estimates matches the top-level work package estimates, as these usually originate from different sources of documentation.

All statistical and mathematical formulas used in this report are gathered from Walpole et al. (2017). Every hypothesis test calculated for a different part of the data will assume the mean or expected value to be zero or perfectly estimated unless otherwise stated in the results. Due to the low sample size of projects, the t-statistic is the selected significance test.

3.4 Industry Data Comparison

Gathering data from similar research poses multiple challenges one needs to be aware of before any kind of comparison should be attempted. This section will briefly discuss some selection criterias and difficulties faced in the search for similar research. The overall research philosophy could be argued to be quite liberal, with a wide range of different studies included. The main object was to gather a decent amount of data from each of the selected project categories to get an overview of the individual industries' representative performance.

The four main categories that were research are infrastructure projects, Offshore, IT-software projects, and Defence projects. Infrastructure was selected for a couple of reasons. Firstly its availability of data and research. It is arguably the project category with the most readily available similar research from a vast amount of different countries and researchers. Secondly, it's a category of projects with decades worth of management and estimation experience and well-documented procedures. An extensive array of the projects undertaken in this category are categorized as low-complexity and predictable, yet many of them are still highly complicated, leaving it a good benchmark for what kind of results to expect from an established and well-versed industry.

Offshore projects are typically recognized as complicated and comprehensive projects that, comparable to TechStick, utilizes a wide range of engineering disciplines but are also particularly heavy on the mechanical side of things. Perhaps will these results bear some similarity and comparable value with the results observed in parts of the analyzed data in this study.

IT- Software projects could, to some extent, be considered to the opposite of infrastructure projects when it comes to experience and established procedures. There is still yearly published a large number of studies suggesting new and innovative ways of estimating the cost and work effort needed to complete different types of projects. In spite of this, it exists a vast amount of research and comparable data within the field, and it should be possible to get an overview of a rough industry standard of how accurate a typical software project estimate is.

The last selected project category is the defense sector. A typical defense project intertwines a large number of different engineering disciplines and combines them in a manner that demands high robustness and reliability. The typical combination of hardware design & production, software integration, electrical work, and safety engineering needed makes these types of projects the most comparable, perhaps along with offshore projects, to those of TechStick. As defense is a vast industry, it should be possible to find similar studies for comparison.

There are many sources of error when looking for comparable data in other studies, but most of them will arguably be related to the selected methods for each study. There are plenty of different ways such studies are angled, and what perspective they presents could have a large impact on the observed results. The first one to mention is that most reports will look into the total project cost, including fixed costs such as material cost, transportation fees, salaries, and more. Even specific- or larger market movements will affect the bottom line of most projects' total cost. As this analysis is a pure work effort estimation analysis, it serves as a good insight into own estimation performance, but a large set of other sources of error need to be accounted for when looking at total cost vs estimate for projects. It is no sure way of knowing what way this might skew results, as fixed costs might introduce unforeseen extra costs or might present opportunities to adjust total cost to compensate for excessive use of work hours.

Another usual take on the problem is the public perspective, as seen through the eyes of the governments and their financial project reporting. This perspective will often come with an early project estimate as their measurement benchmark. Such estimates might have been made by someone else than the executing company, with a different initial scope or plan, and will in addition contain all sources of error mention above.

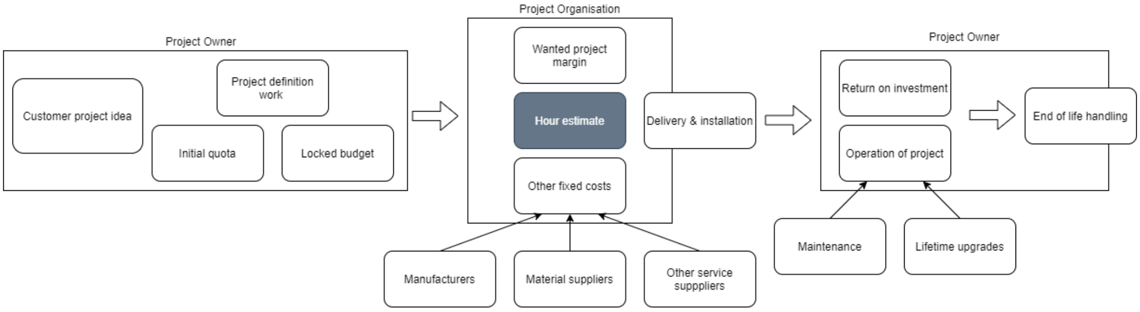

Figure 8: Possible analysis perspectives

Figure 8 shows an illustration of different perspectives possible to select when performing a project overrun analysis. Each box in the figure represents a potential cost, estimate, or both, and hence also represents different factors one might include depending on how grand of an overview one wishes to get with the analysis. The scope of this report is marked with blue. There are of course numerous other factors one might include in addition, but it serves as a good representation of the different perspectives available. Each of these elements will to a large degree have an unknown uncertainty- and estimation distribution associated with it, and hence also contribute with an unknown amount to the final results, depending on how many details are included in each report. A quick example could be the potential difference between the estimation accuracy when looking at the performance of the project organization versus that of the entire project through its lifetime and how many factors that might contribute to differentiating the two results.

These influencing factors will be further discussed in Section 5.

4 Results

This section will consist of two main parts. The first, and largest one, is a presentation and walk-through of the structure and findings from the data set. Here important metrics and analysis specifics will be highlighted along the way, as well as explanations for observed behavior that is important for the reader to know. The second part will be a presentation of the findings from the literature review conducted, with a structured overview of the results from the different categories researched.

4.1 Project Data

This data set consists of 7 projects completed by TechStick that passed the selection criterias for inclusion in the analysis. Each project is given a letter as its identifier, meaning that the portfolio consists of Project A - G, in which the order was randomly set. There is no correspondence between the project names in the prestudy Sandaa (2020) and this report, however six of the projects in this report are overlapping. The reason for this is to make it harder for readers to associate both project cost and hour estimate for any given project in case a reader should by chance have inside knowledge from the company and recognize the data.

Five of the included projects have all their activity estimates available and could hence be called complete data set. Project C and Project F are however missing some of their bottom level estimates, but have in all cases the overall work-package estimate available. Out of the two, Project C is missing a substantial amount of its estimates, while project F is partially complete. The consequences of this will become apparent in time, but note that these projects will still be included in tables where they have an insufficient amount of data available to contribute. This will mostly be marked by a "n/a" to give the reader a better understanding of available data, but they will however not be included in that specific calculation in any way when this is the case. Some other use-cases of "n/a" will also be found but will be explained at the first relevant occurrence.

The following sections will thoroughly present the data set, with some parts seemingly too thorough. This is by choice as the complete data set cannot be made public for privacy reasons, giving the reader a better chance to cross-check numbers or re-calculate the results should it be wanted. Only a handful of very generic bottom-level activity estimates will be disclosed by name and analyzed individually, as most names are far too revealing. Instead, a batch- and category analysis of the bottom-level activities will be done.

4.2 Overall Findings

Table 2 presents an overview of the accumulated hours spent on each project as well as their initial estimate and the estimation accuracy of the individual projects. Most of the projects in the portfolio are in the same order of magnitude in terms of size, with Project A deviating some at is was initially estimated to be twice as large as the next in the data set. In terms of their estimation accuracy, the projects have a wide range of performances, with Project B being noteworthy worse than the rest with an overrun of 162.48%, also accumulating to the largest project in actual hours as well, hence it is expected to see this project affect the means and standard deviations in some of the following calculations quite a bit.

Out of the seven projects included, only two ended up overestimating their needed hours. The overestimations are also significantly smaller than the observed underestimations, leaving the average relative estimation accuracy at -46.21% on the portfolio, however with a fairly large standard deviation. The absolute accuracy is found to be 50.18%, with the standard deviation unexpectedly slightly lower at 54.50%.

Project Overview			
Projects	Estimate	Actual Hours	Estimation Accuracy
Project A	20 821	18 641.25	10.17%
Project B	7 823	20 534.15	-162.48%
Project C	4 165	4 026.80	3.41%
Project D	4 051	5 547.10	-36.93%
Project E	9 155.5	13 142.35	-43.55%
Project F	7 091	8 695.9	-22.63%
Project G	4 173	7 167.4	-71.76%
Average			-46.21%
Standard Deviation			58.45%

Table 2: Overview of project hours and estimation accuracy

This is a perfect opportunity to point out a phenomenon that is important to have in mind for the remaining results. Notice how the total hours in estimates amount to 57 283 hours and the actual is 77 755 hours, leaving only a -37.74% overrun on the entire portfolio, less than the previously mentioned -46.21%. This is due to the fact that all projects are not of equal size and will therefore naturally bear a different weight to the final overall results. Both of these metrics are inherently correct, but they carry a different meaning depending on what data you are looking for, a phenomenon also seen clearly in Bolten et al. (2008). From a portfolio management perspective, it is perfectly valid to evaluate the overall numbers, as this does tell how accurately you have overall performed. This will, however not give any details regarding the performance of the individual projects, as many underlying numbers will in many situations even each other out. This can to some degree, be mitigated by also investigating the different standard deviations, but will still only provide a partial picture of the situation.

This phenomenon will be observed multiple times in this study as we propagate the down through the levels of estimation. We will primarily be investigating the individual accuracies throughout this study and will therefore not weigh any averages along the way, but instead let each project's percentage count equally much, as it is in the authors' opinion that these results are the most representable for the situation and provide the most accurate answers to the research question. These numbers will in most cases also be accompanied by a sum of involved hours as well, to provide the reader a good situational understanding of both individual and overall performance throughout.

4.3 Work-Package Analysis

Each of the projects presented above consists of estimates distributed across multiple different work packages of work by different categories. The following sections will break down the content of the different work packages and analyze their individual performance. Each work package belonging to each project will in term consist of a number of different activities predefined with belonging estimates, which in term sum up to that work package's overall estimate. This will be the metric displayed to the far left. The sum of actual hours consists of two different types of data. The first piece is the used hours associated with one of the activities predefined in the original estimates. The second piece is hours accounted to an activity that is by numbering associated to the work package in question but was not present in the original estimates. A fictional example to illustrate could be that Project A's WBS 1 originally consisted of activities 111, 112 & 113 with their estimates, but during the course of the project, they discovered that they forgot to account for some other work and therefore started logging this to a new activity, 114. The sum of these makes up the actually spent hours on that work package.

Named A. Accuracy and R.Accuracy in the following tables are the absolute and relative accuracies calculated from the metrics mentioned above. As opposed to overall project averages, these are calculated from the the total sums, and not the averages for each activity. The reason for this is simply that the originally unestimated bottom-level activities do not have an original estimate, hence making it impossible to calculate each individual accuracy. This problem will be further investigated in section 4.4.

Some insight into this is presented in the following tables as well. The total Unestimated Activity Hours (UAH), or in other words the sum of hours accounted to an activity not in the original budget, and the amount of unestimated activity hours as a percentage of the total time spent on that work package is presented to the far right. A summary of results from the following sections can be found in Section 4.3.7.

4.3.1 Work Package 1 - Project Administration

Work package 1 is one of the heaviest WP's in terms of actual hours in every included project, giving it a lot of influence on the overall project performance. Five of the projects underestimated the amount of time needed for these various administrational activities, with an average overrun of 24.25% on the hours. This is significantly less than the projects overall, indicating that these activities are pulling the average down. The absolute accuracy is however 44.83%, so this work package hits almost 50% off-target regardless of the direction from estimates.

Noteworthy is the remarkably low amount of hours accounted to unestimated activities, with an average of only 3.74% of the total hours going to unexpected work. These results make sense as the administrational activities are highly repetitive from project to project in the data, and should therefore be one of the easiest to predict. Note that Project C & D have an insufficient amount of activity estimates available for WP 1, and is therefore excluded from the UAH- analysis.

WP 1 - Project Administration

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	2 935	2 185.9	741.1	25.52%	25.52%	-92.4	4.23%
Project B	1 920	3 158.9	-1 238.9	64.53%	-64.53%	0	0.00%
Project C	1 058	1 202.5	-144.5	13.66%	-13.66%	n/a	n/a
Project D	2 335	1 249.1	1 085.9	46.51%	46.51%	-50	4.00%
Project E	2 380.5	2 786.8	-406.3	17.07%	-17.07%	-19.5	0.70%
Project F	1 765	3 658.7	-1 893.7	107.29%	-107.29%	n/a	n/a
Project G	1 222.1	1 701.3	-479.2	39.21%	-39.21%	-166.5	9.79%
Average/ Sum	13 615.6	15 943.2	-2 327.6	44.83%	-24.25%	-65.68	3.74%
Standard dev.				32.76%	52.26%	66.28	3.88%

Table 3: WP 1 data - Project Administration

4.3.2 Work Package 2 + 3 - Engineering

Every included project except one used work package two for all activities related to engineering indiscriminately. One of the projects had a few activities logged to WP 3 as well as 2, leading to the decision to merge the engineering data into a single work package. This will have no noteworthy consequence for the results, but is the reason why WP 3 does not have its own section.

Engineering is where the vast majority of hours are both estimated and spent on almost every project. There is in total 7 526 hours more spent on engineering activities across all the project than originally estimated, with an average overrun of 55.69% from estimates. Notice however that three of the projects actually overestimated and spent less they initially thought on engineering, however Project B has a serious overrun and skews the results considerably in that direction. Another important note is the significant amount of hours that were spent on activities the projects initially did not account for. The three biggest overruns all spent more than 60% of their total engineering hours on unpredicted work, and project D & E even spent more than the actual difference from estimate on unpredicted work. This indicates that the actual estimates made for the activities that they were able to predict, are more than likely actually overestimated. The absolute estimation accuracy for engineering activities are almost double that of the WP 1, with 82.65%. See Table 4 for all details.

WP 2 + 3 - Engineering

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	14 995	11 285.7	3 709.3	24.74%	24.74%	-160	1.42%
Project B	3 125	11 230	-8 105	259.36%	-259.36%	-7 684.5	68.43%
Project C	1 496	1 344.4	151.6	10.13%	10.13%	n/a	n/a
Project D	900	1 585	-685	76.11%	-76.11%	-1 015	64.04%
Project E	4 375	7 944.4	-3 569.4	81.59%	-81.59%	-6 679.9	84.08%
Project F	4 410	1 786.5	2 623.5	59.49%	59.49%	n/a	n/a
Project G	2 465	4 119.5	-1 654.5	67.12%	-67.12%	-74	1.80%
Average/ Sum	31 766	39 295.5	-7 529.5	82.65%	-55.69%	-3 122.7	43.95%
Standard dev.				82.32%	105.49%	3 740.9	39.47%

Table 4: WP 2+3 data - Engineering

4.3.3 Work Package 4 - Support Management

Work package four is mostly utilized throughout the projects for any hours spent on supporting customers and does usually not contain the most significant amount of hours, neither in estimates or actual. The data shows small overall deviations between estimates and actual, with project F as an exception from that. The absolute average is 72.46%, but removing project F from that analysis leaves a 20% accuracy, and a relative accuracy of about 0%.

Both Project C D had zero hours estimated on this work package but ended up spending a few hours anyway. These numbers will naturally result in two contributions of 100% on the UAH-analysis, but each project does still have an average of 339.14 hours spent on activities not initially accounted for. Another interesting observation is that Project E almost hit its target spot on, however non of the hours was actually spent on what they initially thought. It is possible that this is an accounting error, as it seems unlikely that they initially expected to spend over a thousand hours on one or more activity but ended up using zero and almost the exact same on something else instead. This specific work package has a lot of anomalies in its data and shows signs of the fact that it is very few activities on each project that make up the totals, making it hard to draw any conclusions from this. This is also reflected in the high standard deviations in the data. The WP has a total of 900 hours above estimates, which averages out to about a 20% overrun.

WP 4 - Support Management

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	1 343	1 676	-333	24.80%	-24.80%	0	0.00%
Project B	1 649	1 156.1	492.9	29.89%	29.89%	-335	28.98%
Project C	0	185	-185	n/a	n/a	-185	100.00%
Project D	0	24	-24	n/a	n/a	-24	100.00%
Project E	1 150	1 207	-57	4.96%	-4.96	-1 207	100.00%
Project F	345	1 139.2	-794.2	230.20%	-230.20%	-623	59.69%
Project G	0	0	0	n/a	n/a	0	n/a
Average/ Sum	4 487	5 387.3	-900.3	72.46%	-57.52%	-339.14	63.94%
Standard dev.				105.71%	117.32%	444.76	43.12%

Table 5: WP 4 data - Support Management

4.3.4 Work Package 6 - Production Management

All of the included projects expected to use some amount of hours on production planning, with all numbers being in the same order of magnitude. Three out of seven used less time than expected, while the last four used more. Every project shows a reasonable accuracy, except Project G. This project had the least amount of expected hours of them all with 89 hours estimated, yet used five and a half times more than that, making the average accuracy significantly worse. The overall hours in this work package are pretty modest compared to engineering or administrational work, yet the individual accuracies or the overall are only marginally better.

These marginal improvements do leave the work package at only a 5% overrun if you accumulate the numbers, yet the average absolute accuracy is 109% with a standard deviation of 163.7%, with four of the projects more than 40% off target. Project E appears to hit is mark pretty well, with an accuracy of -0.93%, yet almost 60% of the work done in project E was done on activities not initially expected. The overall amount of time spent on unestimated activities is however pretty low, with an average of 28% of the total time across all the projects.

WP 6 - Production Management

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	1 448	808.05	639.95	44.20%	44.20%	-110.55	13.68%
Project B	890	1 398,95	-508.95	57.19%	-57.19%	-535.95	38.31%
Project C	332	254.9	77.1	23.22%	23.22%	n/a	n/a
Project D	816	604.6	211.4	25.91%	25.91%	0	0.00%
Project E	360	363.35	-3.35	0.93%	-0.93%	-211	58.07%
Project F	155	382.5	-227.5	146.77%	-146.77%	n/a	n/a
Project G	89	502.7	-419.7	464.83%	-464.83%	-153	30.44%
Average/ Sum	4090	4315.1	-225.1	109.01%	-82.34%	-202.10	28.10%
Standard dev.				163.74%	181.01%	201.97	22.39%

Table 6: WP 6 data - Production Management

4.3.5 Work Package 7 - Logistics Support

Out of all the so far analyzed work packages, this appear to be the one of the most difficult to estimate accurately. The average absolute accuracy is over 300%, with three of the six project containing data having accuracies worse than 100%. These three projects severely underestimated the needed hours for their activities, and spent a large amount of hours on unexpected work in addition. Project E appears to have done a fairly good estimation to begin with, but spent almost 70% of the work on activities not originally estimated.

Out of all the so far analyzed work packages, this appears to be one of the most difficult to estimate accurately. The average absolute accuracy is over 300%, with three of the six projects containing data having accuracies worse than 100%. These three projects severely underestimated the needed hours for their activities and spent a large amount of hours on unexpected work in addition. Project E appears to have done a fairly good estimation to begin with, but spent almost 70% of the work on activities not originally estimated.

WP 7 - Logistics Support

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	100	1 021.5	-921.5	921.50%	-921.50%	-782.0	76.55%
Project B	239	1 972	-1 733	725.10%	-725.10%	-1793.0	38.31%
Project C	1 283	1 040	243	18.94%	18.94%	n/a	n/a
Project D	0	0	0	n/a	n/a	n/a	n/a
Project E	890	840.8	49.2	5.53%	5.53%	-575.8	68.48%
Project F	246	113.5	132.5	53.86%	53.86%	-36.0	31.71%
Project G	396.9	843.9	-447	112.62%	-112.62%	-343.5	40.70%
Average/ Sum	3 154.9	5 831.7	-2 676.8	306.26%	-280.15%	-706.06	61.68%
Standard dev.				406.98%	428.95%	667.95	24.80%

Table 7: WP 7 data - Logistics Support

4.3.6 Work Package 8 - Warranty

The last work package is strictly reserved to warranty elements and has some properties that make the analysis slightly different. As becomes apparent when looking at Table **??**, non of the included projects had any estimates for hours needed to be spent on warranty handling, however four out of seven projects did need to spend a significant amount of hours on it anyway. Every project undertaken by TechStick will have a fixed sum integrated in its budget to handle any warranty cases, however it is unknown if this fixed sum is also meant to cover the needed work-hours to handle these situations and is simply merged to a fixed cost for simplicity.

Regardless of what the actual case is, it is an interesting observation that the needed work hours to handle warranty situations are remarkably consistent from project to project. Given that the need for these extra hours is close to a binary situation, one could easily include more projects into the analysis of the warranty element that doesn't fit the selection criterias for this exact study, to calculate an expected value need for each project. With our sample, that would be about a thousand hours pr project added to each estimate to normalize the effect of the warranty elements across the portfolio. It is likely somewhat counterintuitive to estimate a need for warranty handling, when you as a project manager, of course, try to plan for a good project delivery as well, and could also open the door to the question of how many other buffers should be added as well.

WP 8 - Warranty

Projects	Estimate	Actual Hours	Difference	A. Accuracy	R. Accuracy	UAH	% UAH
Project A	0	1664.1	-1664.1	n/a	n/a	-1664.1	100.00%
Project B	0	1618.2	-1618.2	n/a	n/a	-1618.2	100.00%
Project C	0	0	0	n/a	n/a	0	n/a
Project D	0	2084.4	-2084.4	n/a	n/a	-2084.4	100.00%
Project E	0	0	0	n/a	n/a	0	n/a
Project F	0	1615.5	-1615.5	n/a	n/a	1615.5	100.00%
Project G	0	0	0	n/a	n/a	0	n/a
Average/ Sum	0	6982.2	-6982.2			-997.46	100.00%
Standard dev.						946.74	0.00%

Table 8: WP 8 data - Warranty

4.3.7 WBS Summary

Table 9 shows an overview of the total hours in each category along with the work packages' individual accuracies and average UAH. The engineering work package is the clear largest category, with more than double the amount of hours than project administration in second place. Engineering alone contains over 50% of all the completed project hours, with both categories combined totaling to about 71%. The remaining categories all have a comparable amount of hours in each of them, making up the final 29%.

Project administration is the clear winner of all performance indicators, with an average of 24% overrun and only 65 hours on average accounted unestimated activities, 3.74% of its total cost. Given the low amount of unestimated hours used, it is also possible that the average accuracies in this WP is a good indicator of how accurate the project is able to predict the needed hours for the activities they do remember to account for initially. Engineering have some large variations within its data set, which becomes apparent when comparing its absolute and relative accuracy, however this work package has a much more difficult time actually predict what they'll need to spend time on initially, with almost 44% of all spent time on activities they did not predict.

WBS Summary

	Total Estimate	Total Actual	Total Difference	Average A. Accuracy	Average R. Accuracy	Average UAH	Average % UAH
WP 1	13 615	15 943.20	-2 327.60	44.83%	-24.25%	-65.68	3.74%
WP 2+3	31 766	39 295.5	-7 529.5	82.65%	-55.69%	-3 122.68	43.95%
WP 4	4 487	5 387.3	-900.3	72.46%	-57.52%	-339.14	63.94%
WP 6	4 090	4 315.05	-225.05	109.01%	-82.83%	-202.10	28.10%
WP 7	3 154.9	5 831.7	-2 676.8	306.26%	-280.15%	-706.06	61.68%
WP 8	0	6 982.2	-6 982.2	n/a	n/a	-997.46	100.00%

Table 9: WBS data - Summary

Out of the remaining four work packages, logistics support appears to be the one with the lowest accuracy on their estimations, combined with more than 60% of their hours spent on activities they were unable to predict initially. Work package four shows accuracy results comparable to engineering, but it's hard to draw any conclusions from due to the few available data points. Work Package six, production management, is generally speaking not very accurate in the estimations, but appears to be able to account for most of the different activities needed from the start of the project.

4.4 Activity Analysis

The top-level, work package, analysis still left a lot of unanswered questions, all though providing a good initial overview. In this section, the lowest level activity estimates will be analyzed in as much detail as meaningfully possible. Here, it will be investigated both the activity estimated and their specific accuracy, as well as go into further details about the forgotten activities. Lastly some specific, and recurring individual activities will be analyzed in detail as well.

4.4.1 Unestimated activities

Table 10 have collected the total amount of UAH for each project, along with how large or a percentage that those hours make out of the total, and their relative accuracy for comparison. As is apparent by now, only 5 out of the seven projects have their complete activity estimate available, making it impossible to use two of the projects in much of the following analysis. Four of those five projects experienced an overrun in hours, while Project A did not.

Overview of Unestimated Hours

Project	Actual Hours	UAH	% UAH	R. Accuracy
Project A	18 641.25	-2 809.1	15.1%	10.47%
Project B	20 534.15	-11 966.7	58.3%	-162.48%
Project C	4 026.80	n/a	n/a	3.41%
Project D	5 547.10	-3 778.0	68.1%	-36.93%
Project E	13 142.35	-8 693.2	66.1%	-43.55%
Project F	8 695.90	n/a	n/a	-22.63%
Project G	7 164.40	-737.0	10.3%	-71.73%
Average		-5 596.8	43.6%	-46.21%
Standard dev.		4 606.0	28.5%	58.45%

Table 10: Overview of unestimated hours in projects

Noteworthy is the observation that three of the projects had a larger percentage of unestimated activity hours than the actual overrun itself. These two metrics cannot be directly compared in hours, as one is a function of the estimate, while the other is a function of the actual hours, however, with the exception of project B, this difference is marginal. The average percentage of unestimated activity hours is 43.6% of the actually spent hours and would be even higher as a function of the estimates due to the average overrun in the portfolio.

Both Project A and Project G stand out from the rest here, with a comparably low percentage UAH. The difference between the two is that Project A also has a good relative estimation accuracy, while Project G has the second worst of them all. This indicates that project A has pretty accurately predicted how much time they would need to spend on the activities they were able to identify, while Project G did not.

Table 11 breaks the results into the different work packages and provides information on how many different activities the various projects and WP's forgot, as well as their size in hours. As expected from the previous results, engineering has by far the largest amount of forgotten activities, with 11.4 forgotten activities on average per project. Project administration which had the second most project hours have the least amount of total forgotten hours, and these forgotten activities have the lowest average size, both per project and activity.

Unestimated Activity Hours

	WP 1		WP 2+3		WP 4		WP 6		WP 7		WP 8	
Projects	#	Sum	#	Sum	#	Sum	#	Sum	#	Sum	#	Sum
Project A	1	92.4	4	160.0	0	0	2	110.55	2	782.0	5	1 664.1
Project B	0	0	23	7 759.0	1	335.0	3	64.0	4	1 793.0	5	1 618.2
Project C	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Project D	4	50.0	6	1 015.0	1	24.0	2	420.1	0	0	6	2 084.4
Project E	1	19.5	23	6 663.4	5	1 207.0	1	211.0	4	575.8	0	0
Project F	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Project G	2	166.5	1	74.0	0	0	2	153.0	3	343.5	0	0
Total	8	328.4	57	15 671.4	7	1 566.0	10	966.7	13	3 494.3	16	5 366.7
Average pr project	1.6	65.7	11.4	3 134.3	1.4	313.2	2	193.3	2.6	698.9	3.2	1 073.3
Average pr Activity		41.1		274.9		223.7		96.7		268.8		335.4

Table 11: Amount of forgotten activities and hours in each project and WP

Despite the large amount of unestimated hours in the engineering work package, it is apparent that the average size of each activity is fairly consistent across all the different work packages, indicating that they have a similar idea of what size an activity should have but are simply not able to predict what work will become necessary. The data also confirms that WP 7 has one of the largest amounts of forgotten activities, not just in hours, but also in number of activities neglected. When investigating WP8, Warranty, it is also interesting to note that every project the experienced the need to handle warranty situations actually opened multiple new warranty activities per project. This is something that could be interesting to investigate further.

4.4.2 Estimated Activities

With the results found in the previous sections, especially the overall amount of unestimated hours spent in each project, it is only natural to take a closer look at the activities that the projects identified during estimation, and how accurate they turned out. Table 12 breaks down each project's estimated activities and presents their average deviation from estimates in terms of hours, both absolute and over-/underruns. Column three shows the percentage of underestimated activities in each project, and the last two columns present the average sizes of over- and underestimations individually. To emphasize, this table excludes any hours accounted to activities not originally estimated.

Activity Estimate Accuracy

Project	Average Estimate Deviation	Average Absolute Estimate Deviation	% Underestimated Activities	Average Overestimate	Average Underestimate
Project A	108.6	295.5	28.9%	285.70	-319.50
Project B	-26.6	162.8	55.2%	285.7	-176.75
Project C	n/a	n/a	n/a	n/a	n/a
Project D	167.7	205.2	27.3%	292.36	-123.07
Project E	336.2	680.8	35.7%	572.63	-89.46
Project F	n/a	n/a	n/a	n/a	n/a
Project G	-121.7	184.4	70.6%	120.1	-238.16
Average	92.8	305.7	43.53%	311.3	-189.4
Standard dev.			18.76%	163.2	92.0

Table 12: Estimation accuracy on activity level

This table gives quite a lot of new and relevant information. The most notable being that it is a pretty even distribution between over- and underestimated activities, with the majority of activities actually being overestimated. Only 43.53% of all activities estimated ended up with an overrun at project completion, with somewhere close to 50% being the expected value for a close to perfectly estimated project. The average deviation from estimates across all activities is also 92.8 hours below estimate, while the average absolute deviation from the estimate is 305.7 Hours. Also, notice how the average overrun is only half the size of the average underrun in all the projects, with only project G averaging higher overruns than underruns. This is however expected given its high overall underrun and few amount of unestimated activities in that particular project. As Figure 9 confirms the majority of estimates made are actually overestimated, with most activities sizing around 250 hours.

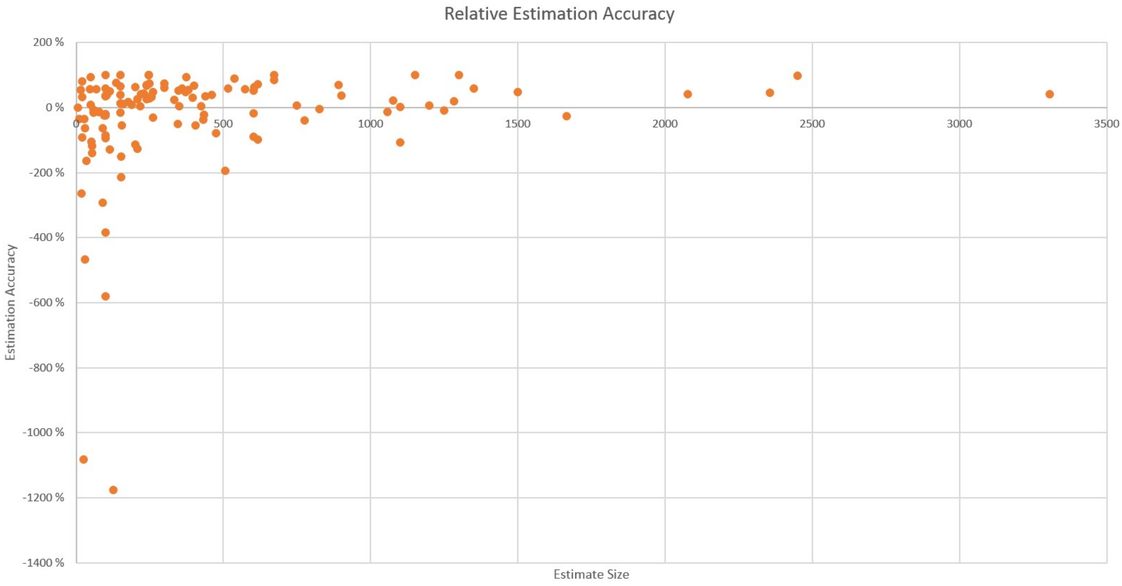

Figure 9: Scatter plot of relative estimation accuracies for activities

Looking at this, there is no apparent trend showing larger overruns on activities estimated to take longer time. In contrast, a few of the estimates below 100 hours in size actually have the most severe overruns. Looking at Figure 10, one gets a better picture of the expected estimation accuracy against the size of the activities. The dotted line is the regression line showing a clear increase in expected accuracy as the size of the estimates increases as well. A few things to notice is however that the amount of estimates made in the below 500 range is significantly larger than those above 500, so assuming a normal distribution around a mean, it is only natural to expect a larger amount of deviations around the area containing the largest amount of estimates.

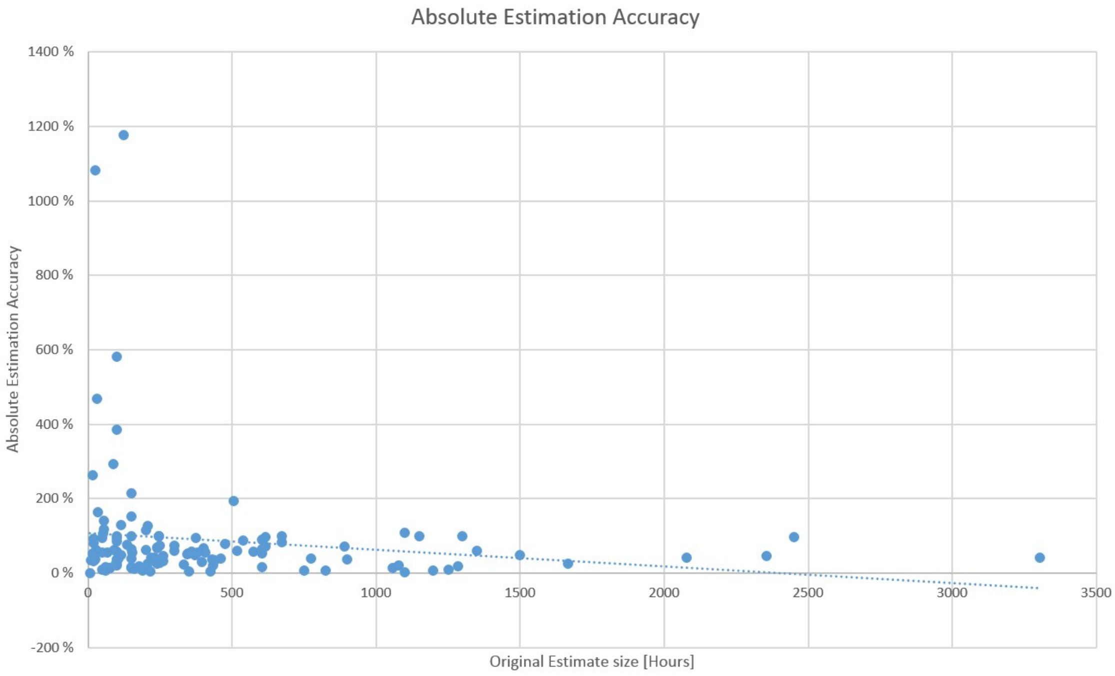

Figure 10: Scatter plot of absolute estimation accuracies for activities

When investigating the trend in Figure 10, one can observe a pretty stable deviation from the estimates with regards to the percentages. The effect this has on the actual hours becomes even more clear when looking at at at Figure 11, as the deviations from estimates all have a significant size on all estimates above 1500 hours in size. Given that the percentages remain decently stable, it does however not look like simply reducing the largest activities into smaller activities would result in any overall savings in terms of hours. This is however strictly seen from a numerical perspective, as a reduced activity size could have many other effects outside of this.

If we however divide the activity estimates into two different groups, those below 500 hours and those above 500 hours, we find some interesting numbers. Those below 500 have an average size of 187.6 hours and an overrun of just 12.8. hours on average, with the absolute deviation being 126.4 hours average. Summing up all the activities, this gives a total overrun of just 1151 hours from estimates, or just -7%. Those above 500 have an average size of 1114 hours and an underrun of 320 hours, giving them a total relative estimation accuracy of 29%. Out of the two groups, the below 500 group have the superior accuracy, while the above 500 group tend to overestimate the needed hours quite a lot. The standard deviations of the two groups are however large, with 219 hours on the below 500 group, and 714 and the above 500 group, making impossible to say with any statistical certainty the two accuracies are not actually the same in the entire population of projects and estimates.

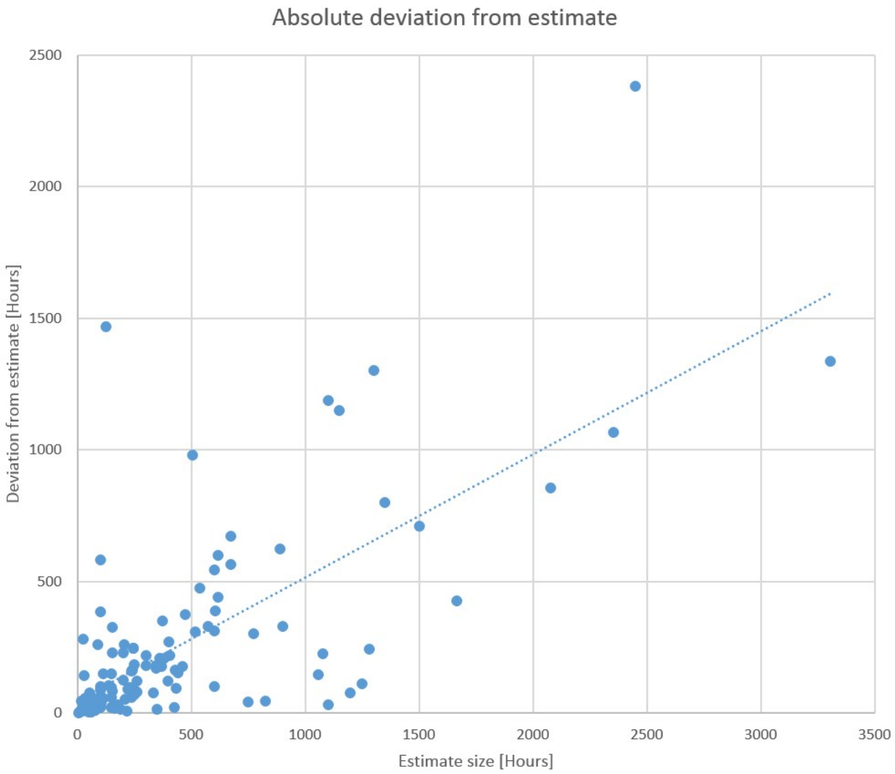

Figure 11: Relation between size of estimate and its deviation

4.4.3 Engineering Activities

After uncovering that the engineering activities represent about 50% off all expenses associated with any project, as well as being the work package with the most forgotten activities by a margin, it makes sense to investigate this closer. In the following analysis, we'll divide the engineering work package into five different engineering disciplines that all are present in the included projects to a varying degree. The following data combines both estimated and unestimated activities. Table 13 shows each of the five engineering categories, as separated by the numbering in the WBS. Sorted by project, the displayed data is the absolute deviation from estimates for each project, as well as the average deviation per activity.

Engineering Disciplines - Estimation Accuracy

	Hardware		Software		System		Test		Electric	
	Abs dev.	Abs avg.	Abs dev.	Abs avg.	Abs dev.	Abs avg.	Abs dev.	Abs avg.	Abs dev.	Abs avg.
Project A	906.2	151.1	2 544.0	318	459.3	229.7	407.5	101.9	585.1	195.0
Project B	194.0	48.5	5 281.2	514.3	1 119.5	279.9	269.5	53.9	176.0	176.0
Project C	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Project D	186.0	62.0	829.0	276.3	n/a	n/a	24.0	24.0	n/a	n/a
Project E	1 031.7	515.9	5 823.6	727.9	2 424.0	186.5	n/a	n/a	n/a	n/a
Project F	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Project G	74	74	n/a	n/a	152.0	50.7	1 561	390.3	n/a	n/a
Average	478.4	170.3	3 619.4	459.1	1 038.7	186.7	565.5	142.5	380.5	185.2
Standard dev.	452.51	197.24	2 349.7	207.1	1 007.9	98.4	682.4	168.2	289.3	13.46

Table 13: Size of estimate deviation in different engineering categories

Electrical engineering shows the least average deviation from estimates per project, with 380 hours. Testing has the best estimation accuracy when evaluated per activity, with a deviation of 142 hours per activity on average. All categories have a similar size on their average deviation per activity, with the exception of software engineering, which deviates with almost 460 hours from each defined activity on average. Software engineering also has the highest amount of average deviation per project, with 3.5 times more than the second-worst one - Systems engineering. The large amount of deviation per project is to some degree explained by an overall larger amount of software activities defined in the included projects, but the deviation per activity is still 2.5 times larger than the same for systems engineering. The reason why no percentage analysis was done on this part of the data set, is that a very large portion of these activities was not originally estimated, as seen in Section 4.4.1. The lack of an original estimate for a majority of the activities makes any sort of percentage analysis very skewed at best, and hard to interpret.

4.4.4 Administrational Activities

The last part of this analysis is a closer look at some of the most common activities included in almost every project, which is the administrational activities. The numbering on each activity shows where the individual activities belong in the WBS, and the name gives a fairly good pointer to what the different activities actually are. As these are included in almost every project and are bottom-level activities, the data paints a good picture of how accurate they are estimated.

Selected Common Activities - Estimation Accuracy					
Activity	Total Estimate [Hours]	Total Actual [Hours]	Total Difference	Average A. Accuracy	Average R.Accuracy
112 - Project Manager	7 449.5	5 833.0	1 616.5	21.19% (23.24%)	16.33% (27.55%)
113 - Project Coordinator	951.5	1626.6	-675.1	153.68% (214.03%)	-114.04% (240.24%)
117 - Contract Support	161.0	193.0	-32.0	77.87% (54.27%)	-27.27% (97.96%)
122 - Project Controller	797.9	1 203.2	-405.3	66.99% (78.37%)	-56.17% (87.98%)
141 - Quality Assurance	931.7	1 506.4	-574.7	115.97% (157.88%)	-89.87% (177.88%)
151 - Configuration Management	435.5	1 005.4	-569.9	150.07% (184.23%)	-130.74% (201.81%)
161 - Material Planning	792.5	422.0	370.5	109.21% (108.85%)	-22.79% (164.49%)
211 - Technical Management	4 646.5	5 157.0	-510.5	54.55% (28.8%)	-13.57% (67.47%)
711 - Logistics Management	1 839.0	836.0	1 003.0	46.22% (23.14)	38.02% (37.40%)

Table 14: Estimation accuracy for different administrational categories. Standard deviation inside the parenthesis

Project management, technical management, and logistics management are the largest activities of the nine included, with project management being the one with the most hours in total. Project management and logistics management is also the two only activities to not underestimate their needed hours, with project management having the best relative and absolute estimation accuracy, with a relative accuracy of only 16.3% below estimated on average. Technical management also has a rather good accuracy of -13.6% overrun. Project Coordinator and configuration management are the two activities with the worst results, both showing more than a 100% overrun on average and an absolute accuracy of more than 150%.

4.5 Industry Data

The following section is a brief summary of the results from the literature review, investigating the four selected sectors for reported cost overruns. Table 15 is the reported overruns from the transportation studies investigated, and Table 16 shows the reported overruns found in offshore-, IT-, and defense projects. Further interpretation of the results will be discussed in Section 5.

As one should note, there is an inconsistency in the reported numbers from the different authors, which will be discussed further in Section 5. Infrastructure was typically divided into three different sub-categories that studies separated the projects on; road-, fixed link-, and rail projects. Road projects have the lowest reported overruns of the three, also with the lowest numbers both in the high- and low end of the scale. Fixed link projects and rail projects both have significantly worse overall results reported than the road projects, with comparable results to each other. The rail projects have both higher and lower average numbers reported than fixed link, while fixed link projects have a lower deviation between its reported overruns.

Infrastructure				
Source	Road	Fixed link	Rail	Overall
Gao and Touran (2020)	n/a	n/a	48.5% Major 24.8% Minor	31.23%
Huo et al. (2018)	22.52%	35.59%	58.08%	39.18%
Cantarelli et al. (2012)	18.6%	21.7%	10.6%	16.5%
Welde et al. (2019)	15%	n/a	8%	14%
Andrić et al. (2019)	10.47%	n/a	21.11%	n/a
Flyvbjerg, Holm, and Buhl (2003)	20%	34%	45%	27.6%
Odeck (2004)	7.9%	n/a	n/a	7.9%
Odeck (2017)	26.9%	n/a	36.3%	34%
P. E. Love, Sing, et al. (2014)	12.49%	15.75%	n/a	13.3%

Table 15: Summary of literature review - Infrastructure

The remaining three categories all have a large range of different reported overruns. Offshore projects have the second-largest deviation in the results, with results from Norway as low as 8% and worldwide results as high as 53%. IT projects appears to be slightly higher on average based on these findings, with results between 12 and 45%. Welde et al. (2019)

is the only report of all, that found an average overestimation in the analyzed project portfolio. The rest of the defense projects have a pretty clear separation between the studies looking into US defense spendings averaging over 40%, and the European studies with way lower numbers

Offshore		
Source	Projects experiencing overruns	Average Overrun
Oljedirektoratet (2020)	17%	8%
Rui et al. (2017)	75%	18.2%
Merrow (2012)	78%	25%
EY (2014)	65%	53%
IT		
Bloch, Blumberg, and Laartz (2012)	50%	45%
Flyvbjerg and Budzier (2011)	n/a	27%
Yang et al. (2008)	51%	12%
Moløkken-Østvold et al. (2004)	76%	41%
Sauer and Cuthbertson (2003)	n/a	18%
Defence		
Welde et al. (2019)	n/a	Overestimated
Bolten et al. (2008)	n/a	38.8%
Younossi et al. (2007)	51%	46%
Swank et al. (2000)	76%	40%
Schwartz (2014)	n/a	32% (median)
NAO (2015)	n/a	11.4%

Table 16: Summary of literature review - Offshore, IT and Defence

5 Discussion

This section will be a discussion of the found data and literature, with the goal of answering the research questions asked in the introduction. Some of the information will also be seen in light of the results found in the prestudy. For a full understanding of some of the parts discussed in this chapter, it is therefore recommended that the author also reads Sandaa (2020), especially Chapter 4 - Results. It will however be attempted to briefly recap the most relevant results when seen necessary, so the recommendation is only for the most interested of readers.

5.1 Research Question 1

What parts of the estimated work experiences the most deviation from the actual work done?

This question can mostly be answered by looking at the data presented in section 4, all though the actual interpretation of the data might not be that straightforward. Overall, the data gives us an average overrun of 46.21% on the hours of each project. Similarly to the results found by Bolten et al. (2008), this data set also gives a larger overrun when examining averages instead of totals, indicating that the combined accuracy of the larger projects is somewhat better than those for the smaller projects, however the difference is only marginal with the total accuracy of -37.74%, less than 10% better than the average. Not enough to indicate any significant difference between the larger and the smaller projects in terms of overall accuracy at this level.

At the work package-level analysis, most of the work packages contained enough data from the projects to see some consistency and trends, with the exception of work package four, which contained too few data points and too many anomalies to draw any conclusion. Engineering and Administration represent the majority of estimated- and used hours in every included project, while the remaining five work packages represent the remaining 30% of project hours. Work Package one, Administration, is by far the most accurately estimated work package of all, with the best relative and absolute accuracy. In addition to this, they forget to estimate very few activities initially, and those few forgotten are typically very small. This is also an observation which makes some sense, as the project manager would expectedly be the person putting the most effort into the estimation process, as well as it is mostly recurring activities from project to project contained within this work package, which should make it more predictable for an experienced project manager. Each project forgets to account for 65.7 hours on average each in WP1, with each activity having an average size of only 41.1 hours; both metrics are the smallest across all work packages.

Looking at the activities contained within WP1, we find that the project manager-activity is the only one that is on average overestimated, with an accuracy of 16.33%. All the remaining activities are underestimated on average, with the project coordinator and the configuration management being the two worst, both using more than twice the estimated hours on average.

Engineering, on the other hand, appears to be one of the hardest work packages to estimate. It has the second-best relative accuracy after WP1, which is good considering it contains over 50% of all project hours, but it appears to have severe difficulties in predicting what activities will be needed. Three of the five projects with sufficient data actually have more hours spent on the activities they forgot, than on the total amount of hours spent above budget. Another one of the five accounted for most activities initially, but severely underestimated their size and had an engineering overrun of almost 70% anyway. The projects forget to estimate almost 44% of the engineering hours they end up needing on average, and the work package has a relative accuracy of -55.69%. In terms of percentages, the engineering work packages perform pretty average, but due to its large size it still represents the biggest overruns of hours in the projects.

Out of the five investigated engineering categories, software engineering has the worst estimates with the highest deviation from estimate both per project and per activity. The projects included do have an overweight in software engineering, which causes some bias in the average deviation per project as there is a tendency to underestimation. Still, the average deviation from the activity estimate is also much bigger. This metric is not affected by the bias, which leaves software with the worst estimates. Systems engineering is the second worst, but still far better than software. The remaining categories all have comparable results, with about 350-550 hours deviation from estimates per project and about 160 hours deviation per activity.

The worst performer of all the work packages is WP7, logistics management. Even though it has the least amount of hours estimated, it has the second-highest amount of unestimated activity hours after engineering, as well as the second largest overrun on total hours - also after engineering. The total overrun is even bigger than that of WP1, in spite of having only one-fourth the amount of initially estimated hours. The accuracy of this work package is -280%.

As illustrated continuously through the results, the data has relatively large standard deviations making it hard to find statistically significant results. The benefit of carefully selecting projects to include, has its downside in the low sample of projects possible to include. This makes it hard to pick up any trends, however including projects with an unknown scope change during its course would almost certainly introduce a negative bias to the results even though it would make the sample more consistent and probably also yield significant results. Simply not the results we are looking for.

5.2 Research Question 2

How does the estimation accuracy compare to other similar industry analyses?

As you might have figured by now, there are numerous reasons why this comparison is hard to do and even harder to get any value from. The first thing to mention is the obvious lack of any studies investigating the exact same topic, with the same selection criterias. As attempted illustrated in Section 3.4, there as many perspectives one can evaluate when doing overrun analyses. Most of the identified studies have investigated the overrun measured from early estimate at handover to project organization, too the delivery back to project owner. The main reason for this appears to be data availability, as most researchers use publicly available data from government sources. This leaves an ocean of potential cost mitigating and accumulating sources between the scope of these studies and this one, sources that are very hard to quantify or statistically measure and subtract. The argument could be made for both cases, that the observed results could be either higher or lower with no easy way of determining which is most likely. One could easily look to all the sampled categories and find examples of reported averages both higher and lower than the number found in this study, for either category.

The natural next thing to consider is the obvious factor that most of these studies examine total project cost, not just hours. This also leaves another layer of potential sources of error between the sampled projects and this study, for which many of the same arguments as in the previous paragraph can be used. Typical identified other sources are currency rates, material costs, shipping etc. (Swank et al. 2000; Hove and Lillekvelland 2016;

Oljedirektoratet 2020; Odeck 2017), which affects the results of this study minimally. This goes for all included projects in the literature review, with the exception of Yang et al. (2008), which is the only other study examining project hours and overruns. The data for that study was however sampled through questioners at a conference, which could very possibly also give it a positive bias in the results.

As was slightly discussed earlier as well, there still does not exist a commonly agreed-upon term for how an overrun should be measured, or what should be counted. While some authors such as P. E. D. Love et al. (2015) and P.Smirnoff and J.Hicks (2008) are very aware of this and raises the issue in their writing, a lot of researchers usually just combine data originating from cost growth, escalation, and overruns, and simply chose one of the terms to represent the combined results. Such practices make it hard to understand exactly what the research is investigating and what is included in the data, unless this is thoroughly explained in the methodology. While some researchers say they investigate cost overrun, it is often hard to tell if the observed results are actually overruns, escalations caused by things such as excessive inflation or growth due to controlled scope changes during the course of the projects. Flyvbjerg, Morris, et al. (2011) attempts to make the argument for optimism bias, quoting "*Conventionally, the following are listed as causes of project underperformance in the literature: project complexity, scope changes, technological uncertainty, demand uncertainty, unexpected geological features, and negative plurality, i.e. opposing stakeholder voices. No doubt, all of these factors at one time or another contribute to cost overruns and benefit shortfalls, but it may be argued that they are not the real, or root, cause. The root cause of underperformance is the fact that project planners tend to systematically underestimate or even ignore risks of complexity, scope changes, etc. during project development and decision-making.*", and while this study presents no data that contradicts this claim, it is noteworthy to observe that many previous overrun-studies by this research cluster make no obvious attempt to eliminate projects from the portfolio which undertook controlled cost growth through contract extension and similar during their lifetime. Studies such as Bolten et al. (2008) make the data and comparison more useful, as different influencing factors and contributions to projects are broken down and presented, however it is hard to be that detailed with any larger quantity of projects.

A couple of other things to consider are project size and geography. Project size is relevant as multiple researchers have investigated so-called megaprojects and found a statistical difference between their relative overrun compared to smaller projects, such as Merrow

(2012) and Huo et al. (2018) and P. E. D. Love et al. (2015), two of which is also included in the literature review. Many of the studies included in the literature review, such as Flyvbjerg, Holm, and Buhl (2003) also found a clear link between overrun and the project duration - a metric this study does not provide the necessary data for to make any comparison. Multiple of the included studies also found significant differences in the geographical results in their data, with usually a slightly lower overrun in European projects (**ey**; Flyvbjerg, Holm, and Buhl 2003; Merrow 2012; Rui et al. 2017; Odeck 2017), with the exception of P. E. D. Love et al. (2015), where Europe had one of the worst measured accuracies.

Now, with all these things in mind, one might try to make some comparisons, but it is in the authors' opinion that it should be used as a relative measurement unit for the results found in this study, not a direct comparison due to the many things mentions above. The results found in the prestudy was, however, from an estimation accuracy looking at total project cost and overruns, making it a much easier comparison. There, we found that TechStick averages 11% overrun on the comparable data. This is about the same accuracy as the European results from both the oil- and defence sectors(Oljedirektoratet 2020; NAO 2015), and comparable to some of the best results from the infrastructure projects as well, which had the lowest overruns of the investigated categories. It is in the author's opinion the industry TechStick operates within is most comparable to the defense- and offshore industry when looking at engineering complexity and type, making the overall performance very comparable to the European industry standard.

Out of the four categories, IT projects were found to have the largest average overrun, with the least deviation in the results. This is comparable to the results found here as well, with software activities suffering from the worst overruns of the different ones. Yang et al. (2008) is the only study that actually gives any insight into estimated work hours, and in contrast to this data set, is the average overrun in the Chinese IT industry only found to be about 12%. Interestingly, Bolten et al. (2008) found the electronics programs to have to lowest overruns of their included programs, which was also the best performing engineering category in this study as well. Lastly, it might be a stretch to draw a parallel between the mechanical engineering category in this study and the accuracy of offshore projects, as they are extremely complex projects containing a large array of different disciplines, but they both appear to be roughly in the middle of the bunch in terms of accuracy.

5.3 Research Question 3

What are probable explanations for the observed results?

It is apparent by now that it is not an obvious, direct relation between how many hours a project overruns and how large the overall project overrun becomes. While the analyzed projects had an average overrun of 46% of the estimated hours, the projects were found to have an average overrun of only 11% in total. It must be said that this is not measured from the exact same sample of projects, and with such a small sample, errors could be introduced because of this. Now while there is no doubt a connection between the two metrics, the projects appear to have a significant buffer stored in the fixed cost values of their estimates, to absorb the hit from the overran hours.

One of the advantages of this study is the strict selection of projects to include in the study, focusing the observed results as much as possible on the actual estimation process and eliminating as many sources of error as reasonably possible. The high amount of forgotten activities during the estimation process in the early phase could be an indicator of too little time spent on the estimations and project planning. It was discovered in the prestudy that TechStick has limited feedback to the project organization after project completion, with regards to how accurately the initial estimates turned out. This, combined with the uncovered fact that many project members base their estimates on estimates they have made for previous projects, could partially explain why some of the estimates are consistently wrong or forgotten.

One of the most interesting results, was the fact that 57% of the estimated activities were actually overestimated, with the remaining being underestimated. In addition to this was the average size of the overestimate almost twice as big as the size of the average underestimate. This could indicate that there is very little optimism bias (Flyvbjerg, Morris, et al. 2011) going about, but rather a project team that adds a little extra in fear of underestimating. Averagely, each activity estimate misses its target by about 93 hours, which is not horrible. However, the accuracy in the estimates is little help if you forget half the needed activities.

A final note is that there appears to be a better estimation accuracy for the activity estimates that are below 500 hours in size, than for those above. It is generally advised to keep the activity sizes way smaller than 500 hours in WBS literature, partially for that exact reason. This observation would also be consistent with that made by other researchers such as Flyvbjerg, Holm, and Buhl (2003) and Gao and Touran (2020), who expected a 4% increase in overrun for each ongoing project year, as larger activity estimates will take longer to complete. Hence a reduction in activity size, at least for the biggest ones, could be a mitigating move. As for the overall project size, there is no apparent trend indicating any larger overruns or worse accuracy for the larger projects, which is in line with findings from research such as Huo et al. (2018) and Flyvbjerg, Holm, and Buhl (2003), however differs from the results found by Odeck (2004) and Cantarelli et al. (2012).

6 Conclusion

This analysis yields many interesting results on different levels. The overall project cost estimation was found to be in line with the top performing studies from Europe within offshore- and defence project, both industries whihc very comparable to that of TechStick. Despite the average cost overrun only being 11%, is the average overrun on the hour estimate more than 46% for the projects estimated. With this observation, it is safe to assume the the poor estimation accuracy of project time is part of the reason why the projects typically overrun their estimated cost.

The project portfolio has more than 50% of its used hours in the engineering categories, another 20% in project administration, and the remaining parentage's distributed evenly across the other five work packages. Project administration is found to be one of the most accurate categories in this study, with an average overrun of about 24%, while logistics management had the worst accuracy with an average overrun of 280%. The activities estimated were found to be mostly overestimated, the only 43.5% of activities running above budget, and the average underrun being almost twice as large as the average overrun. It appears as if the biggest cause for the large overruns in hours has its origin in a large amount of activities that were not accounted for in the early stages of the projects, with 43.6% of all hours spent in the average project, being accounted to unestimated activities.

This study is however not able to find any statistically significant results that can draw conclusions across all the projects the company have completed, due to the low sample size and large deviations within the results.

References

Andrić, Jelena M., Abdul-Majeed Mahamadu, Jiayuan Wang, Patrick X. W. Zou, and Ruoyu Zhong (2019). "The Cost Performance and Causes of Overruns in Infrastructure Development Projects in Asia". In: *Journal of Civil Engineering and Management* 25.3, pp. 203–214.

Arena, Mark V., Robert S. Leonard, Sheila E. Murray, and Obaid Younossi (2006). "Historical Cost Growth of Completed Weapon System Programs". In: *Technical Report 343. Rand Corporation.*

Atkinson, Roger (1999). "Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria". In: *International Journal of Projects Management* 17.6, pp. 337–342.

Bloch, Michael, Sven Blumberg, and Jürgen Laartz (2012). "Delivering large-scale IT projects on time, on budget, and on value". In: *business technology office.*

Bolten, Joseph G., Robert S. Leonard, Mark V. Arena, Obaid Younossi, and Jerry M. Sollinger (2008). "Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs". In: *Monograph 670. Rand Corporation.*

Burek, Paul (2013). "The ABC basics of the WBS Paul Burek". In: *Paper presented at PMI® Global Congress 2013—North America. PA: Project Management Institute.*

Cantarelli, Chantal C., Bert Van Wee, Eric J.E. Molin, and Bent Flyvbjerg (2012). "Different cost performance: different determinants?: The case of cost overruns in Dutch transport infrastructure projects". In: *Transport Policy* 22, pp. 88–95.

Creswell, John W. and J. David Creswell (2018). *Research Design: Qualitative, Quantitative & Mixed Method Approaches.* 5th ed. SAGE Edge Publications.

EY (2014). "Spotlight on Oil and Gas Megaprojects". In: *Ernst Yong.*

Flyvbjerg, Bent, Atif Ansar, Alexander Budzier, and Søren Buhl (2018). "Five things you should know about cost overrun". In: *Transportation Research Part A Policy and Practice* 118.

Flyvbjerg, Bent and Alexander Budzier (2011). "Why Your IT Project May Be Riskier than You Think". In: *Harvard Business Review* 89.9, pp. 23–25.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Søren L. Buhl (2003). "How common and how large are cost overruns in transport infrastructure projects?" In: *Transport Reviews* 23.1, pp. 71–88.

Flyvbjerg, Bent, Mette K. Skamris holm, and Søren L. Buhl (2004). “What Causes Cost Overrun in Transport Infrastructure Projects?” In: *Transport Reviews* 24.1, pp. 3–18.

Flyvbjerg, Bent, Peter W. G. Morris, Jeffrey K. Pinto, and Jonas Söderlund (2011). “Over Budget, Over Time, Over and Over Again: Managing Major Projects”. In: *The Oxford Handbook of Project Management* Oxford: Oxford University Press, pp. 321–344.

Gao, Nan and Ali Touran (2020). “Cost Overruns and Formal Risk Assessment Program in US Rail Transit Projects”. In: *Journal of Construction Engineering and Management* 146.5.

Globerson, Shlomo (1994). “Impact of various work-breakdown structures on project conceptualization”. In: *International Journal of Project Management* 12.3, pp. 165–171.

Hajdu, Miklos and Orsolya Bokor (2014). “The Effects of Different Activity Distributions on Project Duration in PERT Networks”. In: *Procedia - Social and Behavioral Sciences* 119.

Hartley, Keith (2016). “UK defence inflation and cost escalation”. In: *Defence and Peace Economics* 27.2, pp. 184–207.

Heemstra, F. J. (1992). “Software cost estimation”. In: *Information and Software Technology* 34.10, pp. 627–639.

Hove, Kjetil and Tobias Lillekvelland (2016). “Defence specific inflation (DSI) of goods and services: A refinement of concepts and new estimates”. In: *FFI-RAPPORT* 16/00175.

Huo, Tengfei, Hong Ren, Weiguang Cai, Geoffrey Qiping Shen, Bingsheng Liu, Minglei Zhu, and Hengqin Wu (2018). “Measurement and Dependence Analysis of Cost Overruns in Megatransport Infrastructure Projects: Case Study in Hong Kong”. In: *Journal of Construction Engineering and Management* 144.3.

Hussein, Bassam A. and Ole Jonny Klakegg (2014). “Measuring the impact of risk factors associated with project success criteria in early phase.” In: *Procedia - Social and Behavioral Sciences* 119, pp. 711–718.

Jackson, Simon (2002). “Project Cost Overrun and Risk Management”. In: *Proceedings of Association of Researchers in Construction Management 18th Annual ARCOM Conference.*

Jalali, Samireh and Claes Wohlin (2012). “Systematic literature studies: Database searches vs. backward snowballing”. In: *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, pp. 29–38.

Jha, Kumar Neeraj and K. Chandrashekhar Iyer (2007). “Commitment, coordination, competence and the iron triangle”. In: *International Journal of Projects Management* 25.5, pp. 527–540.

Jung, Youngsoo, A.M.ASCE, and Sungkwon Woo (2004). “Flexible Work Breakdown Structure for Integrated Cost and Schedule Control”. In: *Journal of Construction Engineering and Management* 130.5.

Kelly, Jams E. and Morgan R. Walker (1959). “Critical-Path Planning and Scheduling”. In: *Proceedings of the eastern joint computer conference.*

Kemerer, Chris F. (1987). “An empirical validation of software cost estimation models”. In: *Communications of the ACM*, pp. 416–429.

Kolltveit, Bjørn Johs. and Kjell Grønhaug (2004). “The importance of the early phase: the case of construction and building projects”. In: *International Journal of Project Management* 22.7, pp. 545–551.

Lichtenberg, Steen (2000). *Proactive Management of Uncertainty Using The Successive Principle - A Handbook.*

Love, Peter E. D., Olugbenga Jide Olaniran, David Edwards, Oluwole Alfred Olatunji, and Jane Matthews (2015). “Cost Overruns in Hydrocarbon Megaprojects: A Critical Review and Implications for Research”. In: *Project Management Journal* 46.6, pp. 126–138.

Love, Peter E.D. and Dominic D. Ahiaga-Dagbui (2017). “Debunking fake news in a post-truth era: The plausible untruths of cost underestimation in transport infrastructure projects”. In: *Transportation Research Part A* 113.

Love, Peter E.D., Chun-Pung Sing, Xiangyu Wang, Zahir Irani, and Didibhuku W. Thwala (2014). “Overruns in transportation infrastructure projects”. In: *Structure and Infrastructure Engineering* 10.2, pp. 141–159.

Malcom, D. G., J. H. Roseboom, c. E. Clack, and W. Frazar (1959). “Application of a Technique for Research and Development Program Evaluation”. In: *Operations Research* 7.5, pp. 646–669.

Merrow, Edward W. (2012). “Oil and Gas Industry Megaprojects: Our Recent Track Record”. In: *Oil and Gas Facilities* 1.2, pp. 38–42.

Moløkken-Østvold, Kjetil, Magne Jørgensen, Sinan S. Tanilkan, Hans Gallis, Anette C. Lien, and Siw E. Hove (2004). “A Survey on Software Estimation in the Norwegian Industry”. In: *10th International Symposium on Software Metrics*, pp. 208–219.

NAO (2015). “Major Projects Report and Equipment Plan 2014 to 2024”. In: *National Audit Office.*

Niazi, Adnan, Jian S. Dai, Stavroula Balabani, and Lakmal Seneviratne (2005). “Product Cost Estimation: Technique Classification and Methodology Review”. In: *Journal of Manufacturing Science and Engineering* 128.2.

Odeck, James (2004). “Cost overruns in road construction — What are their sizes and determinants?” In: *Transport Policy* 11.1, pp. 43–53.

— (2017). “Variation in cost overruns of transportation projects: An econometric meta-regression analysis of studies reported in the literature”. In: *Transportation* 46, pp. 1345–1368.

Oljedirektoratet (2020). “Prosjektgjennomføring på norsk sokkel”. In: *Rappor tnr. OD-02-20.*

P.Smirnoff, James and Michael J.Hicks (2008). “The impact of economic factors and acquisition reforms on the cost of defense weapon systems”. In: *Review of Financial Economics* 17.1, pp. 3–13.

Pollack, Julien, Jane Helm, and Daniel Adler (2018). “What is the Iron Triangle, and how has it changed?” In: *International Journal of Managing Projects in Business*, pp. 527–547.

Qasim, Iqra, Hanny Tufail, Alia Fatima, Tayyaba Rasool, and Farooque Azam (2017). “Cost Estimation Techniques for Software Development: A Systematic Literature Review”. In: *International Conference on Engineering, Computing Information Technology*, pp. 38–42.

Rehman, Sumaira and Marin D.Guenov (1998). “A methodology for modelling manufacturing costs at conceptual design”. In: *Computers Industrial Engineering* 35.3.

Reichelt, Kimberly and James Lyneis (1999). “The dynamics of project performance: benchmarking the drivers of cost and schedule overrun”. In: *European Management Journal* 17.2, pp. 135–150.

Rui, Zhenhua, Fei Peng, Kegang Ling, Hanwen Chang, Gang Chen, and Xiyu Zhou (2017). “Investigation into the performance of oil and gas projects”. In: *Journal of Natural Gas Science and Engineering* 38, pp. 12–20.

Sandaa, Kristian (2020). “Cost Estimations for Projects in Uncertain Environments: An Accuracy Analysis”. In: *Specialization Project - Unpublished.*

Sauer, Chris and Christine Cuthbertson (2003). “The State of IT Project Management in the UK 2002-2003”. In: *Templeton College, University of Oxford.*

Schwartz, Moshe (2014). “Defense Acquisition Reform: Background, Analysis, and Issues for Congress”. In: *Congressional Research Service.*

Swank, Major William J., Dr. Paul A. Alfieri, Charles K. Gailey III, and Raymond W. Reig (2000). “Acquisition Trend Metrics in the Department of Defense”. In: *Defense Acquisition University. Technical Report TR 1-00.*

Toor, Shamas-ur-Rehman and Stephen O. Ogunlana (2010). “Beyond the ‘iron triangle’: Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects”. In: *International Journal of Projects Management* 28.3, pp. 228–236.

Walpole, Ronald E., Raymond H. Myers, Sharon L. Myers, and Keying Ye (2017). *Probabilities Statistics for Engineers Scientists.* 9th ed. Pearson.

Welde, Morten, Magne Jørgensen, Per Fridtjof Larssen, and Torleif Halkjelsvik (2019). “Estimering av kostnader i store statlige prosjekter: Hvor gode er estimatene og usikkerhetsanalysene i KS2-rapportene?” In: *Concept-Rapport* 59.

Wohlin, Claes (2014). “Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering”. In: *Blekinge Institute of Technology.*

Ou-Yang, C. and T. S. Lin (1997). “Developing an integrated framework for feature-based early manufacturing cost estimation”. In: *The International Journal of Advanced Manufacturing Technology* 13, pp. 618–629.

Yang, Da, Qing Wang, Mingshu Li, Ye Yang, Kai Ye, and Jing Du (2008). “A survey on software cost estimation in the chinese software industry”. In: *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, pp. 253–262.

Younossi, Obaid, Mark V. Arena, Robert S. Leonard, Jr. Charles Robert Roll, Arvind Jain, and Jerry M. Sollinger (2007). “Is Weapon System Cost Growth Increasing? A Quantitative Assessment of Completed and Ongoing Programs”. In: *Monograph 588. Rand Corporation.*

NTNU

Norwegian University of
Science and Technology