Anonymised version

Witteveen+Bos & University of Twente

Analysing and predicting the costs of the Amsterdam Quay Wall renovation project: using parametric estimation.

Civil Engineering Bachelor Thesis

Molin, B.J. (Bruno, Student B-CE)
14/03/2024

Colophon

Title: Analysing and predicting the costs of the Amsterdam Quay Wall renovation project: using parametric estimation.

University: University of Twente, department of Civil Engineering and Management (CEM)

Company: Witteveen+Bos (Deventer, Leeuwenbrug)

Supervisors: Wilco Tijhuis (University of Twente) & Erik Schulte (Witteveen+Bos)

Author: B.J. Molin

Email: B.j.molin@student.utwente.nl

Date: 14-03-2024

Preface

This thesis is the result of the final phase of my bachelor's. I performed this research at Witteveen+Bos in Deventer, where I worked from November 2023 until February 2024.

During this research, I learned a lot, from the basics of contracting, to how a large project like this is achieved and organised in real life by multiple organizations.

I would like to thank Erik Schulte for welcoming me to Witteveen+Bos and being my helpful supervisor there. Furthermore, I would like to thank Wilco Tijhuis for being my supervisor at the University of Twente and for providing me with the necessary feedback for reaching this report.

Furthermore, I would like to thank Witteveen+Bos and the colleagues there for the conviviality and the pleasant lunch walks.

Bruno Molin

Table of Contents

Table of Figures

Table of tables

Summary

In the centre of Amsterdam, a large project is executed. This project is to inspect all 215 kilometres of quay wall in the city and renovate where necessary. To keep up with the maintenance, 2 kilometres of quay wall renovation needs to be done every year. However, currently the budget available is only half that of the necessary budget. Furthermore, the cost differences between the projects in the first batch were very large. Therefore, the municipality of Amsterdam requested Witteveen+Bos to research the costs of the quay wall renovation project.

The goal of this research is to provide insight into the costs of renovating quay wall projects in Amsterdam. This is achieved by finding Cost Estimating Relationships (CERs) in historical data of quay wall renovation projects, which helps to understand the costs in the projects and predict costs in future projects. Furthermore, determine what the costs are of requirements from the municipality of Amsterdam. Finally, to provide recommendations based on the findings in the research and on how to improve the research in the future.

To build the dataset, projects are gathered. Project data from the municipality of Amsterdam and the Witteveen+Bos server is used. Then, the cost data of the projects is structured using a classification system. This is done to see what the project costs are composed of. When comparing the cost composition per project it becomes clear that large differences exist in the costs per project. Furthermore, the cost categories themselves are examined, to find the largest costs and largest contributors to the costs in the project. It is found that on average the foundation costs are the most expensive part of the project. The foundation also has the largest standard deviation, making it the largest contributor to the cost differences.

The Cost Estimation relationships are determined using regression analysis. First possible predictors are determined, second the predictor data is gathered, third the data is normalised for inflation and the side projects and requirements are removed. Then, simple linear, nonlinear, and multiple linear regression analyses are done, and later on, the best results are validated.
The best CER models found are the Contract type and Length to the Costs (€), and the Contract type and retaining area to the Costs per meter (€/m). The contract type is good at explaining the cost differences between the projects, however, the use for prediction is questionable, due to the municipality of Amsterdam switch in the use of contract types around 2020.

Then, the costs of the requirements from the municipality of Amsterdam are looked into by analysing the costs of trees, BLVC measures, Transport HUB & NTC and houseboat costs. It is found that the costs of retaining a large historical tree are the highest.

Based on the findings in the research, recommendations are given to the municipality of Amsterdam. The most important advice is to only use the IPK (Innovation Program Quay walls) and RAW (the standard contract type for civil works) contract types since the SOK (Cooperation Agreement Quay wall constructors) is the most expensive contract type of these three as can be seen In the developed Cost Estimation Relationships.

Then advice is given for further data gathering. Where the most important predictors to gather are given.

The results of the research provide insights into the costs of the quay wall renovation project in Amsterdam. These insights can be used to reduce the costs of future quay wall renovation projects. By making better choices based on the type of contract type, or to work for longer stretches/ increase the construction speed. Furthermore, the advice can be used to improve the research in the future by more precisely gathering and storing data.

1. Introduction

1.1. Introduction

In this chapter the research is introduced, first background information is given, and then the problem statement and the research objective are explained. Which aims to describe what the research focuses on and what the research tries to achieve. Furthermore, the scope and boundaries of the research are addressed, describing what is in the scope of this research and what is not.
This paragraph discusses:

- Background information
- Problem statement
- Research objective
- Scope of the research
- Boundaries of the research
- Research questions

1.2. Background information

On September 1, 2020, a section of the Grimburg wall in the canal of Amsterdam collapsed (Figure 1). This is the cause of a new project from the municipality of Amsterdam to inspect and if necessary renovate the quay walls in Amsterdam (*Peschier*, 2021). The municipality of Amsterdam has to inspect all 215 kilometres of the quay wall in the city and renovate it where necessary. (*Gemeente Amsterdam*, 2022).

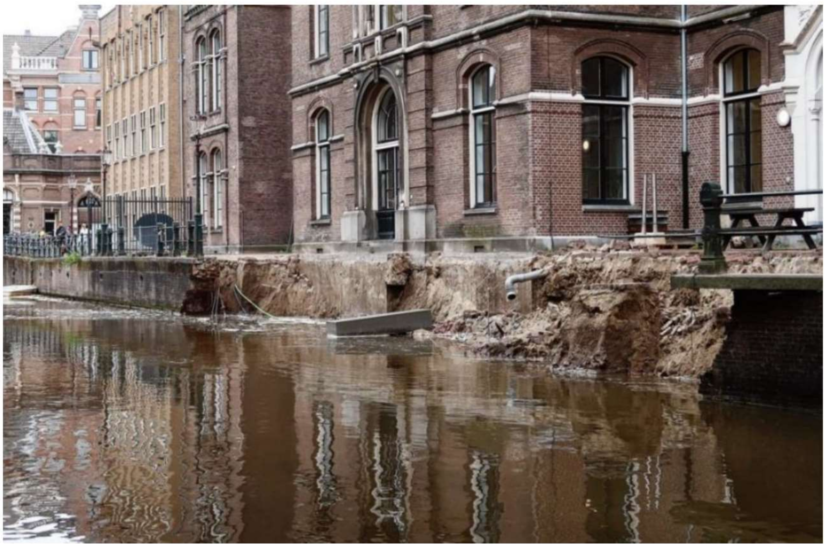


Figure 1 Collapsed quay wall in Amsterdam (source: Herter (2020))

The life span of a quay wall is about 100 years, which means that to keep up with the maintenance of the quay walls, 2 kilometres of renovation needs to be done every year. This is set as the goal by the municipality, to renovate 2 kilometres of quay walls every year. As long as this goal is not reached, the problem of overdue maintenance only grows (*Gemeente Amsterdam*, 2022).

This goal is much more than is currently being done, i.e. In 2018 only 100 meters of quay wall renovation was done every year. Therefore a large scale-up of the process is necessary (*Gemeente Amsterdam*, 2022). Currently, 3 alliances are testing innovative techniques in Amsterdam, which aim to speed up the process of renovating the quay wall whilst also reducing the costs (De Leeuw, 2018).

If these tests prove to be successful a framework agreement will be signed (*Programma Bruggen en Kademuren*, 2023).

According to Gemeente Amsterdam (*2022*), the municipality of Amsterdam has set multiple requirements for the construction of the quay walls. Because of the BLVC-plan of Amsterdam (Bereikbaarheid (Connectivity), Leefbaarheid (Livability), Veiligheid (Safety) en Communicatie (Communication)), the municipality has set requirements e.g. to work from the water to prevent hindering the inhabitants or traffic that travels through the area. Furthermore, the municipality has requirements to save more trees or to also work on the street surface when the quay wall is renovated. It is expected that the requirements of the municipality of Amsterdam influence the costs of the renovation program (*Witteveen Bos*, 2023).

The problem the project currently faces is that the cost estimation for the project came out high. The costs of the first 9 projects of the program currently ongoing or already finished were high, also the estimations for the first phase are high *(Project Intake Formulier,* n.d.). In the years 2023-2026, the budget available was only 50% of what is required to reach the desired ambition of 2 kilometres every year (*Gemeente Amsterdam*, 2022). Furthermore, there was a large difference in the costs per meter between the projects of the first batch.
The large costs and the differences in costs between the projects have a large impact on the planned and necessary renovation operation in Amsterdam. The task looks no longer financially feasible.

Since the costs were so high the municipality of Amsterdam has asked Witteveen+Bos (W+B) to research the costs and provide an explanation for the differences in the costs. Therefore W+B researched the first 9 available projects of the program. W+B looked for Cost Estimating Relationships (CERs), to explain the differences in costs between the projects and better estimate the costs of future projects. However, the results of the first analysis did not prove a single cause of budget overrunning and therefore W+B is interested in finding better CERs.

Furthermore, in the coming years, more quay wall renovation project data from Amsterdam will become available due to more projects being finished. Also, Witteveen+Bos has historical data available on quay wall renovation projects.

This bachelor thesis is executed at Witteveen+Bos in Deventer. The research looks for better Cost Estimating Relationships by expanding the number of projects used for the analysis. The Cost Estimating Relationships (CERs) are determined using parametric estimation. Furthermore, the costs of the 'additional' requirements from the municipality of Amsterdam are analysed. Then finally, based on the analysis, advice is given to save costs and on how to gather data to improve the CERs in the future.

The data that is used in the analysis are the actual cost incurred in the quay wall renovation program and historical cost data available to Witteveen+Bos.

This research is relevant since there is still much work to be done in the project of renovating the 200 kilometres of quay wall, and since the costs estimated are larger than the available resources. This analysis can help the municipality make better-considered choices, due to better estimations and therefore more realistic budgets for the program can be set.

1.3. The problem statement

In the quay wall renovation project in Amsterdam, the ambition that needs to be renovated is 2 kilometres per year. However, due to large costs in the years 2023-2026, the available budget was only half of necessary to reach the ambition of 2-kilometre renovations per year. Furthermore, there are large differences in the costs between the projects in the first batch of the program, it is unknown what is cause of this. Moreover, the municipality of Amsterdam has set requirements for the renovation project, which are expected to also influence the costs, but it is not clear what that effect is.

The problem statement is as follows:
It is unknown what drives the costs in the quay wall renovation program in Amsterdam, why there are large cost differences between different construction locations and what the costs are of the additional requirements from the municipality of Amsterdam.

1.4. The research objective

Witteveen+Bos is interested in finding Cost Estimation Relationships (CERs), to explain the cost differences between the projects of the quay wall renovation project in Amsterdam, and to better estimate the cost in future projects. Furthermore, Witteveen+Bos would like to get insight into the costs of the requirements of the municipality of Amsterdam. Since this program has just started and will continue for many years to come, important recommendations based on this research can be used to improve the project in the future. Furthermore, this is the first analysis. When more projects are known in the future and more data is gathered, the analysis can be improved.

The research objective is as follows:
To find Cost Estimating Relationships (CERs) in the quay wall renovation program in Amsterdam, which can help understand the cost differences between projects, and better predict the costs in projects to come. Furthermore, to provide insight into the costs of the requirements of the municipality of Amsterdam, such as transporting materials over water or retaining trees.
Finally, to provide advice based on the findings of this research and on how to improve the analysis of the costs in the future.

1.5. Scope of the research

The scope of the research is:

To examine the Cost Estimating Relationships in the quay wall renovation program in Amsterdam. Based on the available data to Witteveen+Bos. This includes data from the 9 finished/ongoing projects which are already briefly analysed, newer available data and historical data.

Moreover, the costs of the requirements from the municipality of Amsterdam for the quay wall renovation project are analysed.

Furthermore, advice is given to the municipality of Amsterdam based on all the findings done in the research. This can include, advice on further research or findings which are useful for the quay wall renovation program. Furthermore, advice is given on how to improve the data gathering for the research to be expanded.

1.6. Boundaries of the research

The boundaries of the research are used to further clarify the scope of the study, by removing some elements that might be seen as part of the scope but are not.

Only the replacement of the quay wall is taken into account. Repairing, life-expanding or emergency construction works on the quay walls are not taken into account.

Based on the following statement by Witteveen+Bos, it has been decided to perform the research only on the available cost data at Witteveen+Bos, and not to include any quay wall costs from the literature.

"Literature shows a broad range of construction costs for quay wall construction. First of all, this is often based on newly built quays (like in port expansion) and secondly, this is not inside a dense city centre. Therefore we believe unit rates used in literature are not reliable for cost comparison in the Amsterdam situation." (*Witteveen Bos*, 2023).

1.7. Research questions

This research focuses on the following research questions, these questions act as a sort of common thread through the research. During the research, the answers to the questions are gathered and in the conclusion, the answers are discussed.

- *Where are the project costs composed of and what are the largest contributors to the cost differences?*
    - *What are the project costs composed of per project?*
    - *What are the average project costs composed of?*
    - *What are the largest contributors to the cost differences?*

- *What is the best Cost Estimating Relationship that can be found?*

- *What do the requirements of the municipality of Amsterdam cost?*

- *Which advice can be taken from the analysis?*

- *What data should be gathered for further research?*

1.8. Structure of the report

The report starts by stating the research objective and the research questions. Then relevant context to the quay wall renovation program is given. Continued by the theoretical framework, which explains the steps in the research and provides the theory. Then the chapters go through the steps. First, the project's data is gathered and an analysis of where the costs are in the project is done. Second, the Cost Estimation Relationships process is done step by step. Here the relationship data is gathered, regression analysis is performed and the results are verified. Third the costs of the requirements from the municipality of Amsterdam are analysed, followed by recommendations to the municipality of Amsterdam based on the findings in the report, and on the gathering of data is given. Then a discussion is done, followed by the conclusion. The report ends with final recommendations.

1.9. Resume

It is unclear why the costs are so high in the quay wall renovation project in Amsterdam, why there are such large cost differences and what the costs are of the requirements by the municipality of Amsterdam.

This research focuses on finding Cost Estimation Relationships (CERs) in quay wall renovation projects. Furthermore, the costs of the requirements from the municipality of Amsterdam are addressed. Next to that, advice is given based on the findings of the research.

This is relevant since the costs are higher than the available budget to reach the minimal goal of renovating 2 kilometres every year. Therefore, it is important to see what drives the cost differences between projects and how the costs of the quay wall renovation projects can be reduced.

2. Context

2.1. Introduction

In this chapter, the different contract types and construction types are explained. These have a large impact on the construction process and are therefore used for the Cost Estimation Relationships process. Furthermore, the requirements of the municipality of Amsterdam are discussed, of which the costs are determined later on in the research.

The following topics are discussed:

+ Contract type
+ Construction types
+ Requirements of the municipality of Amsterdam

2.2. Contract type

In quay wall renovation projects there have been different contract types. The contract form can influence the work the contractor has to do, and therefore the costs of the project. This report distinguishes the following 3 contract types: RAW contract, SOK Kademakers and IPK program, which are explained next.

2.2.1. RAW contract

According to Pianoo (*Pianoo, n.d.*), the RAW has already been in use for 30 years in the Netherlands. RAW stands for Rationalisatie en Automatisering Grond-, Water- en Wegenbouw (Rationalisation, Automation, Ground-, Water-, and Road construction). This is the standard specification system in the construction field of Ground-, Water-, and Road construction.

In this contract, the client is responsible for the design and the contractor does the construction, this can be found in Appendix A. With the RAW contract, there is free market tendering, where multiple contractors can make their offer based on the design, to win the project. This process could take up to 6 months and can start after the design is finished (*W+B Cost Expert 1, 2024*).

2.2.2. SOK Kademakers

According to Ingenieursbureau Gemeente Amsterdam (2021), three parties are working within SOK Kademakers. SOK Kademakers stands for Samenwerkingsovereenkomst Kademakers (cooperation agreement Quay wall builders). In this program, parties work with traditional methods. In January 2021, an agreement was signed that for the next 6 years, these parties would each renovate 300 meters of quay wall annually. The parties are:

1. Dura Vermeer Infra Regionale projecten
2. Beens Groep
3. Aannemingsmaatschappij H. van Steenwijk, with subcontractors Mobilis and Van Gelder

The contract type looks like the RAW contract however, it differs in a few ways.

+ In SOK, there is no free market tendering, due to the parties already being known. As stated in 2.2.1. RAW contract, the free marketing tendering process could take up to 6 months, therefore time is saved.
+ Since the parties are already known from the start, the contractor sits at the design table (*De Bouwcampus, 2021*). The idea is that in this way disputes between the contractor and the designer can be solved in an early stage (*W+B Cost Expert 1, 2024*).
+ The % profit and risk is fixed (*De Bouwcampus, 2021)*

+ The rate list is predetermined (*De Bouwcampus, 2021)*

The biggest difference between the SOK compared to the RAW contract is that the parties are known beforehand.

2.2.3. IPK program

According to Gemeente Amsterdam (*2022),* 3 alliances are currently testing their new innovative techniques to renovate the quay wall. This program is called the IPK program (Innovatie Programma Kademuren (Innovation Program Quay Walls)), it was started to speed up the process of renovating the quay walls and to reduce the costs. In this program, contractors came up with new innovative methods, which are tested in a pilot project. If these tests succeed, framework agreements will be signed for 4 years.

According to the Leeuw (*De Leeuw*, 2020), the three participating alliances are:

1. G-Kracht (Van Gelder, Gebr. De Koning en Gieken)
2. Kade 2020 (Sweco Combinatie Midden Delfland, Bouwadviesbureau Strackee en Oosterhof Holman Beton- en Waterbouw)
3. Koningsgracht (BAM Infra Nederland en Royal Haskoning DHV)

In this contract type, the contractors themselves design the construction method of the quay wall. In the current contracts, there is no maintenance responsibility for the contractors (*W+B Supervisor, 2024)*.

2.2.4. Change in tendering method

According to Ingenieursbureau Gemeente Amsterdam *(n.d.),* in 2012 a new tendering method was introduced, EMVI tendering. This is a European tendering law which requires the tendering process to not only be awarded based on the cheapest offer but also to look at what brings the best quality.

This might have caused an increase in the overall costs of the quay wall renovations. However, this influence is difficult to take into account in the comparison. This might explain possible differences, in the analysis in later chapters of the report.

2.3. Construction types

In the research, there has been made a distinction between 4 types of quay wall construction types. These construction types influence the work to be done and can therefore influence the costs. The 4 different construction types in this report are: IPK, L-Wall, Suspended concrete slab and Other construction types, which are discussed in this paragraph.

2.3.1. IPK

The IPK program, as can be seen in 2.2.3. IPK program, uses innovative techniques to build a quay wall. These innovative techniques can be different to the construction method of the traditional techniques or the structure itself. Therefore, in this research, they are put in their own category. It is important to note, however, that all these IPK techniques are different from each other.

2.3.2. L-Wall

The L-Wall is a traditional construction method. A schematic drawing can be found in Figure 2, here it can be seen that the L-wand (in blue) is constructed on the foundation. Usually, the concrete is poured on-site, the soil is then put on the long flat surface of the L-shape (right) until the height of the L is reached and on the small surface (left) masonry is constructed up to the height of the L.

When the concrete is poured on-site, a construction pit (retaining wall) is necessary to keep the construction dry.

2.3.3. Suspended concrete slab

The Suspended concrete slab is also a traditional construction method. The concrete is suspended from the foundation, which can be seen in Figure 2 where the blue part is the suspended concrete. To make this construction possible the concrete is prefabricated, often with the masonry already largely constructed. Only the connection between the suspended concrete slabs needs to be finished by hand, which reduces the time necessary for the masonry to be constructed on-site. Furthermore, there is no need for a construction pit using this method.

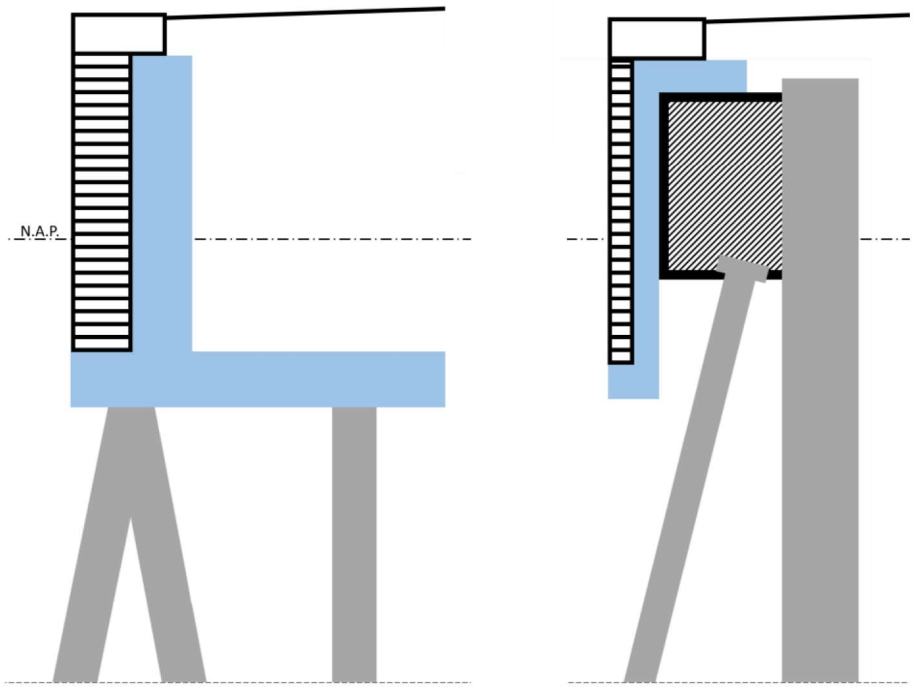


Figure 2: Drawing of an L-Wall (left) and Suspended concrete slab (right), based on the drawings of Bloemgracht (Witteveen+Bos, 5 Juni 2008) and Krom Boomssloot (Witteveen+Bos, 7 July 2015)).

2.3.4. Other

The final category consists of the other projects. These are projects which have a steel sheet foundation and use other finishes like wood or stone covering stone.

2.4 Requirements of the Municipality of Amsterdam

The municipality of Amsterdam has set multiple requirements for the renovation of the quay walls. Of course, there are standard design requirements, e.g. the quay walls being designed to last 100 years (*Gemeente Amsterdam*, *2022*). Also, the requirements for the aesthetics are fixed, since the historical city centre is protected by UNESCO world heritage (*Gemeente Amsterdam*, *2022*).

These requirements are for the construction process of the quay wall, without changing anything to the design. First, there are the BVLC measures, which purpose is to minimize hindrance and nuisance in the city (*Gemeente Amsterdam (a)*, *n.d.*).

According to Gemeente Amsterdam (*2022*), for the construction of the quay walls in Amsterdam. The municipality has set requirements for working from the water. In this way, the roads can stay open and the inhabitants plus traffic have less hinder of the construction works. Another requirement can be to work exclusively with electrical equipment.

According to Verheijen. T. (*2021*), on 1 October 2021, new weight restrictions were introduced in the city centre of Amsterdam. This is done to protect the old quay walls and reduce emissions in the centre. This restriction is for vehicles heavier than 7,5 tonnes, however, up to 30 tonnes an exemption can be applied for. This caused a problem for the construction projects in the centre since these projects often require heavy or many materials. For this problem, transport HUBs are set up. In these HUBs, materials are gathered and shipped over water or in smaller vehicles to their destination in the centre.

According to Gemeente Amsterdam (*2022*), sometimes it is also chosen to combine projects. E.g. also work on the resurfacing of the road together with the replacement of the quay wall. Some other construction works that are often combined are: work on pipelines and cables, improving growing spots for trees, adding charging poles for cars and boats or placing underground garbage containers.

Furthermore, to protect the historical city centre it is required to keep many of the trees. Using traditional construction methods, only young trees could remain in place. This is done by replanting the tree before the construction and planting it back after construction, old and larger trees needed to be cut down. The expectation is that with the new techniques, more trees can remain in place (*Gemeente Amsterdam, 2022*).

The requirements discussed in this paragraph are:

- BLVC measures
- Construction work from water
- Working with electrical equipment
- Supply of materials over water (via Transport HUB)
- Retaining Trees (or improving growing spots)
- Surface works:
    - Resurface of road
    - Underground garbage containers
    - Charging poles for cars/boats

2.5. Resume

This chapter discussed the relevant context of the project. This information is used later on in the research.

There are three types of contracts used for quay wall renovations in this research, these are RAW, IPK and SOK. These different pose different limitations and contractors are involved in different stages of the design process. Two SOK and IPK are new and are replacing the RAW type.
Furthermore, there are 4 different categories for types of constructions considered in this project. These are: IPK constructions, L-Wall, Suspended concrete slab and Others.
Next, the municipality of Amsterdam has set multiple requirements for the construction of the quay walls. These are: BLVC measures, construction from water, working with electrical equipment, supply of materials over water, retaining trees and surface works requirements.

3. Research method

3.1. Introduction

In this chapter, the research method is explained. The chapter starts by introducing Cost Estimation Relationships (CERs), which is the basis of the cost analysis in this research. Then the research framework and the phases of the research are discussed. The chapter discusses the following:

+ Cost estimation relationships
+ Research Framework
+ Phase 1, Gather datasets and the composition of project costs.
+ Phase 2, Cost Estimation Relationships Process.
+ Phase 3, Costs of requirements and advice.

3.2. Cost Estimation Relationships

The cost expert team at Witteveen+Bos is interested in using Cost Estimating Relationships (CERs) to determine the cost drivers in the quay wall renovation projects for the assignment of the municipality of Amsterdam (*W+B Supervisor, 2024)*. Therefore Cost Estimating Relationships are determined for cost estimating in this research and no search for an alternative method is undertaken. To determine the Cost Estimating Relationships the NASA methodology is used. The NASA methodology for Cost Estimating Relationships is 'common practice' at Witteveen+Bos.

According to NASA (*NASA, 2015*) Cost Estimation Relationships (CERs) can be determined using parametric estimation. Parametric cost estimation is typically used when there is a larger set of measurements and only a few parameters of a project are known, for example, the volume and weight of a project and the costs. Parametric cost estimation then uses the relationship between the historical data and the predictors (parameters) to design the CER. It is assumed that the things that affected the costs in the past also influence the costs in the future. This technique does not prove the cause and the effect. However, it assumes causality. Despite this, it is good at predicting costs

How this technique can be applied is explained in Regression analysis

There are certain limitations to this technique, which are discussed below:

3.2.1. Overfitting

According to Mathworks (*Mathworks, n.d.)*, overfitting occurs when the model is too well adapted to the input data (to the underlying pattern but also to the noise) so that the model cannot properly predict new data. This can be seen in Figure 3 on the left: Instead of being a smooth curve, the line zigzags around the points. This can have two causes:
First, the model is too complex. Second, the set of input data is too small for the complexity of the model, or the input data contains irrelevant information.
Also, the opposite can occur, where the model is not properly adapted to the input data. This is called underfitting. This is shown in Figure 3 on the right, where the line does not seem properly adapted to the input data.
A proper fitting model is shown in the middle in Figure 3.

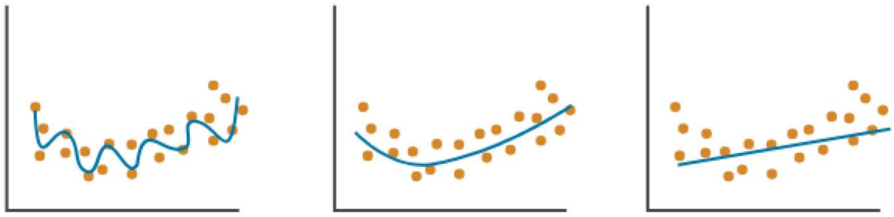


Figure 3 Examples of overfitting (left), right fit (middle), and underfitting (right) (Mathworks, n.d.)

3.2.2. Extrapolation

When using a CER, it is very dangerous to use a CER outside the range of the input data (*NASA, 2015*). According to Shukla. M, (*Shukla. M, Mujumdar. A, n.d.),* this is because it is not known what the effect is outside the range of the input data. It might be that the true regression line, outside of the input data range is completely different.

An example: if you measure the sunlight on earth during 18:00-00:00 and plot a regression line of the amount of light over time. The regression line would imply that it only becomes darker, and the amount of light would never increase. However, as everyone knows in the real world the sun would of course rise again in the morning, and the amount of light would increase again. The regression line can therefore only be used within the input range.

Another example of this problem is shown in Figure 4.
Due to the risk of extrapolation, it is advised to only use the model within the input range. Since the model cannot predict what happens outside this input range.

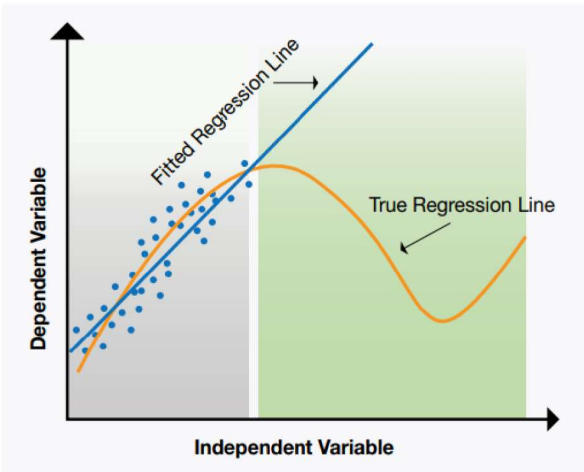


Figure 4 Example of extrapolation (Shukla. M, Mujumdar. A (n.d.))

3.3. Research Framework

In this paragraph the research framework is explained, first, the steps are briefly addressed and then an overview of the process is given.

In Figure 5 the different steps in this research are shown. The breakdown of the process comes down to 3 phases:

1. Phase 1, Dataset collection and the composition of project costs.

2. Phase 2, Cost Estimation Relationships Process.
3. Phase 3, Costs of requirements and advice.

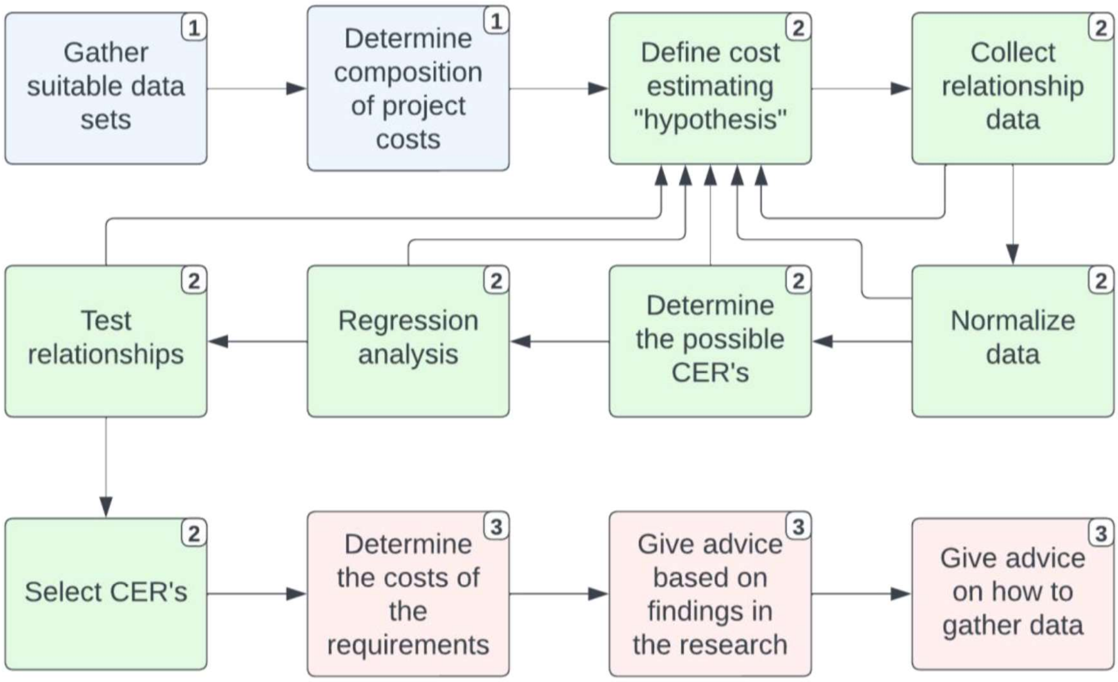


Figure 5 Research Framework

3.4. Phase 1: Dataset collection and the composition of project costs

In this phase, the suitable data sets are determined. The data set is suitable when enough data is known for the parametric estimation process.

Furthermore, an analysis is done to determine the composition of the project costs, where the average costs are within the projects and what costs cause the largest cost differences between projects. This might help with choosing the predictors during phase 2 when searching for CERs.

3.5. Phase 2: Cost Estimation Relationships Process

In phase 2, the process of determining the Cost Estimating Relationships is discussed. For the CER process, the NASA cost estimating handbook is followed *(NASA, 2015)*.
The CER process is iterative, as can be seen in Figure 5 most steps leading back to the first step. In the report, however, all the steps are discussed chronologically without going back and forth.

The following steps are discussed:

1. Define cost estimation "hypothesis"
2. Collect relationship data
3. Normalise data
4. Determine the possible CERs
5. Regression analysis
6. Test relationships
7. Select CERs

1. Define cost estimation "hypothesis"

In this step the possible predictors are thought of, this is done by logically thinking and or brainstorming with experts. If the chosen set does not provide a sufficient result, the next steps lead back to this step.

2. Collect relationship data

Collect the data of the possible predictors.

3. Normalise data

In this step, the data is normalised and prepared for the analysis. E.g. the influence of inflation is removed.

4. Determine the possible CERs

Determine the possible CERs by analysing the data and looking at the scatterplots of the different predictors which were chosen in the first step. Here the best correlating predictors are picked for the next steps.

5. Regression analysis

All the steps are briefly discussed here, a more detailed explanation can be found in Appendix B.

The regression analysis consists of two parts, the first part of the regression analysis is to quantify the relationship between variables using a mathematical expression. The second part is the validation. This paragraph discusses the first part. The second part is discussed in the next paragraph (Test relationships).

The regression analysis starts by making assumptions (5.5.1. Assumptions). These assumptions act as a guideline, in practice often one or more assumptions are violated.

Formula

The formula for a multiple linear regression model is:

$$\boldsymbol{\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \cdots + \hat{b}_k X_k}$$

$\boldsymbol{\hat{Y}}$= estimated value of dependent value (estimated cost)

$\boldsymbol{X_i}$= ith dependent variable (predictors, e.g. length)

k = number of independent variables

$\boldsymbol{\hat{b}_i}$= estimated regression coefficient

Simple linear regression

In Simple linear regression, the formula consists only of the first part since it has only 1 independent variable ($\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1$). The regression coefficients can be determined by using standard formulas. Excel can calculate these coefficients when fitting a regression line through a set of data.

Simple nonlinear regression

When the scatterplot of the data shows a nonlinear relationship, a simple nonlinear regression can be tested. A power function can be used in this case. Another method is transforming the data using logarithms, to make a linear regression on what appears to be a nonlinear regression.

The formula for a power function in simple nonlinear regression is: $\hat{Y} = \hat{b}_0 * X_1^{b_1}$

Multiple linear regression

When using multiple linear regression, the coefficients should be estimated using an iterative process or with a statistical program. Excel can calculate the coefficients using the data analysis tool pack. The best fit is determined by employing the method of least squares. Where the aim is to minimize the sum of squares of the error terms.

As discussed in 3.2.1. Overfitting, overfitting can occur when the model is too complex or when the set of input data is too small for the complexity of the model. Therefore, there should be a maximum amount of independent variables present for the multiple regression analysis. If too many independent variables are used the risk of overfitting increases. According to Statistics Solutions (Statistics Solutions *b, n.d.*), the rule of thumb is that there should be at least 10 observations per independent variable.

6. Test relationships

According to NASA (*NASA,* 2015): The second part of the regression analysis is the validation, this is to define the accuracy of the relationship. This is important since this can tell how good the found relationship is. This is achieved by using various statistics, which are briefly described in this paragraph, how these can be calculated can be found in Appendix C.

The assumptions made in the first step of the regression analysis are also checked in this step.

These are the four methods of analysing the found regression, which are discussed below:

The regression coefficient of determination (R^2)

The regression coefficient of determination (R^2) explains how much the independent variable (predictor) can explain the difference in the dependent variable (costs). E.g. R^2 = 0,7, means that the predictor can explain 70% of the variation in the costs, leaving 30% up to chance.
For multiple regression analysis, the adjusted coefficient of determination is used. This takes into account the loss in degrees of freedom.

The coefficient of determination is used as the first test to determine a good relationship, the predictors with the best R^2 coefficients are used for further validation. The closer the coefficient of determination is to 1, the better. What R^2 coefficient is sufficient depends on the situation.

In Multiple linear regression, the Adjusted Coefficient of determination is used (R_a^2). This value takes the loss of degrees of freedom into account, which is the result of using multiple regression analysis. This makes it possible to compare the results of the simple and the multiple regression.

Standard Error of Estimate (SE)

The Standard Error of the Estimate (SE) tells the deviation of the sample points from the fitted regression line. This is an absolute value and can therefore on itself not be used for evaluation. However, it is used to calculate the coefficient of variation.

Coefficient of variation (CV)

The Coefficient of Variation (CV) tells the degree of variation. It is the Standard Error of the Estimate divided by the mean of the dependent variable ($CV = \frac{SE}{\bar{Y}}$). Which enables it to compare with other models.

The CV should be less than or equal to 20%, preferably less than 10%.

Strength of regression coefficients (hypothesis test)

The hypothesis test is performed to determine if the coefficient ($\hat{b}_i$) of the independent variable is significantly different from 0. The slope of 0 would imply that the relationship is purely chance. Which would imply that X is of no use in predicting Y.
The significance of the intercept ($\hat{b}_0$) is not of importance since this is almost always outside the range of the data, which would impose extrapolation and is therefore not taken into account.

In the case of multiple linear regression, the F-test is done for the entire model. When the existence of the relationship is established. The F-test is done for all the coefficients.

7. Select CERs

In this step, the best CERs are selected.

3.6. Phase 3: Costs of requirements and advice

Phase 3 starts by determining the costs of the requirements set by the municipality of Amsterdam for the construction of the quay walls.

Then this phase advises the municipality of Amsterdam based on findings in the research.

Furthermore, advice is given on how to gather data for further research. Which data should be gathered, what data shows promising results, but also what was not been taken into account due to data not being available.

3.7. Resume

This chapter discussed the theoretical framework of the project.

Cost Estimation Relationships (CERs) are determined using regression analysis. This technique works by using historical cost data and predictors (parameters) to design the CER. Certain limitations to this technique are overfitting and extrapolation. Furthermore, the research consists of three parts. The first part focuses on the gathering of datasets and looks into the composition of the project costs. The second part is about the Cost Estimation Relationship process, where all the different steps are addressed. The regression coefficient of determination (R^2) is used to explain how much the predictor can explain the difference in the dependent variable (costs), the closer to 1 the better.
The third part looks into the costs of the requirements by the municipality of Amsterdam and gives advice, both on findings in this research and on data gathering for future research.

4. Dataset and the composition of project costs

4.1. Introduction

This chapter discusses how the dataset is collected and an overview of the locations Is provided, this dataset is used later in the report to develop Cost Estimation Relationships. Furthermore, an analysis is done to see where the costs are within the projects and what type of costs are responsible for the largest cost differences in the project. The last gives a boost to the next phase since it can help the process of finding the predictors in the CER process.

+ Dataset and reliability of data
+ Project cost composition and largest contributors to cost differences

4.2. Dataset and reliability of data

This paragraph discussed the dataset. The source, filters and assumptions used to reach this dataset are discussed. Then, the paragraph provides an overview of the locations of these projects.
The following parts are part of this paragraph.

+ Dataset
+ Overview of project locations
+ Reliability of data

4.2.1. Dataset

This paragraph discusses the dataset. Stating the sources of the data, and providing an overview of the filters and assumptions made to reach the dataset. However, this phase has led to a scope change, which is discussed in the box below.

Scope change

In this stage an important change has been made to the scope of the study. Instead of only using the actual costs incurred of the project. It has been decided to add other types of costs as well.

This is done based on the following reasoning:

- Witteveen+Bos has performed a research on the first 9 projects available from the municipality of Amsterdam. Here the results were not satisfactory and therefore they were interested to increase the dataset in hope to find new conclusions.
- Due to data scarcity it is not possible to increase the dataset with actual costs incurred. Since Witteveen+Bos does not have actual costs incurred project data of quay wall renovations. Only estimations, budgets or the sum for which the project is awarded is known to W+B.

Due to the fact that Witteveen+Bos is interested in increasing the dataset, and the limitation that there is no possibility to increase the dataset when only using the actual costs incurred within the scope. It has been chosen to also include the detailed design estimate, tender stage client estimate and contract sum. This, in order to significantly increase the number of projects for the analysis.

However, a downside of using other types of costs. Is that the costs estimated in other earlier phases of the project have (higher) variation. Therefore this variation is discussed in 4.2.3. .

The suitable projects from the Witteveen+Bos server are determined first by looking at the database of W+B and then analysing the project files found in the server. A broad explanation of this process can be found in Appendix D, however, the overview of the applied filters is presented below:

- Search terms: 'kademuren' or 'kademuur' (Dutch for Quay wall(s))
- Cost data after the year 2000 [1]
- Quay wall is located inside the city centre [2]
- Quay wall is not part of a (large) harbour
- Contained: detailed design estimate, tender stage client estimate, contract sum and/or actual costs incurred (contract sum + contract variations) [3]

[1] The costs are normalised for inflation in 5.4.1. Normalize for inflation.
[2] To further increase the dataset it has been chosen to also include projects outside of Amsterdam. However, the construction conditions should be more or less the same, therefore this filter is applied.
[3] It has been assumed that in the project data, the most recent tender bid is the contract sum.

Using these filters, 11 projects from the Witteveen+Bos server are selected to be included in the dataset for this research. Including 3 projects outside of Amsterdam.

Furthermore, 14 projects from the municipality of Amsterdam are included in the dataset. Which are 5 more projects compared to the first research by W+B. However, of 2 projects the exact location of the quay is unknown, only the name of the Quay is known.

4.2.2. Overview of project locations

In the following figure, the overview of the project locations can be found (Figure 6).A detailed list of the location and data source can be found in Appendix D Table 14 Overview of the projects used in the analysis As discussed in the last paragraph there are a few projects outside of Amsterdam, two of which are in Haarlem and one in De Rijp.

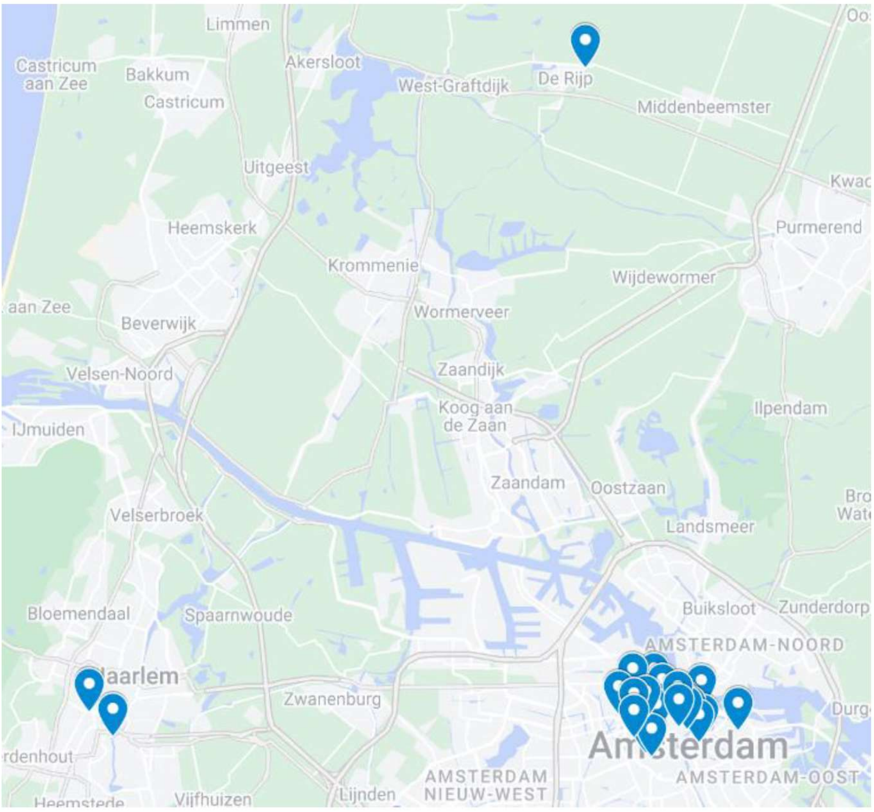


Figure 6 Project locations in this report (anonymised)

4.2.3. Variation

As discussed in the scope change in 4.2.1. Dataset, the initial scope of the study was to only work with the actual costs incurred in the project. That would translate to the contract sum plus contract variations. However, due to data scarcity, it has been decided to add projects with other types of costs. Other types of costs mean that the project is in another stage at the time that the costs were calculated/estimated. Which would result in higher variation due to possible changes in further stages. The differences in the cost type are discussed first and later the variation of these costs is discussed.

Costs differences

- Detailed design estimate. In the design phase, estimations are done for the costs. In this research only the detailed design estimate is taken into account, other earlier estimates are deemed too unreliable. The detailed design estimate is done by the designer and has the largest variation coefficient of the types of costs used in the project.
- Tender stage client estimate, also referred to as client estimate. At the end of the design phase, the designer made the construction specifications, which is a list of all work that needs to be done. The designer then budgets the construction specifications, resulting in the client estimate.
- Tender bid. If the contractor budgets the construction specifications and makes his bid, this results in the tender bid. If accepted, this becomes the contract sum.
- Contract sum. This is the sum of which the contractor signs the contract. In case the project goes according to the construction specifications, the contract sum is the final sum. If there are design changes by the client during the project, the costs are for the client. Unexpected events can cause additional costs, who is responsible for these costs depends on the situation. The additional costs for the client can be found in the contract variations. For this project, the contract sum + contract variations are seen as the actual costs incurred.

Variation

The lower the variation coefficient, the lower the more reliable. Due to the costs being estimated/calculated in different phases by different parties, there are differences in the variation coefficient of the costs. The detailed design estimate has a higher variation coefficient than the contract sum. Based on common practice (*W+B Supervisor, 2024*):

- The client estimate has a variation coefficient of 5% to the contract sum.
- The detailed design estimate has a variation coefficient of 10% to the contract sum.

Here contract variations are out of the scope since there is calculated to the construction specifications.

On average the projects where the contract variations were known, the contract variations were 5% of the actual costs incurred (Appendix E, Table 15). The three highest outliers are (Table 1):

Table 1 Highest percentage contract variations and their reasons (anonymised)

Location	**Percentage contract variations of actual costs incurred**	**Reason of contract variations**
[redacted]	10,4%	Nitrogen problems during construction
[redacted]	9,6%	Incomplete construction specifications at the start
[redacted]	16,0%	Changes in construction specifications

The most and largest contract variations come from changes in the construction specifications, in the table above it is assumed that the nitrogen problems are a one-time event. So if no changes are made here, and no expected events occur, the contract sum is 100% of the total costs. Therefore it is assumed that the contract sum has no variation.

The average variation coefficient of the dataset is on average 1,4% (Appendix E, Table 16), which is relatively small. This is small because there is only 1 detailed design estimate and 5 client estimates out of the 25 projects in total. Due to this small variation coefficient and since it is not possible to include this variation coefficient in the dependent variable using the NASA method for determining the CERs, the uncertainty in the data is not taken into account in this research. This results in lower reliability of the final models. It might be that the models are better or worse than is validated.

4.3. Project cost composition and largest contributors to cost differences

In this paragraph, it is determined what the project costs are composed of, per project and on average. Furthermore, it is determined what the main cost drivers are. This is done by first structuring the data using a classification system. Then it is possible to compare the costs per project and see outliers per project. Then the largest average cost categories are looked at and next, the largest contributors to the cost differences are analysed.

The following chapter discusses:

+ Structuring the data
+ Cost composition per project
+ Average project cost composition
+ Largest contributors to cost differences

4.3.1. Structuring the data

In the data of the projects, there are differences in how the costs are labelled. Therefore it is difficult to compare the projects with each other and see what the project costs are composed of. To determine where the costs are within the project, a classification system is used to organize the costs in certain groups. These groups can then help to make clear what the project costs are composed of.

The system that is used is stated below in Table 2: Cost classification system The costs experts at Witteveen+Bos designed this system for the first analysis on the quay walls. Using this method it becomes possible to differ the different types of costs within each project.

Table 2: Cost classification system Witteveen+Bos (22-9-2023)

nr	Description	nr	Description
11	Surface design/sewage/cables and piping	75	Transport HUB & NTC
13	Trees	77	Side projects (Houseboats, Pont etc.)
21	Construction pit and auxiliary works	81	Monitoring, research and end documentation
23	Auxiliary bridge structures	84	Fees, insurance and bank guarantees
24	Earthworks	87	Risk reservation- posts
26	Dredging activities/cleaning waterway	91	One-off costs and construction site costs
27	Drainage	93	Management + execution costs
31	Demolition quay wall	95	Engineering costs
41	Foundation	96	Overheads
44	Concrete work	97	Profit/risk
51	Masonry and cover stones	9697	Overheads/profit/risk
54	Quay facilities	98	Provisional sum/ place at disposal of
61	Guide piling	99	Contribution
71	Phasing BLVC-measures	index	Indexations

4.3.2. Costs composition per project

To visualize where the costs are within the project, the costs are put in bar charts. Here they are put in larger groups than is done in 4.3.1. Structuring the data, which is done for better visualization. Furthermore, they are divided between the direct costs and the indirect costs. The charts can be found in Figure 8 Direct costs composition per project & Figure 7 Indirect costs composition per project. The locations of the project were already shown in Figure 6.

It can be seen that the column of [redacted] contains no data. This is due to the limited data available from this project, however, this project is used later on in the analysis wherever possible. Furthermore, [redacted] has very low direct but very high indirect costs, this is due to the % engineering costs being very high compared to the other projects. This is also the only detailed design estimate that is present in the database. In the indirect costs, the one-off- and construction site costs in [redacted] are negative, this is due to the residual value of the temporary steel quay wall that was removed at the start of the project.

Furthermore, there are large differences in the composition of costs between the projects.

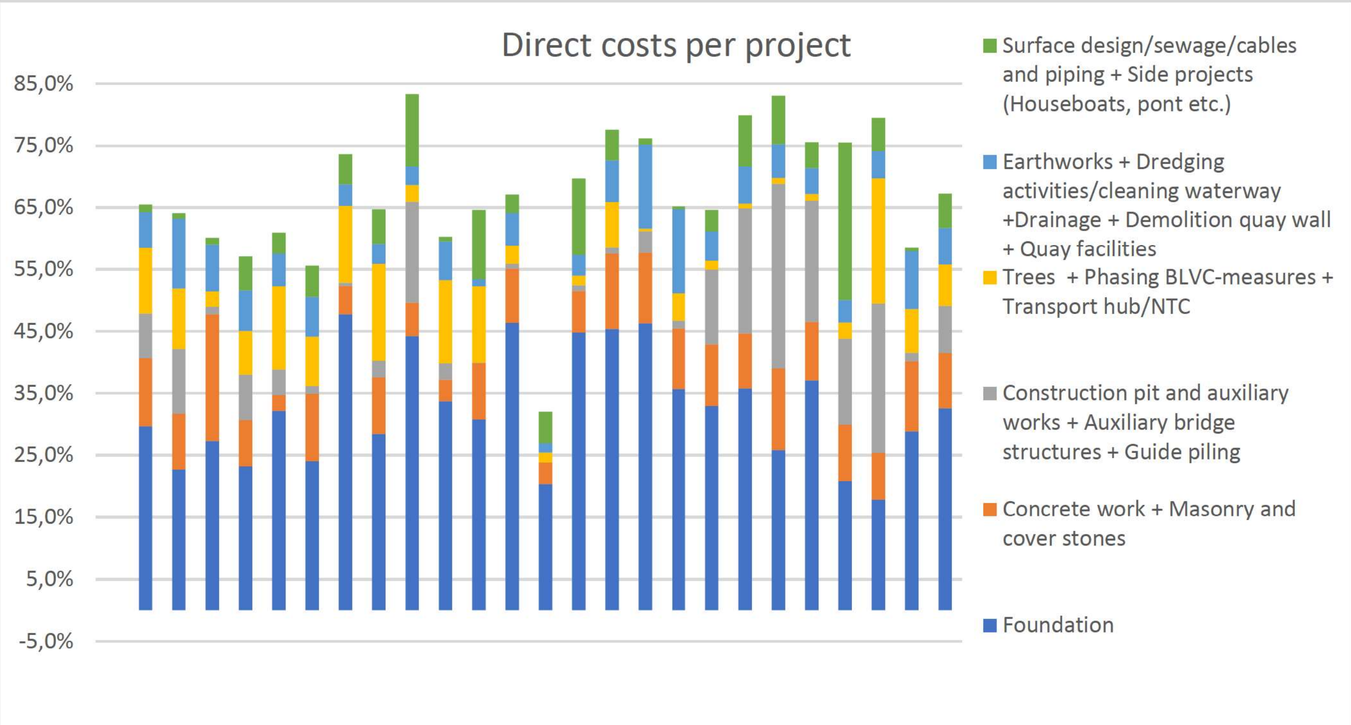


Figure 8 Direct costs composition per project (anonymised)

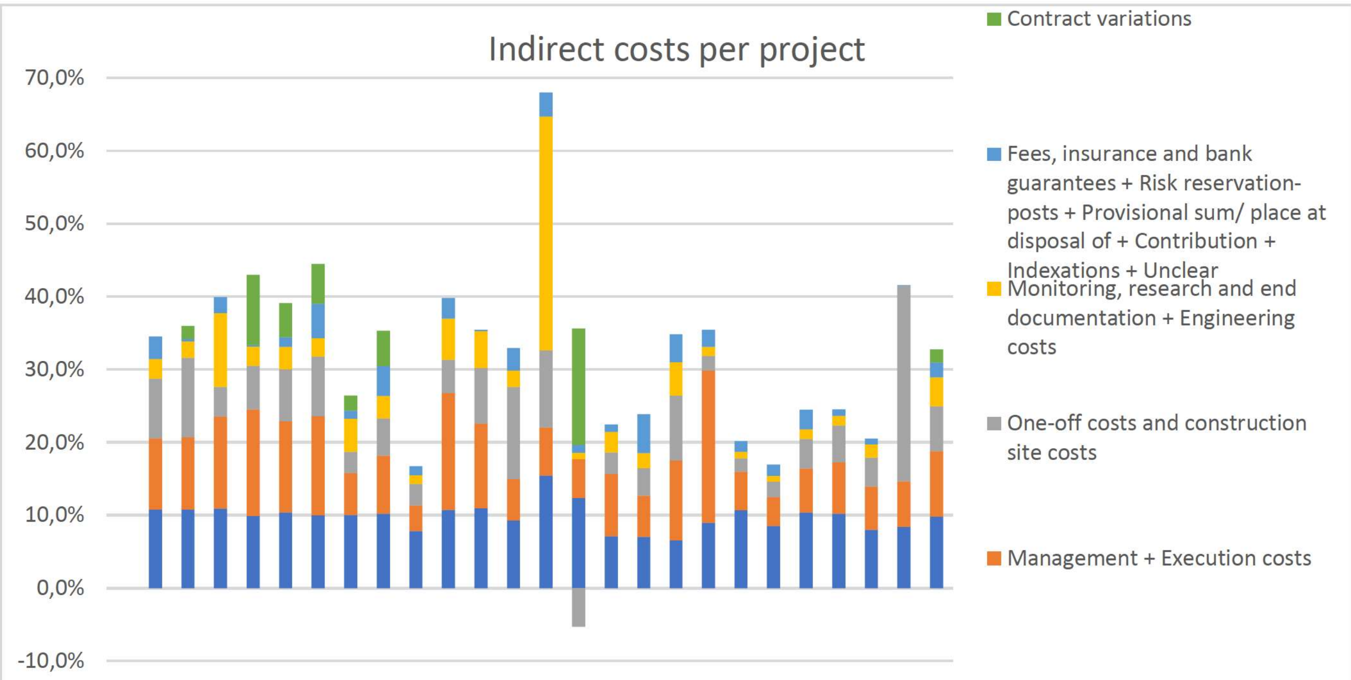


Figure 7 Indirect costs composition per project (anonymised)

4.3.3. Average project costs composition

It is useful to know what the main cost drivers of the project are. This can help understand what drives the costs in the projects. The largest cost categories can be found in the pie chart below, here the average of the 10 largest cost categories is given (Figure 9). The full distribution of costs can be seen in

Appendix F. It can be seen that on average the foundation is the largest cost, followed by overhead/profit/risk and management +execution costs. In total, these 3 costs are responsible for over 50% of the costs.

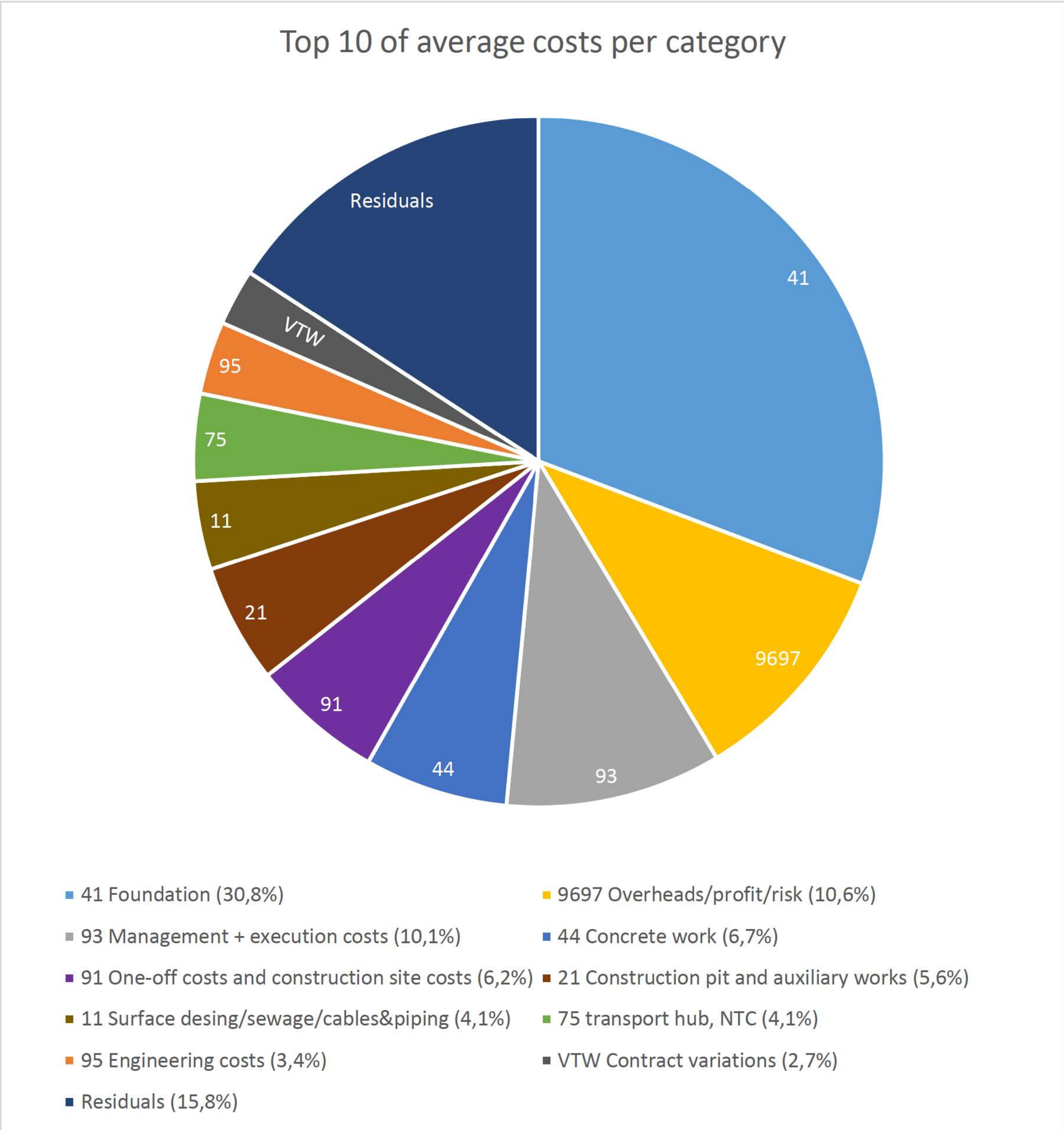


Figure 9 Top 10 average costs per category

4.3.4. Largest contributors to the cost differences

The largest contributors to the cost differences are useful to know when looking for CERs. When it is known what drives the differences in the costs, predictors related to these costs can be looked for.

To determine the largest contributors to the cost differences, there is looked at the standard deviation. The standard deviation is a measure of the variation around the mean, a lower standard deviation would imply that the values tend to be close to the mean, and a higher standard deviation would imply a higher spread of the values.

All costs discussed in this research are exclusive of VAT.

The 10 largest standard deviations of the cost categories are shown in the bar chart below (Figure 10). Here it can be seen that the foundation is the highest contributor with a standard deviation of [redacted], followed by the engineering costs [redacted], which has a low average so therefore surprising to see here. This could be the result of the [redacted] having such large engineering costs (4.3.2.). In the third place is the Management + execution costs [redacted]

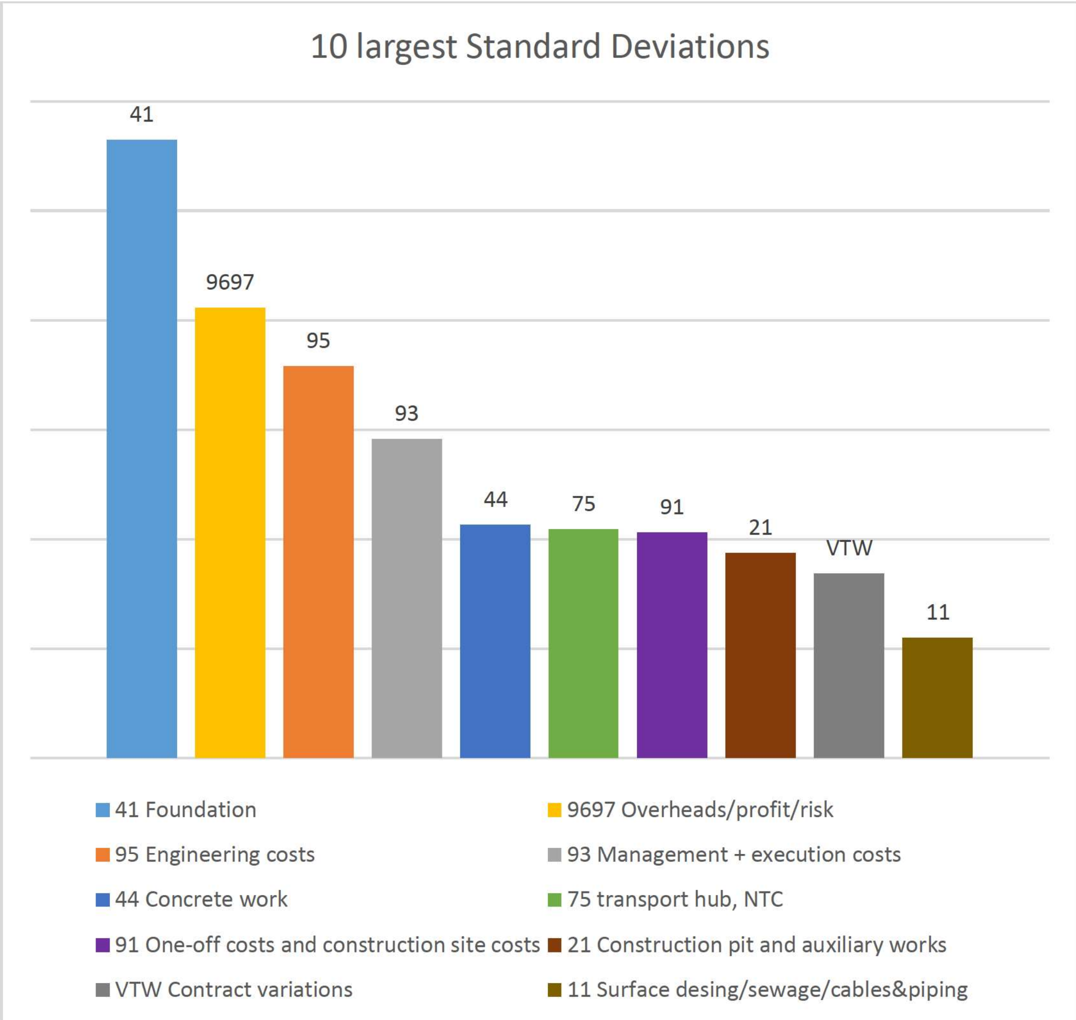


Figure 10 10 largest standard deviations of the costs categories (exclusive of VAT) (anonymised)

4.4. Resume

In this chapter suitable datasets are gathered, this data comes from the municipality of Amsterdam or the Witteveen+Bos (W+B) server. Moreover, a scope change has been made, instead of only using the actual costs incurred in the project, also other types of costs - of earlier project stages - are taken into account. This is done to increase the dataset significantly compared to the first research by W+B. However, this led to uncertainty in the data. On average the variation coefficient is 1,4%, this is rather small and is neglected in this research. There is been chosen to add 3 suitable projects from outside of Amsterdam. In total, there are 25 projects included in the dataset.

Furthermore, there is looked into what the project costs are composed of. Here the data is structured using a classification system. It has been found that there are large differences in the composition of costs between the projects. Furthermore, on average the foundation is responsible for 30,8% of the costs of the project, followed by the overheads/profit/risk (10,6%) and the management + execution costs (10,1%). Which together make up for over 50% of the costs. Moreover, the largest contributors to cost differences are foundation [redacted] engineering costs [redacted] and management + execution costs [redacted]

5. Cost Estimation Relationships Process

5.1. Introduction

This chapter discusses phase 2, which is the process of determining the Cost Estimation Relationships (CERs) for quay wall renovation, this is an iterative process. However, for the sake of the report, it is discussed in chronological order without switching back and forth. The process goes through all the steps of the CER process, the steps are shown below:

+ Cost estimating 'hypothesis'
+ Collection of relationship data
+ Evaluate and normalize data
+ Simple linear regression
+ Multiple linear regression
+ Validation

5.2. Cost estimating 'hypothesis'

In this paragraph, the cost estimation 'hypothesis' are determined. These 'hypotheses' should be seen as the predictors that possibly influence the costs.

The parameters which are expected to be a good predictor of the costs are discussed below:

1. **Length**. It is expected that the longer the length of a project is, the cheaper per meter it becomes. This is due to scaling benefits and the reduced starting costs.
2. **Construction type**. The type of construction has an impact on the construction process and materials. This could therefore influence the construction costs.
3. **Duration**. The longer a project takes to complete the more costs there are e.g. renting materials. Therefore, the duration of a project could influence the costs, e.g. shorter project can have smaller costs.
4. **Type of contract**. Currently, different types of contracts are in the project, which could cause differences in responsibilities and therefore in the costs.
5. **Construction pit**. The use of a construction pit requires the building of a temporary wall to keep water out. This causes large starting costs, however, it could reduce the construction costs of the quay wall itself.
6. **Material use**. This category focuses on the exterior finish of the quay wall since the foundation in all cases is steel. The materials of which the quay wall finishes can influence the costs.
7. **Retaining height**. A higher retaining height could contribute to higher costs. This is due to the added structural challenges.
8. **Retaining area**. This is the product of the Retaining height and the length of the quay wall. This could be a valuable predictor of the costs since it takes into account the total surface of the quay wall. The higher this surface area the higher the total costs of a project. Therefore, the combination of these predictors could be an interesting predictor.
9. **Foundation depth**. As can be seen in 4.3.4. , the foundation is the largest contributor to the costs of the projects. It could be possible that the depth of the foundation is of influence to the costs.
10. **Width canal**. The width of the canal can be of influence to the construction process. A wider canal may give the contractor more working space, and boats can more easily pass.

11. **Traffic**. The traffic that passes by the construction site can be of influence to the construction process. This can make the process more difficult, especially in the busy city centre of Amsterdam.
12. **Working from the water.** Working from the water could mean that there is more working space so it is easier to work and the costs can be reduced. Therefore it is expected that this influences the costs.
13. **Transport of materials over water**. The transport of materials over water could affect the costs, it is expected to be more expensive since materials need to be transported elsewhere probably by trucks, then reloaded onto the ships and then shipped to the construction site.
14. **Working vibration-free** (or reduced). This causes the use of other equipment furthermore the speed at which these machines work is reduced compared to traditional machines.

5.3. Collection of predictor data

In the last paragraph the possible predictors for the Cost Estimation Relationship are discussed, in this paragraph the data of these predictors is gathered. This data Is necessary for determining CERs.

The topics addressed in this paragraph are:

+ Data sources
+ Missing predictor data
+ Predictor data overview
+ Categorical data problem

5.3.1. Data sources

Multiple sources are used in gathering the data for the predictors. Which is discussed in the following paragraph. Also, some assumptions were made for the data to be collected.

These sources are documents available to Witteveen+Bos and online sources. Documents that are most available to Witteveen+Bos are: Design drawings, construction specifications, client estimate, detailed design estimate, contract sum and contract variations. Sometimes other relevant documents were available. Online sources were used for the following information:

- Traffic. No information was found on the use of the roads on the quays, therefore this could not be taken into account. However, there is information on the number of ships using the canals, this is included in the analysis. This information was achieved from Grachtenmonitor Gemeente Amsterdam (*2022*).
- Retaining height. Of the collected data 50% is taken from the documents, and the other 50% is found in the data from Waternet (*15-01-2024).*
- Canal width is found using the same source as the retaining height. However, missing values are estimated using the length calculator from Google Maps (*12-2023*).

Furthermore, some assumptions were made for the data to be gathered.

- The length of the quay wall - if not stated somewhere - can be assumed the same length as the number of meters of cover stones required in the construction specifications.
- The duration of the project - if not stated somewhere - can be assumed to be the duration of the temporary measures.
- The retaining height - the retaining height is measured from the bottom of the canal to the top of the quay wall.

- If the contract type is not stated in the description, then it is assumed that if the project was before January 2021 it is a RAW contract. Since is done since at this date the SOK kademakers program started (as stated in 2.2.2. SOK Kademakers) and the IPK contracts are all known.

5.3.2. Missing predictor data

Some data of the predictors which are discussed in 5.2. Cost estimating 'hypothesis', are not included in the datasheet for further analysis. This is due to the following reasons:

- The vibration-free technique of foundation construction has been used in almost every case, in some cases it was not entirely clear. However, since in all known cases, the technique used was vibration-free, this predictor is not included in the research.
- Working from the water, this was in most cases not clearly stated. Some documents stated that they indeed worked from the water, but due to much uncertainty, working from the water is not included.
- Transport materials have the problem that in a lot of cases where it is not known whether there was transport over the water or how much is transported over water. It was traceable in several cases in the client estimate or contract sum, here it is used in a separate analysis later in the report. However, it is not included in the search for CERs.
- Traffic on the quay walls is not taken into account since no relevant information could be found, as stated earlier in 5.3.1. Data sources. Therefore, only shipping traffic has been taken into account which was found.

5.3.3. Predictor data overview

The overview of the predictor data can be found in Appendix G

. It is important to note that empty cells are missing information. However, due to data scarcity, it has been chosen to use these projects and predictors wherever possible.

5.3.4. Categorical data problem

For the regression analysis to be performed, all values should be numerical. However, the predictors: construction type, material and contractor, use categorical data, which in this case are not numerical. However, there is a workaround (*Statistics solutions (a, n.d.)*.

According to Statistics Solutions *(a, n.d.)*, for the categorical data to be used in the regression analysis dummy variables are introduced. These dummy variables act as a binary value (1 or 0), being present or not present, which enables use for regression analysis. This is achieved in the following way.

In the case of the contract used for the quay wall project. There are 3 categories: SOK, IPK and RAW. These 3 categories were originally stored in one column however, dummy variables split the cells into more columns, creating something like Table 3 Dummy variable example.

Table 3 Dummy variable example

Project	Contract SOK	Contract IPK
A	1	0
B	0	0
C	0	1

In this way, when SOK is 1, the other categories are 0. When both SOK and IPK are 0, then this means that RAW is 1. For this reason, the number of dummy variables is the number of categories **-** 1. Otherwise, this would give the regression redundant information which would result in multicollinearity. Therefore, in Table 3, the RAW is not present as an independent variable.

However, due to there being 13 different contractors in the projects it has been chosen not to use them for the regression analysis. This is because most of the contractors only have one project and therefore pose the risk of overfitting.

Furthermore, it is not possible to use the 'contract type', 'construction type' and 'material' for simple linear regression, due to them being split into multiple independent variables. Therefore they are tested at the multiple linear regression.

5.4. Normalize data

Before the datasets can be compared the data needs to be normalized. This is done in the following paragraph. To find accurate CERs in the construction costs of the quay walls, the costs of side projects and the requirements of the different quay walls are separated. This allows for a more accurate comparison since the side projects and the requirements differ a lot per project location. Finally, the boundaries in which the model can be used are evaluated.

The following topics are addressed:

+ Normalize for inflation
+ Remove side projects and costs of requirements
+ Limits of model use

5.4.1. Normalize for inflation

The impact of the inflation has to be taken before the project can be compared, due to changes in the costs. How this index is determined can be found in Appendix G

Table 21 Normalised costs. The reference point is the index of Oktober 2023, which means that all the costs are calculated as the costs in October 2023. In the figure below the costs index over the years can be found, it can be seen that the costs have risen significantly, however last year the costs also slightly dropped again (Figure 11).

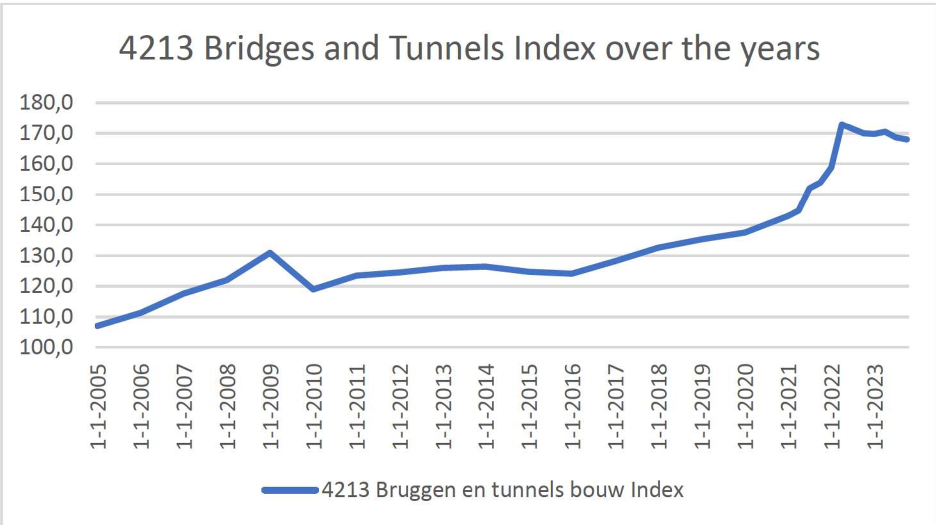


Figure 11 Normalised index for bridges and tunnels (CBS 2023)

5.4.2. Remove side projects and costs of requirements

As explained in 2.3.4. Other

The final category consists of the other projects. These are projects which have a steel sheet foundation and use other finishes like wood or stone covering stone.

2.4 Requirements of the Municipality of Amsterdam, the municipality of Amsterdam has several different requirements. These requirements can differ from project to project and can influence the costs of the project. Moreover, certain side projects are present in the cost data, such as resurfacing of the road or providing maintenance to bridges.
Since these requirements and side projects are additional to the construction of the quay wall they are separated from the dataset. This is done to enable a more clear comparison of the true construction costs of the quay wall projects. Moreover, the costs of the requirements are analysed separately in section 6.2. Costs of requirements.

The removed cost categories are:

- surface design, sewage, cables and piping (11)
- trees (14)
- Phasing BLVC-measures (71)
- Transport HUB & NTC costs (75)
- Side projects (77)

5.4.3. Limits of model use

In this paragraph, the ranges for which it is advised that the Cost Estimating Relationship can be used are discussed.

As explained in 3.2.2. Extrapolation, a CER model can only properly predict inside the range of the input. The range is found by looking at the extreme values of the different predictors. The maximum and minimum values are taken as the boundaries for which the model can be used. These values are given in Table 4.

Table 4 Advised limits of model use

	Length (m)	**(expected) duration (w)**	**Retaining Height (m)**	**Max depth foundation (m below NAP)**	**Width canal (m)**	**Average traffic per day**	**Retaining area (m^2)***	**Length/ duration (m/w)**
Min	44	20	1,6	9	7	1	145,2	1,21
Max	291	104	5,9	24	100	4	1718,2	9,93

* The limit is necessary since taking the product of the extreme length and retaining height would result in too extreme limits for the predictor retaining area. E.g. min length*min retaining height = 70,4 m^2.

5.5. Simple linear regression

In this phase, possible Cost Estimation Relationships (CERs) are determined, by use of regression analysis between the predictors and the costs of quay wall renovations. This is done by looking at scatterplots of the data. Since Excel can perform a regression analysis together with constructing a scatterplot this step is combined with the regression analysis for simple linear regression.

+ Assumptions
+ Sidenotes
+ Results

5.5.1. Assumptions

For the regression analysis to take place, certain assumptions are made (NASA, 2015):

1. There exists a linear relationship between the X (dependent) and Y (independent) so that the independent variable(s) can sufficiently explain the variation in the dependent variable.
2. Independent variables have no error.
3. The residuals (error terms) are independent, normally distributed multivariate random variables with mean zero and constant variance. (This is checked in 5.8.1. Normal probability plot)
4. The error terms are not correlated (this is checked in 5.8.2. Residual plot)
5. The data is homogeneous, which means that the data is from the same type of product
6. Independent variables are independent of each other (This is only in case of multiple linear regression, which is checked in 5.8.3. Testing for multicollinearity)

5.5.2. Sidenotes

Before starting on the simple linear regression it is important to note that some predictor data is missing, which results in holes in the dataset. However, due to the lack of data, it has been chosen to include always as much data as possible, therefore for each case, the maximum amount of data is used. This means that for each predictor that is tested, only the rows where that predictor is missing are removed. This results in a different dataset depending on which independent variables are tested. As discussed in paragraph 5.4.2. Remove side projects, the dependent variables are the costs directly related to the construction of the quay wall itself.

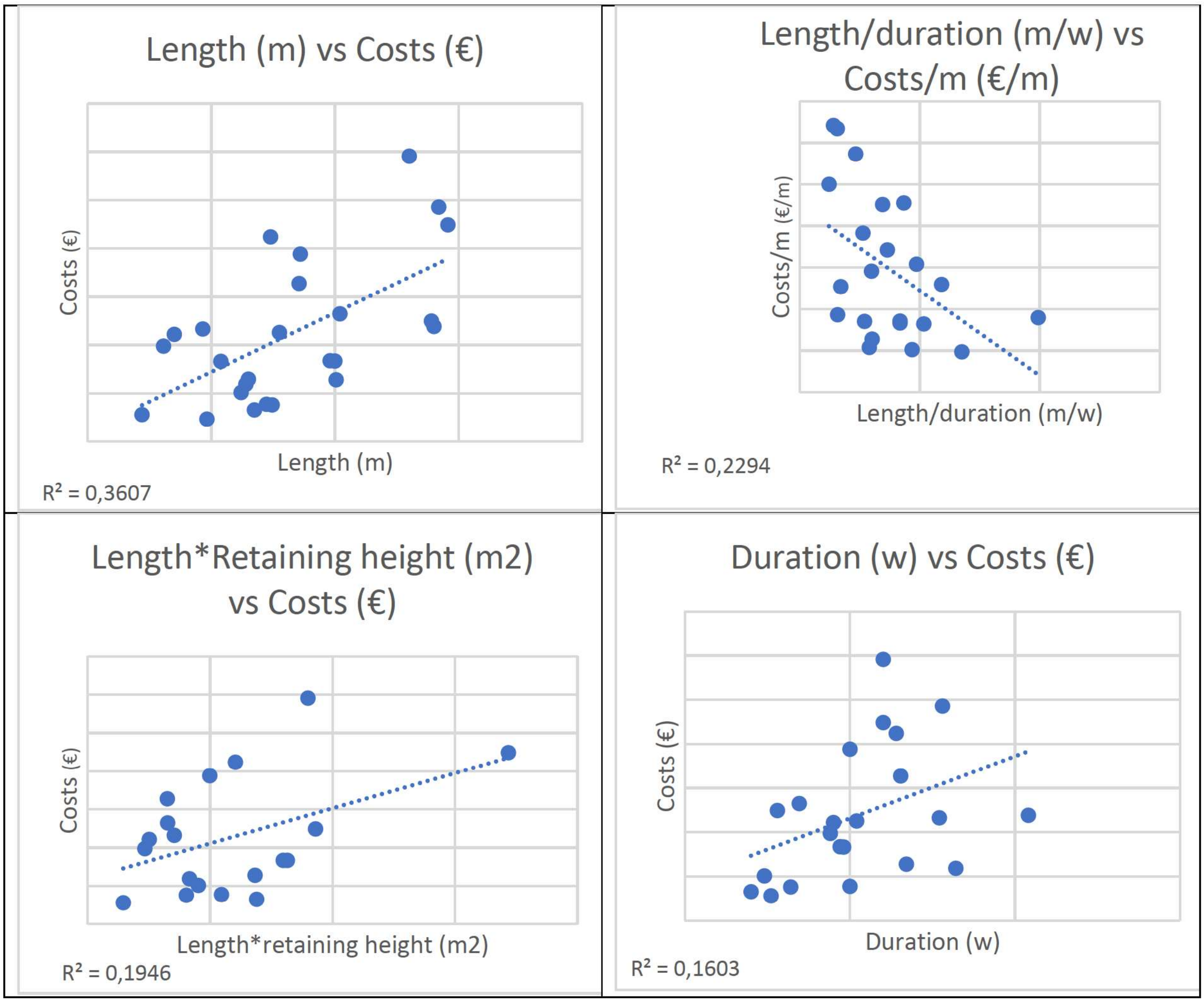
The tests are performed on both the costs per meter and the total costs.

5.5.3. Results

The results of the simple linear regression analysis can be found below. The 4 best results (based on the coefficient of determination, R^2) are shown in Table 5 Scatterplots and fitted Simple Linear Regression , where the R^2 coefficient can be found at the left bottom. The found R^2 coefficients are not too good since the best predictor – the length - only explains the difference in costs for 36%. This means that 64% is left up to chance, which makes this not a good predictor.

The other results can be found in Appendix I Table 22. The R^2 coefficients of the predictors: Construction pit, Width canal, Retaining height, average traffic per day, and max depth foundation, were all underneath 0,12. It is clear that some predictors that were discussed in 5.2. Cost estimating 'hypothesis', are not good at all in explaining the costs.

Table 5 Scatterplots and fitted Simple Linear Regression line of 4 best results (anonymised)



5.6. Simple nonlinear regression

It might be that a better relation can be found using nonlinear regression. Therefore, a visual inspection is done to the scatterplots to see whether a nonlinear pattern exists, where this pattern is spotted the nonlinear regression is fitted.

From the most promising correlations of the last paragraph, there is one graph found where the pattern looked nonlinear. In the costs/m vs the length/duration, a regression in the power form is found to fit better. With the power function, the coefficient of determination improved from 0,23 to 0,33. Making it the best predictor for the costs/m. The graph with the power regression fitted can be found below in Figure 12.

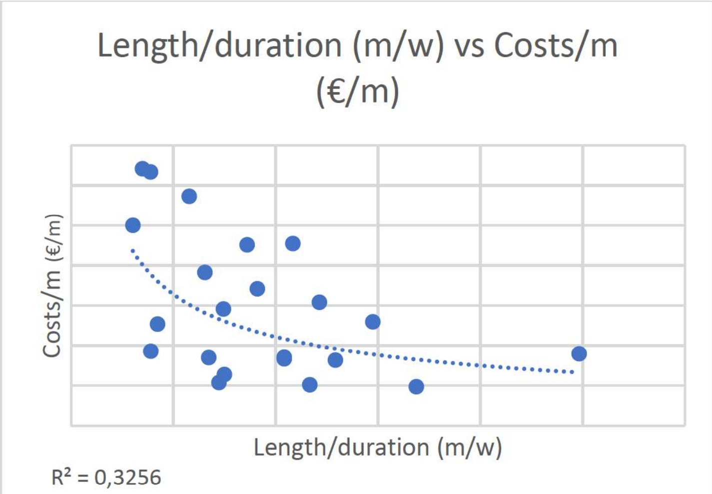


Figure 12 Scatterplot of costs/m vs length/duration and fitted simple nonlinear regression (anonymised)

5.7. Multiple linear regression

There are a lot of situations which require two or more independent variables to accurately describe the process, this is done in multiple linear regression analysis. Furthermore, multiple linear regression analysis also helps to find the precise effects of predictors, what happens when one predictor changes but the other predictor is fixed. In this chapter, the multiple linear regression analysis is performed. The following steps are discussed:

+ Adjusted coefficient of determination
+ Categorical predictors results
+ Maximum number of independent variables
+ Multiple linear regression Results
+ Formula check

5.7.1. Categorical predictors results

As discussed in 5.3.4. Categorical data problem, 3 predictors had to be tested in multiple regression analysis due to the use of dummy variables. The results are shown in Appendix I Table 23. Here it is seen that the contract has a very good coefficient of multiple determination (R_a^2) compared to the other independent variables. The contract to the costs: R_a^2 = 0,55 and the costs/m: R_a^2 = 0,70. These are large improvements compared to the simple linear regression analysis outcomes, where the best R^2 value was the length to the costs with an R^2 of 0,36.

Furthermore, the results of the multiple regression analysis for the Construction type and the Material used are insignificant with their highest R_a^2 being 0,02.

5.7.2. Maximum number of independent variables

As discussed in 3.5. Phase 2: Cost Estimation Relationships Process, the rule of thumb is to use at least 10 observations per independent variable. Therefore, in the total data set of the research, there are 25 observations, so two independent variables can be present according to the rule of thumb.

The categorical predictors are split into multiple independent variables because of to the dummy variables, so one predictor has multiple independent variables. However, there is only one category present each time, and they all represent one string of data, they are treated as one independent variable for the above-mentioned rule. This enables the multiple regression to be tested for a categorical predictor in combination with another predictor.

As explained in paragraph 5.3.2. Missing predictor data, there are gaps in the predictor data. However, due to data scarcity, it has been chosen to use as many datasets wherever possible. This creates a difference in the number of independent variables per predictor combination. The overview of the number of independent variables per combination of predictors is given in Table 6.

Table 6 Number of data points per predictor combination

Nr of data points	1	2	3	4	5	6	7	8	9
1. Length (m)									
2. Duration (w)	22								
3. Construction type	23	22							
4. Lenght/Duration (m/w)	22	22	22						
5. Material	23	22	23	22					
6. Witdh canal (m)	25	22	23	22	23				
7. Max depth foundation (m below NAP)	17	16	17	16	17	17			
8. Length*Retaining height (m2)	19	19	19	19	19	19	15		
9. Contract type	25	22	23	22	23	25	17	19	
Retaining height	19		19						19
Construction pit	25								

Furthermore, the number of independent variables changes because of the different categorical predictors, an overview of this can be found in Appendix J Table 24.

In some cases due to the gaps in the dataset, the number of data points available is slightly lower than 20. Nevertheless, it is chosen to test these 2 predictors. The rule is not a strict rule but a rule of thumb. Furthermore, good relationships are tested in 5.8. Assumptions and validation. So if a good relationship was established which has less than 20 data points, the results were still validated and discussed.

5.7.3. Results of multiple regression analysis

The different combinations are tested for two different dependent variables, for both the total costs and the costs per meter. Two CER models are chosen for further analysis, the best for the total costs and the best for the costs per meter.

The results of the multiple regression analysis on the costs can be found in Table 7 and the multiple regression results for the costs per meter can be found in Table 8, the intersection of the column and row (the predictors) is the coefficient of multiple determination. Here it can be seen that the contract type is in both cases a strong predictor in combination with other predictors. This was expected since the contract is also the strongest single predictor (5.7.1. Categorical predictors results).

Table 7 Results of multiple regression analysis on the costs

Costs (R^2 adjusted)	1	2	3	4	5	6	7	8	9
1. Length (m)									
2. Duration (w)	0,301								
3. Construction type	0,314	0,096							
4. Lenght/Duration (m/w)	0,363	0,157	0,046						
5. Material	0,300	0,055	-0,091	0,001					
6. Witdh canal (m)	0,310	0,248	0,222	-0,010	0,142				
7. Max depth foundation (m below NAP)	0,290	0,040	0,031	-0,078	-0,128	0,198			
8. Length*Retaining height (m2)	0,199	0,231	0,331	0,137	0,174	0,094	0,287		
9. Contract type	0,8356	0,55392	0,44736	0,56901	0,50414	0,57231	0,48799	0,75109	
Retaining height	0,190								0,543883
Construction pit	0,303								

Table 8 Results of multiple regression analysis on the costs per meter

Costs/m (R^2 adjusted)	1	2	3	4	5	6	7	8	9
1. Length (m)									
2. Duration (w)	0,095								
3. Construction type	0,031	-0,036							
4. m/w	0,151	0,154	0,141						
5. Material	0,030	-0,049	-0,134	0,126					
6. Witdh canal (m)	-0,015	-0,068	-0,047	0,228	-0,046				
7. Max depth foundation (m below NAP)	0,060	0,036	-0,042	0,193	-0,054	0,012			
8. Length*Retaining height (m2)	-0,022	0,080	0,020	0,167	-0,006	-0,021	0,040		
9. Contract type	0,79939	0,70739	0,6492	0,77103	0,70827	0,7153	0,76731	0,88102	
Retaining height	-0,0234								0,689882

The best regression model of the costs and the best regression model of the costs per meter are chosen for further analysis as a CER. The formulas are calculated by Excel using the regression analysis function of the data analysis tool pack in Excel. The results are discussed below.

CER model 1 (CER 1): Contract type and Length to the costs

The highest impact on the dependent variable 'costs' was the predictor combination of 'Contract type' and 'Length', this combination resulted in an adjusted R^2 value of 0,836. This means that the predictors 'contract type' and 'length' can explain 84% of the differences in the dependent variable 'costs', leaving 16% up to chance.

The equation is stated below:

$$\boldsymbol{Predicted\ price}\ (€) = -\mathbf{667.690} + \text{[redacted]} * \boldsymbol{X_1} + \text{[redacted]} * \boldsymbol{X_2} + \text{[redacted]} * \boldsymbol{X_3}$$

X_1 = Contract IPK
X_2 = Contract SOK
X_3 = Length (m)

Here it is important to note that the IPK and SOK can both only be 1 or 0, as explained in 5.3.4. Categorical data problem. Furthermore, they refer to the RAW contract type. So the IPK Program is [redacted] euros more expensive than the RAW contract for a project of the same length. The same accounts for the SOK, which is [redacted] euros more expensive compared to the RAW contract.

Keep in mind that in the case IPK and SOK would be 0, (RAW contract is used), the standard costs would be -667.690 euro plus [redacted] euro per meter of quay wall to be constructed. When only 10 meters are constructed, the costs would be negative. This is of course not possible, therefore there are the model limits as discussed in 5.4.3. Limits.

CER model 2 (CER 2): Contract type and Retaining area to the costs per meter

The best predictor combination for the dependent variable 'costs/m' is the combination of 'contract type' and 'retaining area' which explain 88% of the differences in the costs/m.

The equation is stated below.

$$\boldsymbol{Predicted\ price/m}\ (€/\mathrm{m}) = \mathbf{25.187} + \text{[redacted]} * \boldsymbol{X_1} + \text{[redacted]} * \boldsymbol{X_2} - \text{[redacted]} * \boldsymbol{X_4}$$

X_1 = Contract IPK
X_2 = Contract SOK
X_4 = Retaining area (m^2)

As discussed in 5.7.2. Maximum number of independent variables, in some cases the number of data points is less than 20, here there are 19. However, the results are validated in 5.8. Assumptions and validation.

According to CER model 2, the IPK contract is [redacted] euros more expensive per meter than the RAW contract for a project of the same length. The same accounts for the SOK contract is [redacted] euro more expensive than the RAW contract.

5.7.4. Formula check

Furthermore, the behaviour of the CER models is checked in Appendix K. Here it becomes clear that the limits should be set per contract type, these new limits can be found in Appendix K Table 26.

CER Model 1 (Contract type and Lenth to the costs) is constructed to test the costs, however is also analysed for the limits in costs/m (Appendix L Figure 19). Here it is seen that for the RAW contract, the costs/m increase when the length of the project increases, however, this is expected to drop due to scaling benefits and reduced starting costs. This is compared with the RAW input data and here there is no clear relationship found between the length and the costs (Appendix L Figure 20).

5.8. Assumptions and validation

In this paragraph, first, the assumptions done in 5.5.1. Assumptions are checked. Then the relationships are tested (validation) since there are more relevant statistics than the coefficient of determination. The first statistic is already tested (R^2), but more statistics tell something about the relationship. Using the NASA methodology, also the Coefficient of Variation is checked, and the hypothesis test.

The following is discussed:

+ Normal probability plot
+ Residual plot
+ Testing for multicollinearity
+ Coefficient of variation (CV)
+ Hypothesis test

5.8.1. Normal probability plot

According to NASA (*NASA, 2015*), the normal probability plot is used to test the assumption: the error terms are independent and normally distributed. This is tested by making a plot of the predicted costs and the actual costs. Where if the lines are distributed it can be assumed that the error terms are normally distributed.
The results can be found in Appendix M Figure 21 and Figure 22. Here it can be seen that the lines regression line is 45 degrees in both cases. Therefore the assumption of normally distributed error terms seems reasonable.

5.8.2. Residual plot

The following assumption is made: The error terms are not correlated. To check the assumption, residual plots are made. Which can be seen in Appendix N Figure 23 Residual plots CER model 1 and Figure 24 Residual plots of CER model 2.

"A random scatter around the zero line indicates that the errors are independent and random" (*NASA, 2015*). Since in the residual plots it can be seen that the points are randomly scattered around the zero line it seems reasonable to assume that the error terms are not correlated.

5.8.3. Testing for multicollinearity

For the assumption of multiple linear regression, multicollinearity needs to be checked.
According to NASA (*NASA, 2015*), multicollinearity causes two problems:

- The estimated value can have large variations depending on which parameters are in the model (independent variables).
- The regression coefficient can have extremely large standard deviations

This is tested with a correlation test. Where for any pair of variables, the |R| value above 0,7, then there is a relationship between the independent variables that is likely to affect the regression parameters. Therefore the outcome should be rejected.

The correlation test is performed on CER models 1 and 2. The results can be found in Appendix O, Table 27 and Table 28. All values are below 0,7 and therefore there is no reason to reject for multicollinearity.

5.8.4. Coefficient of Variation (CV)

According to NASA (NASA, 2015) The coefficient of variation allows to use of the Standard Error of the Estimate to evaluate a regression line. The CV should be below 20% and preferably below 10%. The CER with the lowest CV is preferred, as long as the CERs are logical, the slopes are significant (hypothesis test) and there is a sufficient coefficient of determination (R^2).

The Coefficient of Variation is calculated using the following formula $CV = \frac{SE}{\bar{Y}}$. So the Standard Error of the Estimate (SE) divided by the mean of the dependent variable. The CV is therefore a measure to indicate the relative error of the estimate. This relative error of the estimate can be used to compare different models.
The SE is given by Excel when performing the multiple regression analysis. Therefore, the Coefficient of variation can be calculated (Figure 13 Standard Error of the Estimate and coefficient of Variation).

	Standard Error of the Estimate(SE)	Average (per model)	Coefficient of Variation
CER 1	[redacted]	[redacted]	26,6%
CER 2	[redacted]	[redacted]	21,2%

Figure 13 Standard Error of the Estimate and coefficient of Variation (anonymised)

The coefficient of variation should be less or equal to 20% for regression analysis, however, due to the absence of other estimators this variation for now is accepted.

5.8.5. Hypothesis test

According to NASA (NASA, 2015), after the first tests it might be assumed that the equation is a valid predictive device. However, there can still be an error in the equation. The predictions might not be precise due to sampling error.

Therefore, the hypothesis test is used to determine the 'strength' of the regression coefficients.
This test determines whether the coefficient of the independent variable is significantly different from zero. In the case of multiple regression analysis, the F-test is used to see if the regression relationship between the dependent and independent variables is statistically significant. A slope of zero would imply that the relationship is pure chance, and a slope not significantly different from zero would imply that knowledge of value X would be of no use to predict the value of Y.

The hypothesis test can be found in Appendix P. All slopes are significantly different from zero. Both for the complete model and for the individual slopes. This means that the effect found is large enough to rule out the chance that it is caused by sampling error.

5.9. Select CERs

The most interesting CERs have been found in the multiple regression analysis. Here there are two CERs found, one for the dependent variable costs and the other for the dependent variable costs/m. Then the assumptions were checked, and the validation was done.

This gave the following results:

- CER 1 uses the contract type and the length of a project as predictors of the total project costs. It can explain 83,6% of the difference in costs and the relative error of the estimate is 26,6%.
- CER 2 uses the contract type and the product of the length and retaining height as predictors of the project costs per meter. It can explain 88,1% of the difference in costs/m and the relative error of the estimate is 21,2%.

Furthermore, the effect of the slopes in both models is large enough to rule out the chance that it is caused by sampling error.

Based on the analysis it is advised that CER model 2 is used, since it can explain more of the difference in costs and the coefficient of variation is lower.

Furthermore, the CERs that are determined and validated in this chapter are excluding VAT and excluding: the costs of surface design, trees, phasing BLVC-measures, Transport HUB & NTC cost and side projects.

5.10. Resume

First, the cost possible predictors are discussed, then the sources of the predictor data, some assumptions are made for the data to be gathered and some possible predictor data is not found and is therefore missing. The categorical data problem is explained and the solution to use dummy variables is discussed. Then the data is normalized for inflation. Also, side projects and the costs of requirements are removed, this is done to enable a better comparison of the construction costs. Furthermore, the limits between the CER model can be used are shown.

The regression analysis has been performed for simple linear regression, simple nonlinear regression and multiple linear regression. The best predictor for the simple linear regression is the Contract type, which explains 55% of the difference in the total costs and 70% of the difference in the costs/m. For the multiple regression, the best results were: The contract type and the length to the costs, which explained 84% of the difference in the costs. And the contract type and the retaining area to the costs/m which, explained 88% of the difference in the costs/m.

Then these results were validated and their assumptions checked.

6. Costs of requirements and advice

6.1. Introduction

This chapter discusses phase 3 and consists of three parts, the first part is the analysis of the costs of the additional requirements set by the municipality of Amsterdam. In the second part advice based on the findings of the research is given. The third part of the chapter is about advising the data gathering for further research.

This chapter consists of the following parts:

+ Additional requirements
+ Advice based on the findings in the analysis
+ Advice on data gathering

6.2. Costs of requirements

In this paragraph, the costs of the requirements set by the municipality of Amsterdam for the quay wall renovation projects are analysed.

Not all requirements in the quay wall renovation project are additional, some requirements are required by plans or (local) regulations. Nevertheless, they are discussed in this chapter. To understand what these requirements cost to implement.

There is no overview of which requirements are present at each location. Therefore, the analysis is based on the requirement costs found In the project data.

In a lot of projects, there was work done on the surface of the quay, mostly these were works on the resurfacing of the roads or different changes on the surface. However, to keep the scope of the requirements clear for the quay wall construction. It has been chosen to only analyse the requirement costs which are directly related to the quay wall construction and not the surface works. The following requirements are discussed:

- Trees
- BLVC-measures
- Transport HUB & NTC
- Houseboats

6.2.1 Trees

Within the trees, there are 4 categories on which actions are performed with the trees. These are:

- Retaining trees IPK. These (larger) trees are retained, and this is within the IPK program.
- Replacing trees. This is the full cost of removing the tree, planting a new one and improving the growing spot for trees.
- Protection of trees. This is the category for small trees, which can be protected during construction and where more easily be worked around.
- Costs of improving the growing spot of trees.

If a monumental tree had led to a stiffer (heavier) quay wall structure, in the RAW and SOK contracts the additional costs for such a tree are unknown because the extra costs are 'hidden' in substructure and structural costs items (*W+B supervisor, 2024*). In the IPK program, however, these costs are quantified and are therefore labelled as 'retaining trees IPK'.

The results are found below in Table 9, the costs are all given per tree.

Table 9 Costs of requirements related to trees (anonymised)

Category	Average	Standard deviation
Retaining trees IPK	[redacted]	[redacted]
Replacing trees	[redacted]	[redacted]
Protection of trees	[redacted]	[redacted]
Improving growing spot	[redacted]	[redacted]

6.2.2. BLVC measures

As discussed in 2.4 Requirements of the Municipality of Amsterdam , the BLVC measures are measures to minimize hinder and nuisance in the city. However, not all BLVC measures are placed in this category, some of them are expected to be in the management costs or other categories. However, there can be done a brief analysis of the shipping measures and the costs of closing the streets, traffic controllers, traffic measures and plates for pedestrians to walk or cars to drive over.

The costs are divided into costs per week and costs per meter. Here it is chosen to place them in the category where the ratio of the standard deviation divided by the average is the smallest. The costs per week can be found in Table 10 and the costs per meter in Table 11.

Table 10 BLVC measures costs per week (anonymised)

Costs per week	Average	Standard deviation
Fencing	[redacted]	[redacted]
Walking- and driving plates	[redacted]	[redacted]
Shipping guidance	[redacted]	[redacted]
Standby boat costs	[redacted]	[redacted]

Table 11 BLVC measures costs per meter (anonymised)

Costs per meter	Average	Standard deviation
Shipping signs	[redacted]	[redacted]
Matrix shipping signs	[redacted]	[redacted]
Shipping measures	[redacted]	[redacted]
Traffic measures	[redacted]	[redacted]
Traffic controllers	[redacted]	[redacted]

6.2.3. Transport HUB & NTC

This category includes the requirement of transporting materials over water. The costs of the transport HUB & NTC (Wet Transport Cycle) are very large.

The costs of this category largely differ depending on the project, many projects use different ways of stating these costs. However, in the newer contract types (SOK and IPK), a few projects used the same format. Therefore this format is used to calculate the average costs of this requirement.

The costs of the transport HUB & NTC are expressed as both the costs per week and the costs per meter. This is done since the costs can better be determined as the cost per week since the ratio of the standard deviation/average is lower. However, to give a sense of what it costs per meter of quay wall, the costs per meter is also added.

The costs of the transport HUB & NTC are shown in the table below.

Table 12 Average costs per week of transport HUB & NTC (anonymised)

Transport HUB & NTC costs	Average	Standard deviation
Costs/w	[redacted]	[redacted]
Costs/m	[redacted]	[redacted]

Furthermore, a pattern appears, when more meters per week are constructed the costs per meter of the quay wall drop. This is visible in Figure 14.

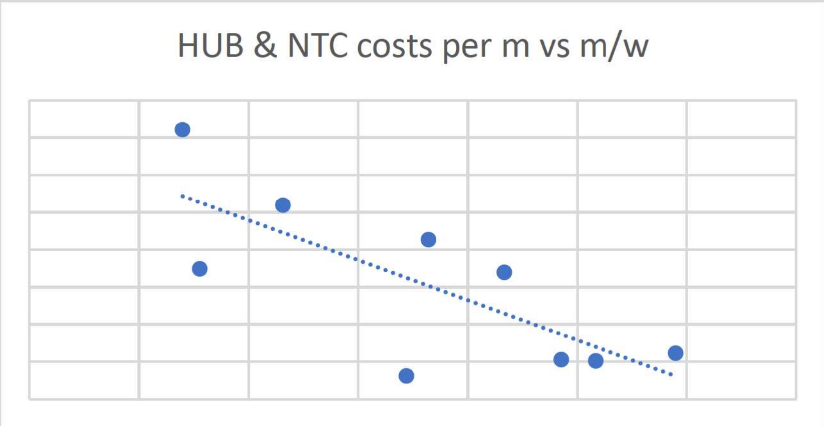


Figure 14 Scatterplot and linear regression line of Transport HUB & NTC costs/m vs m/w (anonymised)

6.2.4. Houseboats

Very little is known about the houseboats. It seems that they are largely not included in the costs of the project. Since only one case of moving a houseboat is known in [redacted] Here the moving of a single houseboat costs [redacted] euros. However, the costs of moving one houseboat is not representative of the costs for moving multiple houseboats, due to the startup costs being present in this one houseboat.

6.2.5. Over the years

What is interesting to see is an increase over the years in the costs per meter of the BLVC-measures, Trees and Transport HUB & NTC (Figure 15). The introduction of the new weight limit of vehicles in the centre of Amsterdam in 2021 can have contributed to this increase (explained in 2.4 Requirements of the Municipality of Amsterdam).

Furthermore, the introduction of the new contract types could have added to this increase since it looks like there is a correlation between the SOK contract and the increase in costs/m for these measures. [redacted] remains rather low compared to the other SOK contracts, but as seen in Table 16 this is the only detailed design estimate present. Therefore, it could be that details are missing in this detailed design estimate.

Figure 15 Scatterplot of Costs of BLVC, Trees and Transport HUB & NTC (per meter) over the years (anonymised)

6.3. Advice based on the findings in the analysis

In this paragraph, advice is given to the municipality of Amsterdam, based on the findings in this research. Advice is given about:

+ Database
+ Evaluation
+ Contract type
+ Further advice

6.3.1. Database

The municipality of Amsterdam first asked Witteveen+Bos to research the costs of the quay wall renovations. This research started for Witteveen+Bos by gathering data for the research, this process consisted of contacting multiple employees within the municipality of Amsterdam for project data. Here W+B gathered the necessary information to research these projects.

Furthermore, at this time it was unknown that an employee of the Ingenieursbureau of the municipality of Amsterdam was also conducting research into the costs of the quay walls. Later on, this research was shared with W+B. During this research, these databases are compared to see whether information from the research of this employee could be added to this database. This comparison gave the following results:

- 15 projects, that were present in this database were not in Amsterdam's database. Of which 12 projects were in Amsterdam
- 2 projects, were included in this database however Amsterdam did not use them since they were missing data.
- 6 projects where in both databases
- 2 projects, were in the report of Amsterdam, but not present in this database.

Before comparing this database had 23 usable projects, whilst the research from Amsterdam had 8 usable projects. However, it is not clear what the scope of the study of Amsterdam was. After the comparison, this database had 25 projects.

Peculiarly, a company outside of the municipality of Amsterdam systems can gather more data on projects that the municipality of Amsterdam has given orders for and paid for. It looks like Witteveen+Bos is better at storing project data, therefore making it possible to create a larger dataset in this research.
Furthermore, the municipality of Amsterdam asked Witteveen+Bos to perform research on the costs. However, the data then needed to be gathered. As explained earlier, the data needed to be gathered in many different places within the municipality. This implies that the municipality of Amsterdam does not properly store this data in a central location. The data of these projects should be sufficiently stored to perform analysis in the future and learn from projects.

Right now, it looks like the information is stored, but not in a clear central database. Since the research of the ingenieursbureau of Amsterdam collected not half of the data that was done for this research. Especially in the case of construction projects which quickly cost millions of euros this is disorderly. Therefore it is advised to gather data from quay wall renovations and store them on a central database.
However, it should be added that it is not clear what the scope was of the research of this employee. This could explain why there were fewer projects in that database, however, he still had 2 incomplete data of projects where Witteveen+Bos had gathered the complete set of data.

6.3.2. Evaluation

In the first part of the research on the costs of the quay walls performed by Witteveen+Bos, part of the research was interviewing employees who worked on the projects. These interviews resulted in new insights into why certain costs were high in projects.

It is recommended that for further research all the projects are evaluated on how the costs can be, what can be done more efficiently and other possible improvements. Since there is still a lot of work to be done this is especially helpful in the early stages, since then different design choices can be made to save costs and speed up the process.

6.3.3. Contract type

As discussed in chapter 5. Cost Estimation Relationships Process, the contract type is the best predictor, this means that this predictor is the best at explaining the differences in the costs. In this paragraph, the differences in the costs of the contracts are provided with possible explanations and advice based on the contract type is provided.

First, the costs per meter and the year are plotted with the contract type. It can be seen that in recent years new contracts have taken over and with this change also the costs per meter and the spread in the costs increased drastically (Figure 16). Other influences which could have contributed to cost differences have already been removed in 5.4.2. Remove side projects**.**

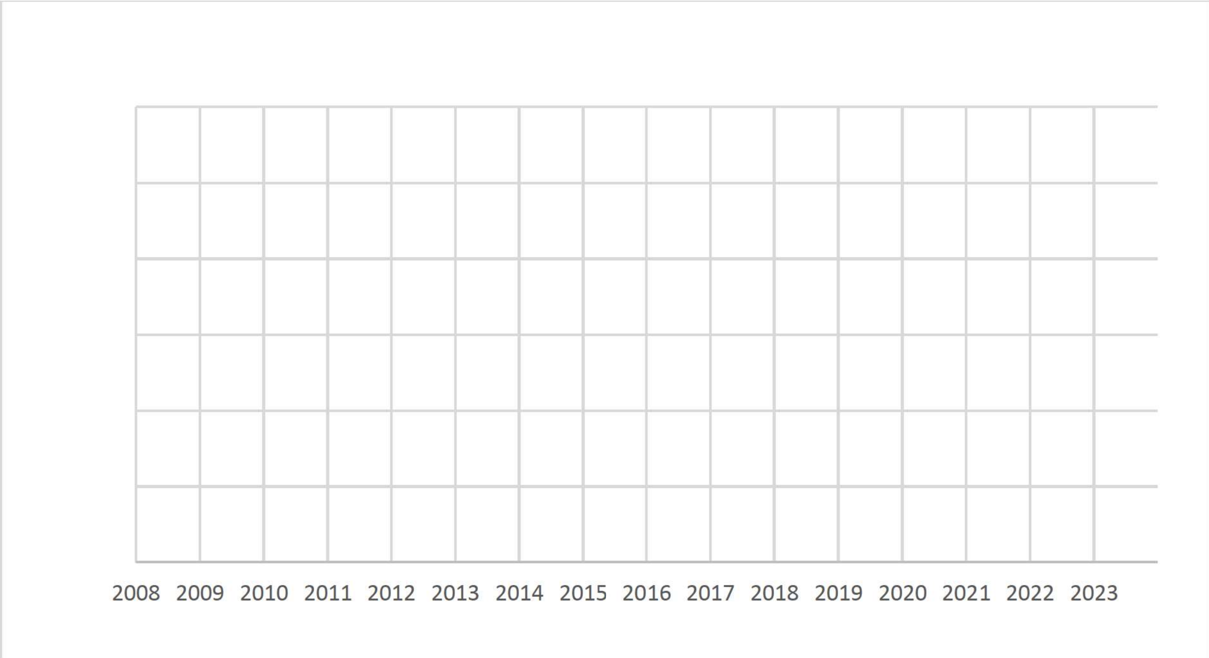


Figure 16 Scatterplot of costs per meter over the years, by contract type and their averages (anonymised)

Possible explanations for the differences in costs are provided.

RAW

The details of the RAW contract have been discussed in 2.2.1. RAW contract. In this contract the client has the responsibility for the design and the contractor builds exactly this design. Furthermore, there is free market tendering, which means that there is competition to win the contract. Which can be won according to the EMVI-tendering legislation (2.2.4. Change in tendering method), based on the costs but also added value. Here, the contractors have to bid realistically since there is competition.
On average the RAW costs [redacted] euro/m.

SOK

The SOK has many similarities with the RAW contract, the key difference is that there is no competition for each project to win the project. As discussed in 2.2.2. SOK Kademakers, the SOK contract has 3 contractors signed for 6 years, during this period there are no other competitors. The idea is that: having fixed contractors enables quicker project execution since the tendering phase, which can take up to 6 months, can be skipped. Furthermore, the contractors can sit at the design table and already in an early phase, disputes can be solved.

However, there are downsides, according to EU legislation: *"Free competition is a key element of an open market economy. It stimulates economic performance and offers consumers a broader choice of better-quality products and services at more competitive prices." (EUR-Lex, n.d.)*

The lack of free competition can make the prices higher, since there is no competition contractors can ask for higher prices overall. The municipality of Amsterdam tries to counter this with the predetermined rate list, and fixed percentage profit and risk. Furthermore, since the contractor sits at the design table, he could influence the list of works to be done. Here the contractor could try to reduce the risks in the project in a way that results in more work to be done. Since the contractor can calculate a margin over the work it executes and due to the reduced risk, this is profitable in two ways for the contractor. This extra work increases the costs, whilst it may have been cheaper to accept certain risks.

Free competition would make the contractor's budget lower and try to reduce the costs. While in this case there is no incentive for the contractor to try and minimize the costs. Therefore, it is important in this contract type to check if it is necessary what the contractor is doing and whether the contractor is doing what it is saying.
On average, the SOK contract costs [redacted] euro/m. Moreover, according to CER model 2, when the impact of the retaining area is removed, the SOK contract is [redacted] euros per meter of quay wall more expensive than the RAW contract type.

IPK

As explained in 2.2.3. IPK program, the IPK program uses newly designed construction methods. These methods are intended to work faster and cheaper than traditional methods. However, to realize these new innovative techniques, investments were made. These investments are in the costs of the quay wall. When the investments are being earned back over time, it is expected that the costs will decrease.

In the IPK program, in case a fixed budget is assigned. The contractor is motivated to work more efficiently, work faster and use less material, in this way their profits will rise. This is possible because these contractors are working with their design and methods, therefore they can keep innovating and improving their efficiency.

The costs of the IPK program may seem high in the current state, however some things should be taken into account. First, in the IPK program, the engineering of the quay wall is in the costs of the quay wall, while in the RAW and SOK, the costs of engineering are separate for the designer. Therefore, to make the comparison more fair the engineering costs should be added to the costs for the SOK and RAW contracts.

On average, the first 3 IPK projects cost [redacted] euros per meter. Moreover, according to CER model 2, when the impact of the retaining area is removed, the IPK contract is [redacted] euros per meter quay wall more expensive than the RAW contract type.

Advice on contract type

It is advised to abandon the SOK contract, despite the benefits of the faster process. When the impact of the retaining height is removed, the SOK is [redacted] euros per meter more expensive than the IPK program, even [redacted] euros per meter more expensive than the RAW contract. Furthermore, if this method is still chosen, check whether the contractor says what it is doing and check whether the changes opposed by the contractor are really necessary.

It is advised to use the IPK and RAW contracts, over time the costs of the IPK are expected to drop. If/when the costs of the IPK become lower than the RAW, switch fully to the IPK program.

The IPK program is currently [redacted] euros per meter more expensive than the RAW contract, however, the IPK has included engineering costs, which are missing in the SOK. Also, the investment costs have to be earned back, therefore it is expected that the costs of the IPK program shrink over time. The IPK program is also expected to work faster and with less hindrance than before due to the new innovative techniques.

The RAW contract is the cheapest, here the competition reduces the costs. Furthermore, the contractor is not present at the design table, which might create disagreements about the design and more iterations over the design cycle. But due to the reduced costs more meters of quay wall can be renovated.

6.3.4. Further advise

It is expected that the costs per meter will drop when working on longer stretches, however this is not directly supported by data.

But it can be logically motivated. The total one-off costs, mobilization and demobilization costs stay the same, but the costs are lower per meter when longer stretches are constructed. It would be even better if the contractors could work continuously, without having to demobilize after renovating each section. Furthermore, due to scaling, materials can be bought in higher volumes, which results in lower per-unit costs. Also, the costs of the transport HUB & NTC are expected to shrink due to scaling.

Furthermore, the costs per meter are expected to drop when the construction speed is higher (m/w). This is supported by Figure 17 below, where a slight correlation can be seen.

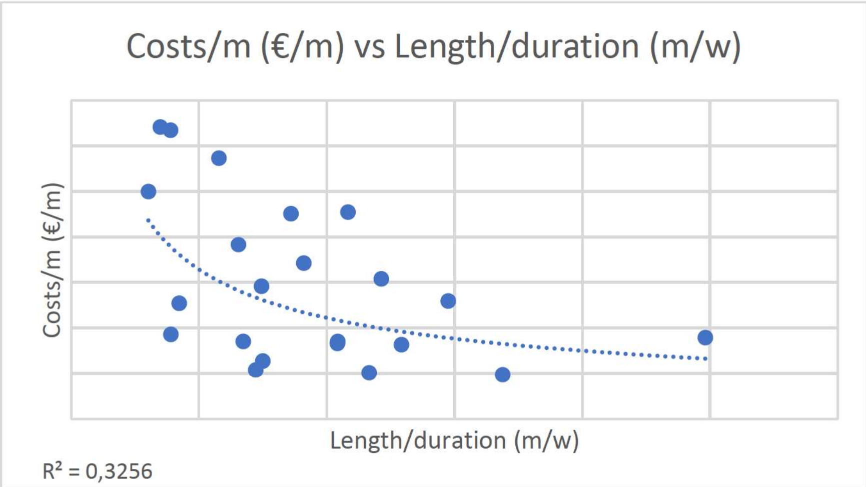


Figure 17 Scatterplot and a fitted nonlinear regression line of costs/m vs length/duration (anonymised)

Furthermore, it is advised to compare the quay wall renovations in Amsterdam to other cities. The 3 projects outside of Amsterdam cost on average [redacted] euro/m, and the projects in Amsterdam cost [redacted] euro/m. However, the contract types outside of Amsterdam where all RAW contracts and therefore this is not a good comparison. Nevertheless, comparing the quay wall renovations in Amsterdam to other cities might provide new insights into what makes the costs of quay wall renovations in Amsterdam so high, and how they possibly can be reduced.

6.4. Advice on data gathering for future research

In this paragraph, advice is given on what data should be gathered for further research. This includes possibilities to improve the current dataset but also which data definitely must be gathered for further research and further advice. The paragraph discusses the following:

+ Data gaps
+ Missing predictors
+ Most promising predictors
+ Further advice on data

6.4.1. Data gaps

First, there are gaps in the data that is used for this research. The missing data are the gaps in Table 18. Many gaps are found in [redacted] these are the two projects of which the exact location of the quay is unknown and are missing information.

The predictors with incomplete data are (excluded the above-mentioned projects):

- Duration (therefore the length/duration is missing data)
- Retaining height (therefore retaining area is missing data)
- Max depth foundation
- Average traffic per day (outside of Amsterdam)

Due to the gaps in data, any combination with one or more of these predictors in the analysis has reduced input data for the regression analysis. Depending on which predictors, the dataset changes. This makes the comparison of the different predictor combinations less transparent, which might result that not the best predictor(s) being chosen. Therefore it is advised to fill the data gaps that are in the input data.

6.4.2. Missing predictors

There was also data which has not been tested in this research due to unavailable data or uncertainty, as discussed in 5.3.2. Missing predictor data:

- Working from water
- Transport materials over water
- Vibration-free working
- Traffic conditions on quay

These predictors are expected to have an influence, however are not tested. Therefore, it is advised to gather and clarify this data, to use them as predictors in future research.

6.4.3. Most Promising Predictors

For the data gathering of the future projects the most important predictor data to gather is determined in this paragraph. The 3 strongest predictors for the simple regression are used and the 3 best combinations in the multiple regression analysis are taken.

Based on the coefficient of determination (R^2), the 3 strongest predictors of simple regression in this research are (5.5. Simple linear regression & 5.6. Simple nonlinear regression):

- Contract type (0,70 adjusted R^2)
- Length (0,36 R^2)
- M/w (0,23 R^2) (nonlinear 0,33 R^2)

In multiple linear regression, the 3 strongest predictor combinations are (5.7.3. Results of multiple regression analysis):

- Contract & Retaining area (0,881 adjusted R^2)
- Contract & Length (0,836 adjusted R^2)
- Contract & m/w (0,771 adjusted R^2)

Based on the strongest predictors the most important predictor data to gather is:

1. Contract type
2. Length
3. Retaining height (for the retaining area)
4. Duration (for the m/w)

For future projects it is advised to gather at least the data of the most important predictors, however, when possible try to gather all the predictors stated in 5.2. Cost estimating ‘hypothesis’, for future analysis. When more predictor data is known other relationships can be tested, which might be better than the relationships found in this research. If the dataset grows with enough projects, it might be possible to add another predictor and it might be that a predictor that currently has not proven itself might be of added value.

6.4.4. Further advice on data

Furthermore, it is advised to retain more (different project data. Currently, the information that is available per project differs. Next to the costs most projects contain drawings and construction specifications however some projects miss these documents. Therefore it is advised to retain the following information: Construction specifications, drawings, client estimate, contract sum, contract variations, and detailed design estimate. Furthermore, relevant information about the project should be stored, such as other reports of the project, interviews with inhabitants etc.

Finally, collect basic data uniformly. For example, project location is stated in different ways in different projects. Sometimes the street and house numbers are stated, in other documents, it is only the quay code that the municipality of Amsterdam assigned to the quay. Uniformly collecting basic data to prevent confusion in the future.

6.4.5. Resume

Advice has been given to the predictor data to be gathered to improve this research. Furthermore, it is advised to gather at least the most important predictor data, however, it is preferred to gather all the predictor data which is stated in 5.2. Cost estimating 'hypothesis'. Furthermore, it is advised to look for a possible predictor in the foundation costs, retain more relevant project data and collect data uniformly.

6.5. Resume

In this chapter first, there is looked into the costs of requirements. The most expensive requirement is that of retaining a monumental tree. Next to trees, BLVC measures, Transport HUB & NTC and houseboat costs are discussed. Furthermore, there has been an increase in the costs per meter of BLVC, Trees, Transport HUB & NTC over the last few years. This can be linked to the introduction of the weight restrictions for vehicles in the centre, and the introduction of the SOK contract type.

Then advice based on the findings of this research is given. The most important advice is to work with the RAW contract and the IPK program, but not with the SOK contract. Furthermore, it is expected that the costs will drop when the project length and construction speed are increased.

Then, advice on data gathering for future research is given. Here advice is given on how to improve the current dataset, but also on the most important predictors to gather data for in future projects.

7. Discussion

The research started with the scope of only using the actual costs incurred in the project. However, there was very limited additional data on the actual costs incurred compared to the first analysis by Witteveen+Bos. Therefore, it was chosen to also include other types of costs: detailed design estimate, client estimate and contract sum were included. However, when including other types of cost the reliability of the input data is reduced. The average variation coefficient is 1,4%, which is not taken into account in this research. This makes the results less reliable.

Eventually, there is only one detailed design estimate project present in the data. This is [redacted] where the other projects have between the +/- 50-80 % direct costs of the total costs, this project has 32% direct costs (4.3.2. Costs composition per project). This makes this project rather odd compared to the other projects and question its reliability.

Furthermore, historical data is used in the research. The oldest project is from 2008. Using data this old, it might be that this data is no longer representative. Due to changes in the construction method and other changes in the construction field that might have had an influence. For example, the change in tendering method (2.2.4. Change in tendering method).

Assumptions have been made to gather as much data for the analysis as possible, this is discussed in 5.3.1. Data sources. Or different sources are used for the same predictor. This can make the results less reliable. Furthermore, there are gaps in the predictor data as can be seen in Appendix G

. However, due to limited data availability, it is chosen to test each predictor for as much data as possible. This results in a changing dataset, where the dataset differs over the different predictors, this makes the comparison between the predictors less transparent.

The most important predictor found in the search for Cost Estimation Relationships is the contract type. However, the RAW contract is no longer used and is replaced by the IPK and SOK contracts. One key assumption of using Cost Estimating Relationships is: ‘the things that affected the costs in the past also influence the costs in the future’ (3.2. Cost Estimation Relationships). However, due to the disappearance of the RAW contract by the introduction of the IPK and SOK contracts. This assumption is not met. Therefore, the chosen CER can be used for explaining the cost differences in the past, but its reliability for predicting the costs is questionable. Other predictors are too weak (a too low coefficient of determination) to provide sufficient predictions.

The behaviour of model 1 is not very realistic, as explained in 5.7.1. Categorical predictors results, the RAW contract becomes more expensive per meter when the length increases. This is not realistic since it is expected that the costs/m will drop when the project is longer.

CER model 2 uses the predictor ‘retaining area’. In this model, the coefficient for the retaining area is negative [redacted]. This means that for the same length of a project when the height is increased, the costs per meter are lower. However, in practice this should be the exact opposite, when the height is increased the costs will rise. The cause for this can be found in the relation of the retaining height and the length, to the costs.

- Increasing the retaining height has a positive impact on the costs and costs per meter
- An increase in the project length has a positive impact on the costs but a negative impact on the costs per meter.

Since both the retaining height and the length have a positive relation to the costs of the project, this product, therefore the retaining height can be used for the total costs of a project. Next, the retaining height has a positive relation to the costs per meter, but the length has a negative relation to the costs per meter. The retaining height can therefore not be used for the costs per meter.

Based on the dataset of the research, the predictor ‘retaining area’ proved to be a good predictor. However, since it cannot be motivated why this works and the explained behaviour above, this model is considered not realistic.

In the costs of requirement, the costs of improving the growing spot of a tree is more expensive than the costs of replacing a tree. However, part of replacing trees is also improving the growing spot. This is odd since this implies that it is cheaper to replace the entire tree and improve the growing spot instead of only improving the growing spot. But this is likely caused by many contractors budgeting differently and since this set of data is rather small, sampling error could have occurred.

8. Conclusion

Looking back at the research objective in section 1.4. The research objective was to find Cost Estimating Relationships, to explain the cost differences between projects and to better predict the costs in projects to come. Furthermore, to provide insight into the costs of the requirement of the municipality of Amsterdam. Finally, to provide recommendations based on the findings of the research and on how to improve the analysis of the costs in the future.

To achieve this, research questions were set up in section 1.7. Research questions. These questions are answered throughout the report. The main findings are:

- The largest costs within the projects are the: foundation, overheads/profit/risk and the management & execution cost. Furthermore, the highest contributors to the cost differences are the foundation, the engineering costs and the management + execution costs.
- The best Cost Estimating Relationship found is the relationship between the contract type together with the retaining area to the costs per meter. This CER model can explain 88% of the difference in the costs.
- The costs of the requirements from the municipality of Amsterdam are analysed, the most expensive requirement was to retain historical trees, next to trees there is looked at the costs of requirements to the BLVC-measures, Transport HUB & NTC costs and the houseboat costs.
- The SOK contact type is the most expensive contract type. Based on CER model 2, when the impact of the retaining height is removed, per meter the SOK contract costs [redacted] euros more than the IPK contract and [redacted] euros more than the RAW contract (5.7.3. Results of multiple regression analysis). Furthermore, the costs of a project are expected to decrease when the project length is longer and when more meters per week are constructed, correlation shows a slight decrease in the costs. (5.6. Simple nonlinear regression, Figure 12).

It is now known where the highest costs are in quay wall renovation projects and what the highest contributors to the cost differences are. This is new insight and can be used to better understand the cost differences. Furthermore, when using the findings of the research the costs of future projects can be reduced. This can be done by working for longer stretches, more meters per week are constructed and when the SOK contract is abandoned. Moreover, there are new insights into what the requirements of the construction process cost. This can be taken into account when setting the requirements for a new construction process, therefore better-considered choices can be made.

9. Recommendations

In chapter 7. Discussion certain shortcomings of the research were discussed, which leads to the suggestions discussed in this chapter.

It is suggested that for further research, the variation coefficient of the different types of costs (Actual costs incurred, contract sum, client estimate and detailed design estimate) are taken into account for the regression analysis. Moreover, when possible, it is preferred to only use the actual costs incurred in the project.

Next, it is suggested to perform the analysis per contract form, right now the contract form is used as a predictor. But when enough data is gathered, it might be better to analyse per contract type. The different contract types could cause different slopes in CERs, but right now all the slopes are the same, which is not very realistic.

Furthermore, when more projects are known, it is recommended that the RAW contract be removed from the analysis. When analysing only the IPK and SOK contract types a better model for estimating can be constructed, this is due to the RAW contract no longer in use.

Next, fill the gaps in the predictor data, or when enough data is known remove the predictors with data gaps. This will make the comparison between the different CERs more reliable.

It is suggested to gather more project data from the last few years, especially on the newer contract types such as the SOK and IPK. Future projects will provide more data, however, currently the research can already be improved by gathering this data. Moreover, in regression analysis, it is always beneficial to have more data, since this makes the analysis more reliable. When possible, interview the project managers and try to discover what could be improved to save costs or speed up the construction process.

It is recommended that Witteveen+Bos looks into the detailed design estimate of [redacted], according to Witteveen+Bos the variance of the detailed design estimate is 10%. However, the costs of project [redacted] largely differ from most projects. This is seen in '4.3.2. Costs composition per project ' where this project had a much higher percentage of indirect costs than the other projects. It is advised to compare the detailed design estimate to the actual costs later on in the project.

Furthermore, it is interesting to compare the quay wall renovation in Amsterdam to other cities. The 3 projects outside of Amsterdam cost on average [redacted] euro/m, and the projects in Amsterdam cost [redacted] euro/m. This might give interesting new insights.

Look for a predictor regarding the foundation costs, since the average foundation costs are the highest in the project with an average of 30,8% of the costs (Figure 9). Discuss with experts which relevant data could be gathered for further research, or what possible predictors could be.

It is recommended to replace the predictor 'retaining area' in CER model 2 with the predictor 'length'. This is more realistic since it can be properly explained and still fits the model well based on the adjusted coefficient of determination. However, it should be validated again.

It is recommended that the requirements are further analysed, for this it is necessary to keep better track of them in future projects.

10. Reference list

CBS. (2020). Basisverlegging inputprijsindexcijfers Grond-, weg- en waterbouw (*Base shifting of input price index figures Civil engineering).* Retrieved from https://www.cbs.nl/nl-nl/longread/discussion-papers/2020/basisverlegging-inputprijsindexcijfers-grond-weg-en-waterbouw-van-2010-100-naar-2015-100/bijlage-1

CBS (*2023, November 30*). Grond-, weg- en waterbouw (GWW); inputprijsindex 2000 = 100, vanaf 1979 (*Civil engineering, input price index 2000=100, from 1979*). Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81139ned/table?ts=1707157529152

CROW (n.d.). Contract typen Schema (*Contract types Scheme*) . Retrieved from https://kennisbank.crow.nl/KennisModule#119550

De Bouwcampus. (2021, June 29). Bijeenkomst Community portfolioaanpak: SOK kademakers Amsterdam (*Community portfolio approach meeting*). Retrieved from https://debouwcampus.nl/bestanden/transitietrajecten/Portfolio_van_Bruggen/documenten/verslag_portfolioaanpakbruggen_communityportfolioaanpak_SOKkademakers-290621.pdf

De Leeuw, M. (2020, June 17). Wie bouwt het slimst, snelst en goedkoopst? Gevecht om monsterklus langs Amsterdamse grachten is beslecht (*Who builds the smartest, fastest and cheapest? Battle for massive job along the Amsterdam canals*). Retrieved from https://www.cobouw.nl/285039/wie-bouwt-het-slimst-snelst-en-goedkoopst-gevecht-om-monsterklus-langs-amsterdamse-grachten-is-beslecht

EUR-Lex. (n.d.). Competition, Summaries of EU legislation. Retrieved from https://eur-lex.europa.eu/summary/chapter/08.html

Gemeente Amsterdam. (2022). Gefundeerd herstellen, actieplan bruggen en kademuren 2023-2026 (*Grounded restoration, action plan for bridges and quay walls 2023-2026*). Retrieved from https://amsterdam.nl

Gemeente Amsterdam. (n.d.-a). Bereikbaarheid, leefbaarheid, veiligheid en communicatie. (*Accessibility, livability, safety and communication*). Retrieved from https://amsterdam.nl

Gemeente Amsterdam. (n.d.-b). Dashboard programma Bruggen en Kademuren (*Dashboard program bridges and quay walls*). Retrieved from https://experience.arcgis.nl

Grachtenmonitor Gemeente Amsterdam. (2022). Grachtenmonitor 2022 (*Quay Wall monitor 2022*). Retrieved from https://amsterdam.raadsinformatie.nl/document/12412936/1/09012f9780b3d678

Ingenieursbureau Gemeente Amsterdam. (2021). Kadevernieuwing van start met drie marktpartijen (*Quay Renovation starts with 3 market parties*). Retrieved from https://www.amsterdam.nl/ingenieursbureau/nieuws/nieuwsberichten/nieuws/2021-jan-jun/kadevernieuwing-start-drie-marktpartijen/

Ingenieursbureau Gemeente Amsterdam. (n.d.). EMVI. Retrieved from https://www.amsterdam.nl/ingenieursbureau/onze-manier-werken/emvi-gunnen-waarde/

International Society of Parametric Analysts (ISPA). (2008). Parametric Estimating Handbook (4th ed.). Retrieved from https://www.scribd.com/document/106939711/Parametric-Estimating-Handbook

Mathworks (n.d.). What is overfitting? Retrieved from https://www.mathworks.com/discovery/overfitting.html

NASA. (2015). NASA Cost Estimating Handbook (CEH): Appendix C - Cost Estimating Methodologies (4.0). Retrieved from https://www.nasa.gov/ocfo/nasa-cost-estimating-handbook-ceh/

Pianoo (n.d.). RAW bestek en raamovereenkomst (*RAW specifications and framework agreement*). Retrieved from https://www.pianoo.nl/nl/sectoren/gww/inkopen-gww/bouworganisatievormen/raw-bestek-en-raamovereenkomst

Scribbr. (2022, August 19). T-tabel | Stappenplan & Voorbeelden (*T-table* | *Roadmap & Examples*). Retrieved from https://www.scribbr.nl/statistiek/t-tabel/

Shukla, M., & Mujumdar, A. (n.d.). Extrapolation Using Regression: Challenges and Solutions. Retrieved from https://www.mphasis.com/content/dam/mphasis-com/global/en/home/innovation/next-lab/Mphasis_Extrapolation_using_Regression_Challenges_and_Solutions_Whitepaper.pdf

Statistics Solutions. (n.d.-a). Dummy Coding: The how and why. Retrieved from https://www.statisticssolutions.com/dummy-coding-the-how-and-why/

Statistics Solutions. (n.d.-b). Sample Size Formula. Retrieved from https://www.statisticssolutions.com/dissertation-resources/sample-size-calculation-and-sample-size-justification/sample-size-formula/

University of Washington. (n.d.). Biostatistics F-tables. Retrieved from https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/F-Tables/

Verheijen, T. (2021, November 24). Amsterdam gaat op slot voor zwaar en vervuilend verkeer. Hoe creatief is de bouw (*Amsterdam is closed to heavy and polluting traffic. How creative is construction*)? Retrieved from https://www.cobouw.nl/300876/amsterdam-gaat-op-slot-voor-zwaar-en-vervuilend-bouwverkeer

Waternet. (2024, January 15). Legger waterschap Amstel, Gooi en Vecht (Map of the Amstel, Gooi and Vecht water board). Retrieved from https://waternet.maps.arcgis.com/apps/MapSeries/index.html?appid=07b1c0250a1043bab51a4584b129a4e1

Witteveen+Bos. (2023, November). Meeting supervisor.

Witteveen+Bos. (2023, September 22). WKP210 - Onderzoek kosten 9 uitvoeringsprojecten (*research the costs of 9 implementation projects*).

Witteveen+Bos. (2024, February 18). Feedback Supervisor

Witteveen+Bos. (n.d.). Project Intake Formulier (*Project intake form*) SOK ID PBK -WP210.

Witteveen+Bos Cost Expert 1. (2024, January 25). Interview.

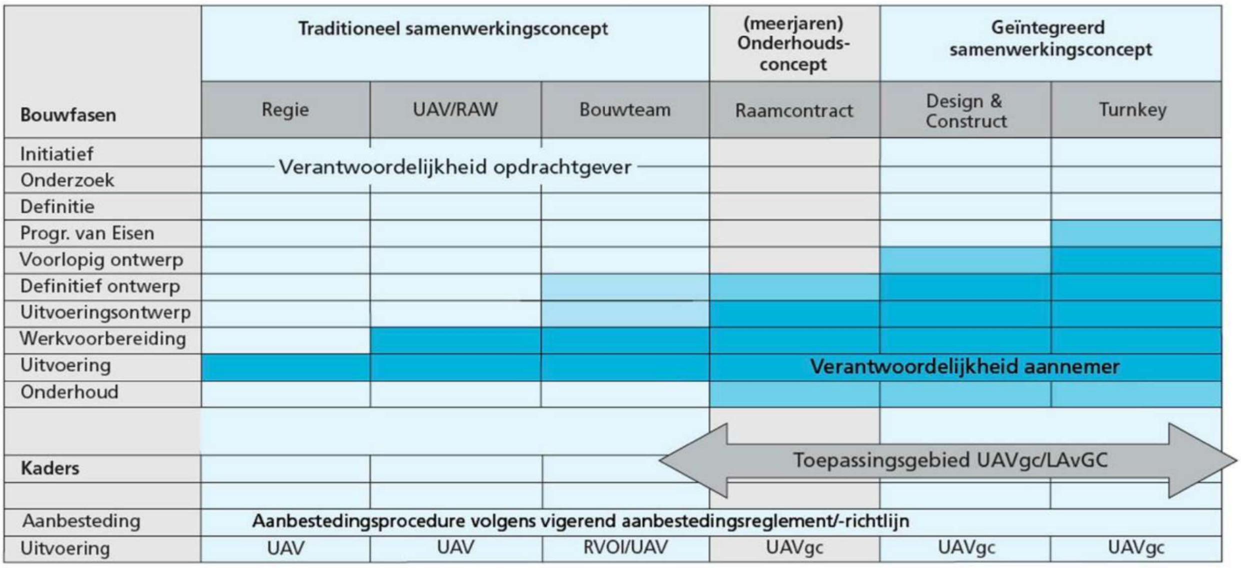
Witteveen+Bos. (2008, Juni, 5). Details Bloemgracht

Witteveen+Bos. (2015, July, 7). Construction Specifications drawing Krom Boomssloot

11. Appendices

Appendix A

Table 13 Contract type table (CROW, n.d.).



De Model Basisovereenkomst met bijbehorende UAV-GC 2005 kunnen worden toegepast als voorwaarden voor opdrachten waarin *ontwerp en uitvoering* zijn geïntegreerd, desgewenst gecombineerd met *meerjarig onderhoud*. Met voorgaande versie – de UAV-GC 2000 – zijn bouwbreed praktijkervaringen opgedaan in de zgn. proef- en implementatieperiode en zijn vragen en opmerkingen binnengekomen. Dit heeft geleid tot inhoudelijke en redactionele aanpassingen van de administratieve voorwaarden. Het betreft geen ingrijpende aanpassingen, maar wel enige juridische optimalisatie en uitbreiding van de toelichting. De *UAV-GC 2005* markeert de afronding van de proef- en implementatieperiode. De ervaringen, vragen en opmerkingen uit die periode zijn verwerkt in de CROW-publicatie 212 *UAVgc: ruim baan voor innovatieve contracten* die daarmee inzicht geeft in de consequenties van de toepassing.

Bij de afbeelding: naast de aangegeven contractvormen zoals bouwteam, design&construct en turnkey worden ook aanduidingen gehanteerd waarin de beginletters van Build, Maintain, Finance, Operate, Transfer e.d. voorkomen. Het juridisch administratief kader van de UA-GC 2005 kan binnen deze contractvormen eveneens gehanteerd worden voor wat betreft de inkadering van de ontwerp- en uitvoeringswerkzaamheden. Het aangegeven toepassingsgebied van de UAVgc is tevens het toepassingsgebied van de in ontwikkeling zijnde Leidraad Aanbesteden van Geïntegreerde Contracten (LAvGC). Deze leidraad gaat uit van de toepasselijkheid van de Europese aanbestedingsrichtlijnen, waarop ook nationale reglementen en (toekomstige) wetgeving zijn gebaseerd.

Uw vragen bij het gebruik, uw ervaring met de toepassing, alsmede uw eventuele commentaren of verbetervoorstellen kunt u kenbaar maken door deze via de website www.uavgc.nl in te brengen. Bovendien kunt u ten behoeve van eigen gebruik vanaf deze website de Model Basisovereenkomst downloaden. Voor eventuele vragen of ondersteuning bij het gebruik kunt u zich wenden tot CROW, Postbus 37, 6710 BA Ede, tel. 0318-695300.

Appendix B

Regression analysis

The steps from the regression analysis are from NASA (*NASA,* 2015).

Assumptions

Before starting with the regression analysis some assumptions need to be made:

1. There exists a linear relationship between the X (dependent) and Y (independent) so that the independent variable(s) can sufficiently explain the variation in the dependent variable.
2. Independent variables have no error.
3. The residuals (error terms) are independent, normally distributed multivariate random variables with mean zero and constant variance.
4. The error terms are not correlated.
5. The data is homogeneous.
6. The independent variables are independent of each other. (If multiple linear regression).

Sidenote: In practice, it occurs often that one or more assumptions are violated, e.g. Independent variables are not completely without error or the sample size is so small that the normality assumption is invalid. However, it is important to check the assumptions. This is done in 5.8. Assumptions and validation. The independent variables are tested earlier.

Formula

The next step is to do the regression analysis. Regression analysis is what makes the parametric cost estimation possible. Often regression analysis is referred to as the Least Square Best Fit (LSBF) method, also known as the Ordinary Least Squares (OLS).

Regression analysis consists of two parts. The first part is to quantify the relationship between variables using a mathematical expression. The second part of the regression analysis is about the accuracy of the relationship and the validation. First, the first part is discussed.

In simple linear regression, the formula is: $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1$, these coefficients can be determined by using standard formulas. Excel can calculate this in a formula when fitting a regression line.

Furthermore, when the scatterplot of data shows a nonlinear relationship simple nonlinear regression can be tested. Data can be transformed using Logarithms to straighten it or using a power function. The following form can be used for the power function ($y = AX^B$).

The general formula for the Multiple regression model is:

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{b}}_0 + \hat{\boldsymbol{b}}_1 \boldsymbol{X}_1 + \hat{\boldsymbol{b}}_2 \boldsymbol{X}_2 + \cdots + \hat{\boldsymbol{b}}_k \boldsymbol{X}_k$$

$\hat{\boldsymbol{Y}}$= estimated value of dependent value (estimated cost)

$\boldsymbol{X}_i$= ith dependent variable (predictors)

k = number of independent variables

$\hat{\boldsymbol{b}}_i$= estimated regression parameters

To perform the statistical regression analysis with multiple variables the procedure is slightly different from simple regression analysis (1 variable). Therefore there is no exact formula for determining the values of the regression parameters. They should be estimated using the sample data. Using the method of least squares to determine the minimized sum of the squared error terms. Eventually, a system of normal equations remains.

The formula for the error terms:

$$\boldsymbol{e}_i = Y_i - \hat{Y}_i$$

$\boldsymbol{e}_i$ = error

Y_i = observed value of dependent value (actual cost)

Here the assumptions need to be checked before the CER model can be used:

1. Functional form is spelled correctly
2. Values of the independent variables are known without error
3. Residuals or error terms, $e_i = Y_i - \hat{Y}_i$, are independent, normally distributed multivariate random variables with mean zero and constant variance.
4. The error terms are not correlated
5. The data are homogeneous
6. The independent variables are independent of each other (Multiple regression)

Even if some assumptions are not met. The CER model can still prove valuable.
The assumptions act more as guidelines than rules.

Appendix C

Test relationships (validation)

According to NASA (*NASA,* 2015):

This is the second part of the regression analysis, about the accuracy of the relationship, and the validation. This is done by various statistics. These are the four methods of analysing the found regression:

1. The regression coefficient of determination (**R^2**)
2. Standard error of estimate (SE)
3. Coefficient of variation (CV)
4. Strength of regression coefficients (hypothesis test or T-tests)

Regression coefficient of determination

To determine the goodness of fit of the regression line the coefficient of determination, **R^2** is used. **R^2** can have a value from 0 to 1, the closer to the 1, the better the fit.

The **R^2** is a number which can explain how much of the deviation of the dependent variable can be explained by the independent variable. E.g. an **R^2** value of 0,7 means that the independent variable can explain 70% of the variation in the dependent variable, leaving 30% over to chance.

Coefficient of determination:

$$\boldsymbol{R^2} = \frac{\sum(\boldsymbol{Explained\ Deviations})^2}{\sum(\boldsymbol{Total\ Deviations})^2} = \frac{\sum(\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})^2}{\sum(\boldsymbol{Y} - \bar{\boldsymbol{Y}})^2}$$

$\boldsymbol{Y}$= Actual value

$\hat{\boldsymbol{Y}}$= Dependent variable, the estimated value

$\bar{\boldsymbol{Y}}$= Mean value of actual ***Y.***

Coefficient of multiple determination:

Since the project deals with multiple variables the Coefficient of determination also needs to be adjusted to the coefficient of multiple determination.

$$\boldsymbol{R_a^2} = \boldsymbol{1} - \frac{\frac{\sum(\boldsymbol{Unexplained\ Deviations})^2}{\boldsymbol{n-k-1}}}{\frac{\sum(\boldsymbol{Total\ Deviations})^2}{\boldsymbol{n-1}}} = \boldsymbol{1} - \frac{(\boldsymbol{1} - \boldsymbol{R^2})(\boldsymbol{n} - \boldsymbol{1})}{(\boldsymbol{n} - \boldsymbol{k} - \boldsymbol{1})}$$

n = number of data points

k = number of independent variables

Unexplained deviations = $(\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})$

Standard Error of the Estimate (SE)

The SE tells the deviation of the sample points from the fitted regression line. This is an absolute value and can therefore on itself not be used for evaluation. However, it is used to calculate the coefficient of variation.

$$SE = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - k - 1}}$$

The sample size (n) must exceed the number of independent variables (k) by at least one, otherwise, there is no 'real number' solution. If the sample size is one bigger than the number of independent variables the SE formula has no solution. So therefore it is important to have the sample size exceed the number of independent variables by two.

Coefficient of Variation (CV)

The CV tells the degree of variation. It is the Standard Error of the Estimate divided by the mean of the dependent variable ($CV = \frac{SE}{\bar{Y}}$). Which enables it, to compare with other models.

The CV should be less than or equal to 20%, preferably less than 10%.

F-test or Hypothesis test

The F-test test whether the regression relationship between the dependent variable and the set of independent variables is statistically significant. The F-test is done for the entire CER model. When the existence of the relationship is established. The F-test is done for all the coefficients.

The hypothesis test is performed to determine if the coefficient ($\hat{b}_i$) of the independent variable is significantly different from 0. The slope of 0 would imply that the relationship is purely chance. Which would imply that X is of no use in predicting Y.
The significance of the intercept ($\hat{b}_0$) is not of importance since this is almost always outside the range of the data, which would impose extrapolation and is therefore not taken into account.

The F-test consists of the following steps

1. Formulate hypothesis
2. Choose desired level of significance
3. Find F_p (model case) or T_p (coefficient case)
4. Calculate F_c or T_c
5. Decide on the hypothesis
   This can be to reject the hypothesis and, therefore conclude the slope is statistically significant. Or fail to reject the hypothesis, Therefore conclude that the slot is not statistically significant.

Appendix D

Data from Witteveen+Bos server

In the database, there has searched for the terms: "kademuren" OR "kademuur", from the year 2000, and the restrictions that the detailed design estimate, tender stage client estimate or contract sum had to be available. However, in this stage, it became clear that not all data had been added properly to the database. Since a few projects that were known earlier were not shown in the results from the search, it became clear that data was missing. Therefore the filters of detailed design estimate, tender stage client estimate or contract sum were turned off.

This resulted in a dataset of 127 projects. This has been filtered to see which projects might be valuable. The projects were quickly filtered on the following:

- Quay wall is located inside a city centre (which was often clear from the title in the database)
- Quay wall is not part of a (large) harbour

This filter passed 45 cases, 43 were unclear and 39 were removed.

Then each case was looked into in detail. This was done on the server. The database is more of an overview of all projects e.g. to quickly find a project in a certain category. Where all documents are stored on the server. That is why for the detailed analysis this is done in the server.

Then a problem arose.
"According to the ISO9001 quality system of W+B, the project manager is responsible for the (digital) project archive. As the cost management department is supportive of each project we're not in charge of the full project. If over the past years, a project is closed, data is moved to the archive after 5 years. Data then is out of reach from the cost management department and can only be retrieved by project manager or project director."(*W+B Supervisor, 2024*).

In this case could be retrieved from the server, however, the server location has been changed. Therefore the search began to find the projects on the server. Many locations were found, however, some were not found so the number slightly dropped.

At this stage, each map was checked if it contained suitable data: The availability of the detailed design estimate, tender stage client estimate (hereafter referred to as 'client estimate') or contract sum was necessary. Furthermore, some cases included a detailed design estimate, however the location of the project was nowhere to be found and contained no further information. So these were therefore removed.

It has been assumed that the most recent tender bid of a project can be used as the contract sum. This had to be done since it was not always confirmed that the bid was accepted.

It has been decided that projects outside of Amsterdam could also be used. This is done to further increase the number of projects. Here, however, the situation should be comparable to the situation in Amsterdam, e.g. in a dense city centre.

After inspecting each project, 14 cases remained. However, after researching them better 3 were removed. The project of [redacted] was removed due to the detailed design estimate being mainly based on the average costs of the [redacted] project available from Amsterdam. The other 2 locations were in Deventer, it was decided to remove these locations due to their differences: there was no traffic on the quay walls, the quay walls were located next to a large river and the design differed, the sum of these factors made the decision to remove these projects.
11 projects from the W+B server remained.

Data received from the municipality of Amsterdam

The first research of W+B on the costs of the quay walls included 9 recent projects from Amsterdam. However, in the meantime, the municipality of Amsterdam provided three additional projects. In two projects the actual costs incurred were known, and in 1 project only the contract sum was known.

Additional data from the municipality of Amsterdam

An engineer from Amsterdam did his research on the costs of quay walls, which was shared with Witteveen+Bos. From this dataset, 2 additional projects were added to this dataset. However, little information is known about these projects and their exact location is unknown (it is known at what quay they are located, however not the exact spot) so their usability is limited.

Table 14 Overview of the projects used in the analysis (anonymised)

Project	City	Location	Data source
[redacted]	Amsterdam	[redacted]	Amsterdam[1]
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam
[redacted]	Amsterdam	[redacted]	Amsterdam (2)[2]
[redacted]	Amsterdam	[redacted]	Witteveen+Bos[4]
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Amsterdam (2)
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Amsterdam (2)
[redacted]	Haarlem	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Haarlem	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	De Rijp	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Witteveen+Bos
[redacted]	Amsterdam	[redacted]	Amsterdam (X)[3]
[redacted]	Amsterdam	[redacted]	Amsterdam (X)

[1] Amsterdam: Municipality of Amsterdam (anonymised)

[2] Amsterdam (2): is the second batch received from the municipality of Amsterdam

[3] Amsterdam (X): is the data from the analysis of X from the municipality of Amsterdam

[4] Witteveen+Bos: is the data from the Witteveen+Bos server

Appendix E

Table 15 Percentage contract variations of total cost (anonymised)

Project	Percentage contract variations total costs
[redacted]	10,4%
[redacted]	0,0%
[redacted]	1,8%
[redacted]	0,0%
[redacted]	9,6%

[redacted]	4,6%
[redacted]	5,5%
[redacted]	2,0%
[redacted]	4,9%
[redacted]	0,0%
[redacted]	16,0%
Average	**5,0%**

Table 16 Types of costs and variation coefficient per project (anonymised)

Project	Type of costs	Variation coefficient
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	Actual costs incurred	0%
[redacted]	contract sum	0%
[redacted]	client estimate	5%
[redacted]	contract sum	0%
[redacted]	detailed design estimate	10%
[redacted]	Actual costs incurred	0%
[redacted]	contract sum	0%
[redacted]	contract sum	0%
[redacted]	contract sum	0%
[redacted]	contract sum	0%
[redacted]	client estimate	5%
[redacted]	client estimate	5%
[redacted]	client estimate	5%
[redacted]	client estimate	5%
[redacted]	contract sum	0%
[redacted]	contract sum	0%
Average		1,4%

Appendix F

Table 17 Average percentage and Standard deviation of project costs per class

Classification nr	Percentages	Average %	Stdev %
11	Surface design/sewage/cables & piping	4,1%	3,3%
14	Trees	0,8%	1,2%
21	Construction pit and auxiliary works	5,6%	5,6%

23	Auxiliary bridge structures	0,3%	0,7%
24	Earthworks	2,2%	2,2%
26	Dredging activities/cleaning waterway	0,4%	0,8%
27	Drainage	0,4%	0,5%
31	Demolition quay wall	2,2%	1,7%
41	Foundation	30,8%	17,0%
44	Concrete work	6,7%	6,4%
51	Masonry and cover stones	1,9%	0,9%
54	Quay facilities	0,6%	0,7%
61	Guide piling	0,2%	0,7%
71	Phasing BLVC-measures	2,3%	2,3%
75	transport HUB & NTC	4,1%	6,3%
77	Side projects (Houseboats, Pont etc.)	0,5%	1,3%
dk	**Direct costs**	**63,2%**	**51,6%**
81	Monitoring, research and end documentation	1,9%	1,8%
84	fees, insurance and bank guarantees	0,3%	0,7%
87	Risk reservation- posts	0,6%	1,2%
91	One-off costs and construction site costs	6,2%	6,2%
93	Management + execution costs	10,1%	8,8%
95	Engineering costs	3,4%	10,8%
96	Overheads	5,6%	4,6%
97	Profit/risk	3,4%	3,8%
9697	Overheads/profit/risk	1,4%	4,0%
98	Provisional sum/ place at disposal of	1,0%	1,1%
99	Contributions	0,1%	0,1%
index	indexations	0,1%	0,2%
0	Divers / Unclear	0,0%	0,1%
VTW	Contract variations	2,7%	5,1%
ik	**Indirect costs**	**36,8%**	**48,4%**

Appendix G

Table 18 Predictor data sheet (removed for anonymization)

Appendix H

Source

For the indexation of the data in this project, the CBS index numbers are used. Another possibility was to use the indexation or the CROW, however, this data is much more specific than the data of the CBS. The construction costs are not specific enough for the CROW index to be applied. The CBS is more general, however even for this index, certain assumptions are made to use this source for the index.

Index category

To determine which index to use the classification system of paragraph 4.3.1. Structuring the data, can be used. Here it can be seen what the large cost drivers are. To make it more clear what the main costs drivers are in

Appendix F. Here it can be seen that the largest direct costs are in: Foundation (32,6%), Construction pit and auxiliary works (7,1%) and Concrete work (6,7%). The 3 largest indirect costs are: Management +execution costs (9%), One-off costs and construction site costs (6,1%) and Overheads (5,7%).

Since the largest contributors to the costs are known this can be used to determine the best Index category from the CBS. For this the weights of the 'Grond,-weg, en waterbouw' (*CBS, 2020*) are compared (Table 19).

It would be preferable to match the average percentage of the classification directly to the weights that determine the index groups. However, they are ordered differently. For example, in the project, the costs for steel are not shown as steel costs but are embedded in the foundation costs. However, in the foundation costs, there are also costs of machinery, labour and concrete for 'groutankers'.

The following motivation is used to find the best index:

- The foundation costs of the project are on average 32,6%. This consists of Steel, Concrete ('Grout ankers'), machinery and labour. Where it is assumed that machinery, labour and concrete play a large role in costs.
- The cost of concrete work is 6,7%, however, the Masonry and cover stones is put in the same category and add up the concrete in foundation. There are of course also labour and machinery costs involved however these are expected not to be as high since most panels are prefabricated.
- It is expected that a lot of costs in quay wall renovation projects are that of labour. This is because of the many challenges the city faces in constructing in such a densely populated and historical area.

Two candidates remain using this analysis. This is 4212 for railway construction and 4213 which is for bridges and tunnels. 4212 matches the concrete/metal expectations better, however here the costs of groundwork are too high while 4213 matches the expectations for labour better. Eventually, it is chosen to use 4213 for the index number. This is also because, in the initial research performed by Witteveen+Bos, it was assumed that 4213 was the best fit. Therefore, due to both indexes having the motivation to choose them, but with the expert insight it was decided to pick 4213.

Omschrijving	42/43 Grond-, weg- en waterbouw	4211a Wegen met open verharding	4211b Wegen met gesloten verharding	4212 Boven- en ondergrondse spoorwegen	4212a Spoorwegen Aanleg	4212b Spoorwegen Exploitatie/onderhoud	4213 Bruggen en tunnels	4221 Werken voor vloeistoffen	4291 Waterbouwkundige werken	4312 Bouwrijpmaken van terreinen	4321 Elektrische installatie
Zand en grind	12	10	12	5	6	2	0		31	38	
Textielvlies	0								0		
Houtproducten	0				1	0	1		0		
Diesel en smeeroliën	3	3	5				1	3	4	10	1
Verf en verfproducten	0					0					
Rubber en kunststoffen	2	0	0			0	0	8			
Betonmortel, betonproducten en bitumineuze producten	25	32	44	12	11	4	19	45	22		4
Metalen	2			18	21	14	17				3
Constructies van metaal	6		2		1	0	3				28
Elektronische en optische producten	5		0	9	3	4					29
Elektrische apparatuur	2			6	3	1					8
Machines en werktuigen	6	4	8	14	13	7	9	6	3	7	3
Auto's en aanhangers	2	2	4				0	1	1	10	1
Andere transportmiddelen	1								4		
Energie	0				1	5					
Verzekering en belasting	0					1					
Dienstverlening w.o. kosten transport en softwarelicenties	0				2	11					
Arbeid	33	49	25	37	39	50	49	38	35	35	23
Totaal	100	100	100	100	100	100	100	100	100	100	100

Table 19 Index weights (CBS 2020)

Table 20 Index values CBS 4213 Bruggen en tunnels bouw (CBS 2023)

	4213 Bruggen en tunnels bouw
2005 January	107,1
2006 January	111,4
2007 January	117,6
2008 January	122
2009 January	130,9
2010 January	119
2011 January	123,5
2012 January	124,5
2013 January	126
2014 January	126,4
2015 January	124,7
2016 January	124,1
2017 January	128,2
2018 January	132,6
2019 January	135,3
2020 January	137,5
2021 January	142,9
2021 April	144,7
2021 Juli	152
2021 Oktober	153,9
2022 January	158,8
2022 April	172,9

2022 Juli	171,4
2022 Oktober	170
2023 January*	169,7
2023 April*	170,5
2023 Juli*	168,6
2023 October*	168

*Table 21 Normalised costs (*referred to October 2023)*

Appendix I

Table 22 Coefficient of determination (R2) output of Simple Linear Regression

R^2 output of Simple Linear Regression	**Costs/m (€/m)**	**Costs (€)**
Length (m)	0,05	**0,361**
Duration (w)	0,03	**0,16**
Construction pit	0,01	0,00
m/w	**0,23**	0,00
Witdh canal (m)	0,00	0,11
Retaining Height (m)	0,00	0,00
Average traffic per day	0,01	0,00
Max depth foundation (m below NAP)	0,12	0,04
Retaining area (m2)	0,05	**0,19**

Table 23 Adjusted coefficient of determination of Construction technique and Material

Independent -dependent variable	**R^2 adjusted**
Construction - costs	0,0222
Construction - costs/m	-0,0181
Material - costs	0,0046
Material costs/m	-0,0074
Contract - costs	0,5537
Contract - costs/m	0,7037

Appendix J

Table 24: Number of independent variables per predictor combination

Nr of independent variables	1	2	3	4	5	6	7	8	9
1. Length (m)									
2. Duration (w)	2								
3. Construction type	4	4							
4. Lenght/Duration (m/w)	2	2	4						
5. Material	3	3	5	3					
6. Witdh canal (m)	2	2	4	2	3				
7. Max depth foundation (m below NAP)	2	2	4	2	3	2			
8. Length*Retaining height (m2)	2	2	4	2	3	2	2		
9. Contract type	3	3	5	3	4	3	3	3	
Retaining height	2		4						3
Construction pit	2								

Appendix K

Before continuing the formula is checked whether the formula makes sense. For this check, the boundaries from paragraph 5.4.3. are used.

- **CER Model 1:**
  - Limits of the input in length between 44-291 m.
  - The original limits of the costs were between [redacted] euro.
  - The original limits of the costs/m was between [redacted] euro.
- **CER Model 2:**
  - Limits of the input retaining area between 145-1718 m^2.
  - The original limits of the costs/m was between [redacted] euros.

The limits are tested for both models, the results can be found below in Table 25. It can be seen that some limits pose extreme values. The lower limit in model 1, causes the costs per meter for the SOK to be [redacted] euros. Where the original highest value was [redacted] meters, this is too extreme.

For model 2 the RAW in the upper limit even goes negative, which would imply that the client would receive money from the contractor in case of construction. Therefore limits are reviewed.

Table 25 Limit test of CER model 1 and 2

CER model 1	Costs	Costs/m	CER model 2	Costs/m
Lower limit (44m)			**Lower limit (145m2)**	
IPK	[redacted]	[redacted]	IPK	[redacted]
SOK	[redacted]	[redacted]	SOK	[redacted]
RAW	[redacted]	[redacted]	RAW	[redacted]
Upper limit (291m)			**Upper limit (1718m2)**	
IPK	[redacted]	[redacted]	IPK	[redacted]
SOK	[redacted]	[redacted]	SOK	[redacted]
RAW	[redacted]	[redacted]	RAW	[redacted]

It looks like the problem of the extreme values is caused by the use of dummy variables. The dummy variables cause there to be only one contract form at a time. Therefore the limits should be linked to that of the present contract. Therefore depending on which contract type, other limits are necessary. In, Table 26 below the new limits are given. In Figure 18 these limits are visualized.

Table 26 Advised model limits per contract type

CER Model 1 limits (Length, m)	**IPK**	**SOK**	**RAW**
min	171	61,5	44
max	204	291	280
CER Model 2 limits (Retaining area, m^2)	**IPK**	**SOK**	**RAW**
min	324,9	233,7	145,2
max	498,8	1718,2	931,3

The behaviour of the models within the limits is shown below in Figure 18.

Figure 18 Behaviour of CER models 1 and 2 within Limits (removed for anonymization)

Appendix L

Figure 19 Behaviour of CER model 1 (costs/m vs length) within limits (removed for anonymization)

Figure 20 Scatterplot and fitted linear regression line per contract type Costs(/m) vs length, (removed for anonymization)

Appendix M

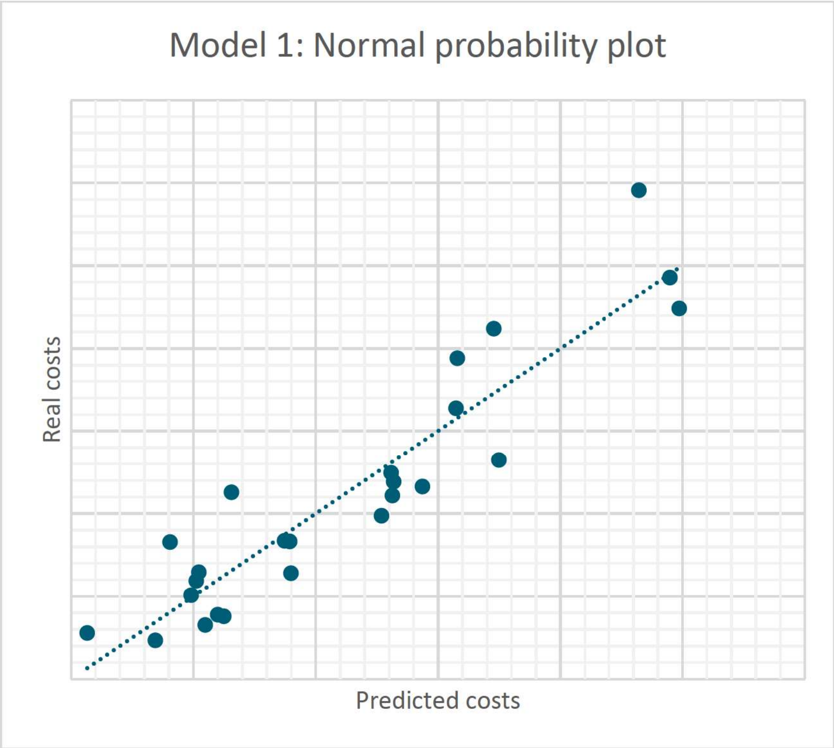


Figure 21 CER Model 1 Normal probability plot (anonymised)

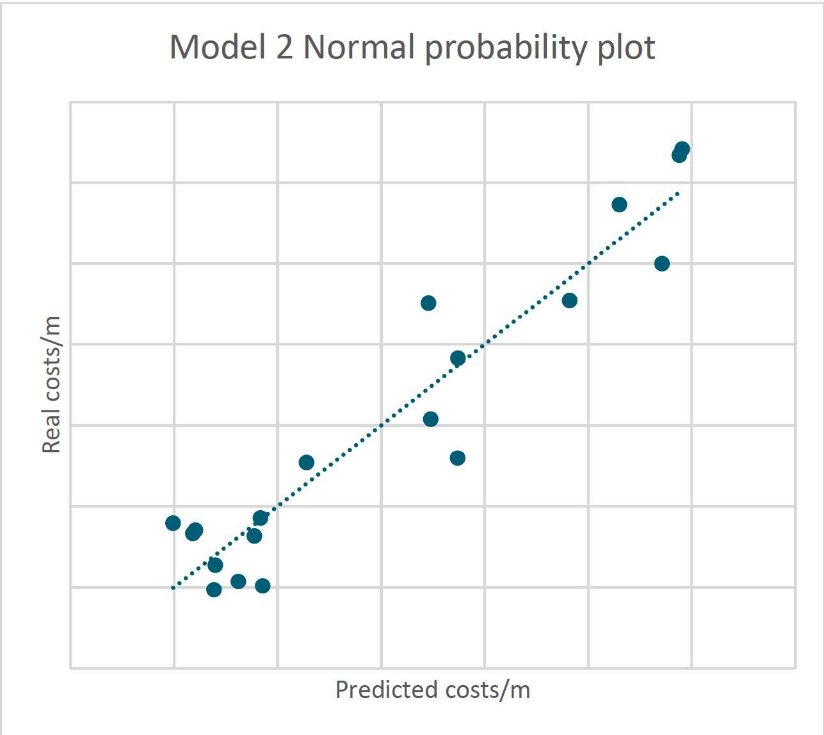


Figure 22 CER Model 2 Normal probability plot (anonymised)

Figure 23 Residual plots CER model 1 (anonymised)

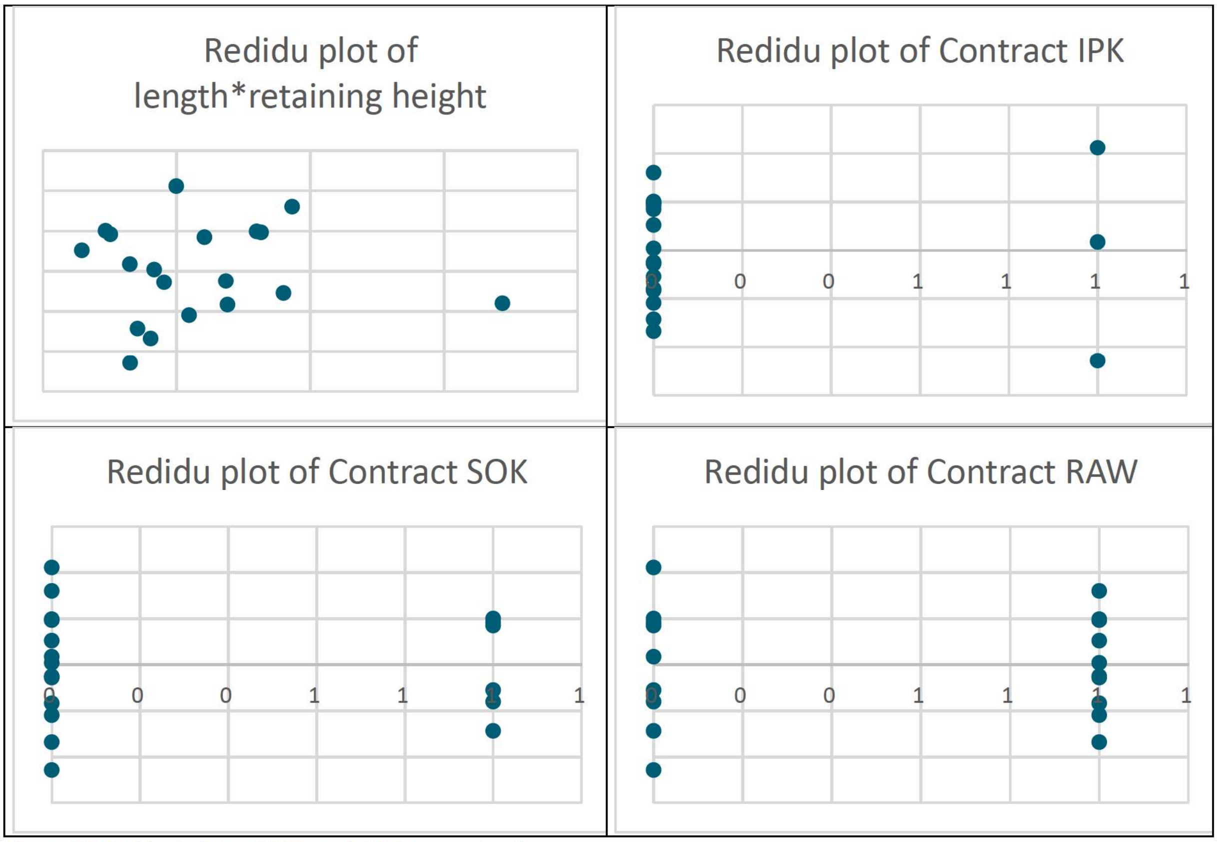


Figure 24 Residual plots of CER model 2 (anonymised)

Appendix O

Table 27 Correlation test Model 1

Correlation test (R)	**1**	**2**	**3**
1. IPK	**1**		
2. SOK	-0,230	**1**	
3. Length	0,091	0,067	**1**

Table 28 Correlation test model 2

Correlation test (R)	**1**	**2**	**3**
1. IPK	**1**		
2. SOK	-0,294	**1**	
3. Retaining area	-0,248	0,180	**1**

Appendix P

Hypothesis test Multiple regression model:

The hypothesis tests are performed simultaneously for both models.

	Nr of measurements (n)	Nr of independent variables (k)	Adjusted R^2
Model 1	25	3	0,836
Model 2	19	3	0,881

1. H_0: $b_1=b_2=0$
   H_1: Not all $b_k=0$
2. Significance (α) = 5%
3. Value of the F_p value is found in the F-table (Figure 25):
   Model 1: $F_p = F_{(1-\alpha,\ k,\ n-k-1)} = F_{(1-5,\ 3,\ 25-3-1)} = F_{(0,95,3,21)} = 3{,}07$
   Model 2: $F_p = F_{(1-\alpha,\ k,\ n-k-1)} = F_{(1-5,\ 3,\ 19-3-1)} = F_{(0,95,3,15)} = 3{,}29$
4. Calculate $F_c = \frac{R^2/k}{(1-R^2)/(n-k-1)}$
   Model 1: 35,58
   Model 2: 37,02
5. Conclusion
   Because In both models $F_c > F_p$, it can be concluded that the independent variables are statistically significant.
   Now the existence of the regression analysis is established.

Hypothesis test independent variables:

1. H_0: $b_i=0$
   H_1: Not all $b_i \neq 0$
2. Significance (α) = 5%
3. The value of the t_p value (two-tailed test, since the opposite of hypothesis is always not equal to) is found in the T-Table (Figure 26):
   Model 1: $t_{(1-\alpha/2,\ n-k-1)} = t_{(1-5/2,\ 25-3-1)} = 2{,}385$
   Model 2: $t_{(1-\alpha/2,\ n-k-1)} = t_{(1-5/2,\ 19-3-1)} = 2{,}433$
4. The t_c value is given by Excel

Model 1	t_c value
Contract IPK	4,4043643
Contract Sok	8,1145568
Length (m)	6,2263598

Model 2	t_c value
Contract IPK	4,051149
Contract SOK	11,2728
Retaining area	-3,73329

5. Conclusion

   Both in model 1 and in model 2: for all independent variables $|t_c| > |t_p|$. Therefore, the null hypothesis is rejected and all the slopes **are** statistically significant.

	Numerator degrees of freedom																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	6.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

*Figure 25 Critical values (percentiles) for the **F** distribution. Upper one-sided 0.05 significance levels; two-sided 0,10 significance levels; 95 per cent percentiles (University of Washington,* (n.d.))

Vrijheidsgraden *(df)*	.2	.15	.1	.05	.025	.01	.005	.001
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	636.619
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.599
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.924
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.610
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.869
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.959
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.408
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.041
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.781
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.587
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.437
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.318
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.221
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.140
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.073
16	1.337	1.512	1.746	2.120	2.473	2.921	3.252	4.015
17	1.333	1.508	1.740	2.110	2.458	2.898	3.222	3.965
18	1.330	1.504	1.734	2.101	2.445	2.878	3.197	3.922
19	1.328	1.500	1.729	2.093	2.433	2.861	3.174	3.883
20	1.325	1.497	1.725	2.086	2.423	2.845	3.153	3.850
21	1.323	1.494	1.721	2.080	2.414	2.831	3.135	3.819
22	1.321	1.492	1.717	2.074	2.405	2.819	3.119	3.792
23	1.319	1.489	1.714	2.069	2.398	2.807	3.104	3.768
24	1.318	1.487	1.711	2.064	2.391	2.797	3.091	3.745
25	1.316	1.485	1.708	2.060	2.385	2.787	3.078	3.725
26	1.315	1.483	1.706	2.056	2.379	2.779	3.067	3.707
27	1.314	1.482	1.703	2.052	2.373	2.771	3.057	3.690
28	1.313	1.480	1.701	2.048	2.368	2.763	3.047	3.674
29	1.311	1.479	1.699	2.045	2.364	2.756	3.038	3.659
30	1.310	1.477	1.697	2.042	2.360	2.750	3.030	3.646
40	1.303	1.468	1.684	2.021	2.329	2.704	2.971	3.551
50	1.299	1.462	1.676	2.009	2.311	2.678	2.937	3.496
60	1.296	1.458	1.671	2.000	2.299	2.660	2.915	3.460
70	1.294	1.456	1.667	1.994	2.291	2.648	2.899	3.435
80	1.292	1.453	1.664	1.990	2.284	2.639	2.887	3.416
100	1.290	1.451	1.660	1.984	2.276	2.626	2.871	3.390
1000	1.282	1.441	1.646	1.962	2.245	2.581	2.813	3.300
Oneindig	1.282	1.440	1.645	1.960	2.241	2.576	2.807	3.291

Figure 26 Critical values of T for the two-tailed test (Scribbr, 19-08-2022)