Tenders in the Dutch construction industry

The impact of client decisions on the tender competition

Thesis

S.J. Groen

Delft University of Technology

Tenders in the Dutch construction industry

The impact of client decisions on the tender competition

by

S.J. Groen

Student Name	Student Number
S.J. (Bas) Groen	4373324

Research committee members:

Chair:	Prof. dr. H.L.M. Bakker
First supervisor:	Ir. L.P.I.M. Hombergen
Second supervisor:	Ph.D. LL.M. S. Renes
Company mentor:	Drs. D. Zwerk (Rijkswaterstaat)

Cover: Construction of the Gaasperdammertunnel, by Ballast Nedam
Style: TU Delft Report Style, with modifications by Daan Zwaneveld

TU Delft

Preface

Before you lies the master thesis ”Tenders in the Dutch construction industry: The impact of client decisions on the tender competition”. With this thesis, the master’s degree for the Master of Science in Construction Management and Engineering is obtained at the faculty of Civil Engineering of the Delft University of Technology. The research presented has been executed from July 2022 to July 2023 in close collaboration with Rijkswaterstaat.

From the start of my bachelor of Civil Engineering, my interest in infrastructure and large constructions increased. The scale, risks, complexity, technological level, and value of infrastructure projects often leave me impressed. Something that motivated me to start as a work student at Rijkswaterstaat during my master Construction Management and Engineering. Due to the combination of a job at Rijkswaterstaat and an education on the construction sector, I kept up to date closely on the procurement problems and disputes in the industry. It intrigued me how clients and contractors interact in the sector and I wanted to further understand the drivers behind this behaviour. As a result, the idea sparked me to graduate on a topic related to procurement in the Dutch construction sector.

Now, officially one year after the start of my thesis an can look back on an eventful year. With the thesis as a central theme in my daily life, some impacting events occurred. I moved to Rotterdam during the graduation process and I searched to secure my first serious job. But because of the support of people around me, I believe I was able to complete the graduation process successfully. Therefore I would like to take the opportunity here to thank the people that contributed to this research.

Firstly, I want to thank my research committee members from the TU Delft. My chair professor Hans Bakker, first supervisor Leon Hombergen, and second supervisor Sander Renes gave essential support in the writing process, shared their extensive knowledge on the construction market, and provided essential support in the application of data science methodologies.

Secondly, I would like to thank the ICG department of Rijkswaterstaat for providing a position to graduate. ICG made me feel welcome and at home right from the start. Because of the positive environment, ICG played a major role in my graduation process. I am very grateful for the sincere interest of my ICG colleagues and that they have treated me as a full member of the department. Also, I would like to especially express my gratitude to my company supervisor Danny. Despite personal circumstances, Danny made sure to put effort into welcoming me at ICG and provided me with helpful feedback on my research and graduation process.

Thirdly, I want to thank TenderNed and the EIB for providing essential data for the research. Duuc, Rick-Jan, and Maël from TenderNed supplied additional information on how many competitors registered to execute specific projects. Nylas from the EIB helped by sharing information on market conditions in the Dutch construction industry. Without this information, a large number of projects would have to be left out of the research.

Last but not least, I would like to thank my friends and family. My friends because they contributed to the wonderful time I experienced during my studies and the friendships we maintain. My family because they have always supported me in my decisions and helped me in building my own life.

For now, I hope you enjoy reading my thesis!

S.J. Groen
Rotterdam, July 2023

Summary

As a result of decreasing competition within the procurement of large construction projects in the Dutch construction sector, the need arises to understand the mechanisms behind the competition. Especially how decisions by clients on the procurement of construction projects impact the number of competitors. When this is understood client can contribute to solving the problems in the construction market. In an attempt to supply clients with insights to impact the number of competitors, this thesis tries to answer the main research question ”To what extent do client decisions in the procurement of construction projects in the Dutch construction sector affect the number of competitors?”.

In an attempt to answer the question, it has been tried to create a model that predicts the number of competitors in the tender phase of construction projects. To be able to make predictions, data on factors that potentially affect the number of competitors is inserted in a regression analysis. Values are included that represent the variables project type, project value, project duration, client, contract type, market conditions, need for work, and tender method. The regression coefficients determined in this analysis can reveal how individual variables affect the predicted number of competitors. But the usability of these coefficients depends on the performance of the analysis set-up, i.e. the used regression model, chosen variables, and input data.

It turned out that the achieved prediction performances were unsatisfactory. The best-performing model reached a prediction accuracy of 44.9%, just 8.7% better than a constant guess. As a result, it is concluded that the relationship between the client factors and the number of competitors could not be quantified. Additional research on the predictability within the data affirmed that it would be hard to make any predictions. The Entropy and Mutual Information within the data, identifying the predictability, showed that the independent variables in the data could only solve approximately 30% of the ’chaos’ within the distribution of the number of competitors.

The unsatisfactory model accuracies are an interesting result, given the fact that, for almost every factor considered influential on the number of competitors, a value has been included in the calculations. The question arises whether even a relation exists between client factors and the number of competitors. On top of that, the identified marginal relations pointed at the market conditions as most influential to the number of competitors. Further indicating that clients have a marginal effect on the competition within tenders. Therefore, it seems that clients are unable to solve the decreasing number of competitors in the construction industry.

In order to show the absence of a relation between the number of contractors and client factors, further research is necessary. It is yet impossible to prove this absence because improvements can still be made to the utilized data and methodology.

Contents

List of Figures

List of Tables

1 Introduction

1.1. Background information

In the past years, Rijkswaterstaat has observed that the competition within tenders of the Dutch construction sector has decreased (Rijkswaterstaat, 2019). Although the difference is small and only measured on over 250 million Euro projects, the development is in line with recent contractor statements. For example, BAM, one of the largest Dutch contractors, has expressed it will no longer participate in projects of over 150 million euros and in which the contractor is liable for damages or extra costs (Benjamin, 2022). Contractor Heijmans, another Dutch heavyweight, shared in a similar statement it will no longer tender for billion Euro mega-projects (Klumpenaar, 2022).

Contractors point partly at Rijkswaterstaat as the cause of this development, saying the projects allocate too much risk and too little profit to the contractors (Rijkswaterstaat, 2019). This is something that cannot be seen separately from major disputes on risk and costs in recent large construction projects. For example, in 2018 the Botlekbrug needed additional work due to mistakes in the constructive strength calculations, leaving the contractors running at a loss (van Nieuwenhuizen, 2018c). Also in the same year, due to disputes on risks and costs, the contract of Knooppunt Hoevelaken was dissolved (van Nieuwenhuizen, 2018b) and mediation has been invoked in the project Aanpak Ring Zuid Groningen (van Nieuwenhuizen, 2018a).

The examples above are characteristic for the problems with the market dynamics in the Dutch construction sector. It is in the interest of Rijkswaterstaat and the public that the Dutch construction sector is healthy, innovative and competitive (Rijkswaterstaat, 2019). But it is expected that when both Rijkswaterstaat and the contractors keep interacting in the same way, competition may disappear in the Dutch large construction industry (Rijkswaterstaat, 2019). Eventually, this could lead to the failure of future tenders and hinder Rijkswaterstaat in fulfilling its core task, maintaining the Dutch infrastructure.

1.2. Problem statement

The problem is that the relation between the number of competitors in the tender of construction projects and client decisions regarding procurement is unknown. However this knowledge is essential to understand the underlying mechanisms and would contribute to solving the declining competition. So far, the only mentioned factors affecting number of competitors are risk and profit. It must be noted that these factors are solely derived from interviews with contractors and no objective research supports the statements (Rijkswaterstaat, 2019). Quantitative research to validate the qualitative findings has not been conducted yet and therefore a research gap exists.

1.3. Research questions

In an attempt to understand the link between the decreasing competition in the Dutch construction sector and client decisions, the following main research question should be answered:

- **To what extent do client decisions in the procurement of construction projects in the Dutch construction sector affect the number of competitors?**

To guide the research in answering the main research question the following sub-questions should be answered step by step. Sub-question one and two initiate the literature research. While sub-question three starts the explanation of what method is used. Subsequently, sub-question four initiates insight into the used data, after which sub-question five introduces the presentation of the results.

1. What is known about competition within the tender phase of construction projects, and factors affecting this competition?

2. What drives the contractors' bid decision within tenders?

3. Which method could be used to investigate what connection exists between client decisions and tender competition?

4. What data is available to utilize in the research?

5. What results does the model show and which relations are identified between client decisions and the number of competitors?

1.4. Research method

In order to answer the main research question, quantitative research will be performed on historical tender data. This is done in the form of a regression analysis that is used to disclose relations between the number of competitors of a project and factors considered important. The goal of the regression is to predict the number of competitors based on the available data. Depending on the performance of these predictions the calculated regression factors disclose the influence of client decision factors on the number of competitors. For completeness, other environmental factors that are considered important for the procurement process are included in the regression analyses as well.

Given the fact that the number of competitors is count data, i.e. a non-negative integer, performing a regression and modeling meets some specific challenges. First, the model's result must be non-negative. Secondly, the regression model must be able to cope with samples that are concentrated on a few small-number discrete values. The most popular for modeling count data is the Poisson model, especially because of its simplicity (Lemonte et al., 2020). But it is widely known that the Poisson model performs poorly on over-dispersed and zero-inflated data (Valle et al., 2019). Over-dispersion means the variance of the used data is larger than the mean. This generates problems because the Poisson model assumes equi-dispersion, meaning the variance is equal to its mean (Lee et al., 2021). Zero-inflated means the used data consists of many zero counts, also something the Poisson model can not adequately model (Feng et al., 2020). Given that real count data is often over-dispersed, but not by default zero-inflated, the Poisson model is unsuitable in most cases. One often used alternative to account for this over-dispersion is the Negative-Binomial model (Valle et al., 2019; Ver-Hoef & Boveng, 2007). Because it is expected that the to be used quantitative tender data is over-dispersed, the Negative-Binomial model will be used next to the Poisson model.

1.5. Relevance

The relevance of the research is twofold. The result of the thesis could be both of practical and scientific importance.

The practical use lies in the fact that clients in the construction sector could use the gained knowledge in future project procurement. When clients have insight into how decisions affect the number of competitors, they could adjust their tenders to contribute to a financially healthy and competitive market. For example, when it is found that projects above a certain value receive less competition, it could be beneficial to split the projects into smaller projects. Or, when a specific type of work receives a limited number of bids it could be decided to package this work with a project that generally receives a high number of bids.

The scientific use is that no quantitative research has yet been conducted on what drives competition in the construction market. Examples found on the topic of competition and bid decisions mostly base their findings on qualitative data, as explained in chapter 3. Furthermore, the thesis contributes to the utilization of a new data source in research on project procurement, namely TenderNed. Prior to the introduction of TenderNed in 2010 procurement data was not widely available for scientists. The introduction of this specific data source in research brings new research possibilities to the scientific field of procurement in the construction industry.

1.6. Thesis structure

The thesis structure follows the order of the sub-questions stated above. Chapter 2 and chapter 3 represent a literature review about respectively, competition in general plus the effect of client and environment on this competition, and the bid-decision of contractors which contributes to the resulting competition in a tender. The research method and utilized mathematical model are laid out in chapter 4. Chapter 5 describes what data is used in the research. In chapter 6 it is shown what results are gathered from the applied method. Chapter 7 subsequently displays a discussion on the found results. Finally, the conclusions of the research are shared in chapter 8 after which recommendations are formulated (chapter 9).

2 Factors affecting competition

This chapter is written to answer the sub-question ”What is known about competition within the tender phase of construction projects and factors affecting this competition?”. The competition for a given project is the final result of the decisions from both the client and potential contractors. Within the environment, considering economic and legal circumstances, the client decides what is to be procured and how this is done. Taking the same circumstances plus the client’s decisions into account the potential contractors decide to participate or not. This latter decision results in an actual number of bids or registrations on a given project, both representing the number of competitors.

Projects in which competition is used to find a contractor are almost without exemption procured by clients linked to the government. These public clients are obliged by law (“Aanbestedingswet”, 2012) to make use of a tender process in the procurement of goods, services, and works. Depending on the value and the to-be-procured object/service, even specific tender procedures could be prescribed as well (Proportionality guide, (Heurkens et al., 2022)). Due to the fact that this thesis focuses on competition in the Dutch construction industry, this immediately implies the research takes procurement into account from public clients in the ”works” category.

Logically the decisions by the client could be of large influence on the number of competitors. The client has an influence on where a project is constructed, what exactly is built, with what techniques, what the project size is, which contract is used, how long the construction should take, and what tender procedure is applied. The configuration of these decisions determines which contractors are capable of executing the work, and therefore which part of the construction market is able to compete in the tender.

Market circumstances out of the clients’ influence could affect the number of competitors as well. For example, the economic situation, availability of materials, and availability of workers could alter the number of tender participants. The influence could either be directly on a project or indirectly because the full market shrinks or expands. Also, specific legislation could enlarge or shrink the number of construction projects. Just as with the market circumstances this directly influences the tender procedure of a specific project or indirectly has an effect on the construction market as a whole.

In this chapter, a more in-depth explanation of different factors and their possible effects on the competition are laid out. Section 2.1 explains what client factors could do to the competition, while section 2.2 describes the potential influence of environmental factors on the competition. Section 2.3 shows what can be concluded from these insights.

2.1. Client factors

The introduction of this chapter describes a list of client factors that could possibly affect the number of competitors, namely: project location, type of work, project size, contract type, and tender procedure. Subsequently, a more elaborated explanation is provided on the potential effects of the individual client factors on the number of competitors.

Project location

The project location could be viewed in different levels of detail. The most detailed description points towards the exact construction site itself while a more rough description only discloses the country or area of construction.

The exact construction site has influence on the execution of the work, and as a result on the part of the market able to complete the work. For example, when the site is located next to a river or canal this opens the opportunity for maritime service companies to execute parts of the job as well. Thus a larger part of the construction market is able to compete for the work. When a construction site lies in the center of a big city with limitations on site size and traffic delays this has implications on the execution phase as well. Not every contractor will be capable of coping with these limitations.

The designated area of the to-be-constructed project has some other implications for the tender competitors. When contractors are physically located in the same area as the project this gives them some advantages in the allocation of materials, machinery, and workers. For example, in the case of asphalt production facilities, when a contractor is not able to produce near enough to the project it is very hard to complete road works with sufficient quality. Depending on the project area this results in a selection of the construction market being better equipped to complete the project.

Type of work

The type of work is one of the key factors affecting the potential number of competitors. Logically not all contractors are able to install offshore windmill farms or have experience with digging tunnels. How specialistic the job is to select the slice of the construction market that could execute the work. The more general the work the more contractors are able to compete. It must also be noted here that it is common practice for contractors to hire sub-contractors for specialistic work. But this is at the expense of a part of the profit margin which is why contractors who are able to execute the job themselves gain an advantage.

Topics linked to the type of work are the complexity of a project and the demands to apply new construction techniques. The complexity of a project usually refers to how difficult it is to complete the job. This difficulty is generally an accumulation of coping with many stakeholders, a challenging project location, and executing different types of work within a large project. A higher complexity will limit how many parties are able to take the construction assignment. The new construction techniques generally have another effect on the competition. Contractors are often keen to gain experience with new construction techniques to gain an advantage over competitors in future tenders with the same demands. As a result, the number of competitors rises when clients demand new construction techniques with the prospect of procuring future projects with similar techniques.

Project duration

The project duration refers to the time a client allows the contractor to construct a project. This decision could have a variety of implications for the contractor. Most importantly when fines are imposed as a result of exceeding the deadline, this increases the financial risk of a contractor. Especially when the deadline is tight. Another implication could be that as a result of a short project duration a relatively large amount of workers must be employed in a short time span. In the case of a very long project duration, a project could become vulnerable to economic changes.

Project size

The project size could describe the physical size of the project but usually, it refers to the project value. A big project is a valuable project. A higher value has some implications for potential contractors. Firstly, higher values increase the project risk for the contractor. Depending on the remuneration scheme, contractors generally are responsible for financing the work until the work is finished. As a result, contractors must be financially healthy enough to compete for large-value projects. Secondly, higher values result in higher transaction costs. As a rule of thumb joining a tender costs around 2% of the project value (Dudkin & Välilä, 2006). Although public clients reimburse a part of the transaction costs to losing parties this puts financial strain on the competitors. The third implication for higher-value projects is that the complexity increases. As mentioned in the ”Type of work” section complexity rises when the number of stakeholders increases. In large projects the main contractor will have to deal with more surrounding stakeholders, more internal stakeholders (sub-contractors), and often more types of work. This makes large projects demand some specific project management capabilities from their main contractor. The larger the project the fewer contractors are capable to execute the project.

Tender procedure

As stated in the introduction public clients are forced by legislation to apply specific tenders on works depending on the project value. In the Dutch construction market, the guidelines are prescribed in the ”Aanbestedingswet 2012” (“Aanbestedingswet”, 2012), ”Aanbestedingsregelement werken 2016” (Rijksoverheid, 2020), and ”Gids Proportionaliteit” (Heurkens et al., 2022). The legislation uses value thresholds to determine which tender procedure is considered proportional to apply to the procurement. Roughly speaking there are seven tender procedures (PIANOo, 2022a, 2022c, 2022d);

1. National private (1 on 1):

   The client asks a single contractor to execute the work and come up with a cost estimate. This procedure is allowed for projects with a value below 150.000 euros. Also, there is no obligation to select foreign contractors.

2. National multiple private:

   The client objectively asks multiple contractors to submit a bid on a specific work. The project must be awarded to the economically most advantageous bid. The procedure is allowed on projects with a value between 150.000 - 1.500.000 Euro. Just as with the national private procedure, there is no obligation to include foreign contractors in the procedure.

3. National open:

   The client publishes the work and allows the national construction market to submit bids. The project will be awarded based on predefined award criteria. The national open procedure is applied on works with a value between 1.500.000 - 5.382.000 Euro (the EU threshold).

4. National restricted:

   The client publishes the work and allows the national construction market to register for the project. In order to register for the job contractors must meet set requirements. The client subsequently selects up to 5 contractors to prepare an actual bid. This procedure is used when a lot of bidders are expected on the national open procedure. The preselection mitigates the transaction costs for both the client and competitors because fewer bids are made and analyzed. The same value boundaries as with the national open procedure apply to the restricted procedure. Similarly, the project will be awarded based on predefined award criteria.

5. European open:

   The client publishes the work and allows the European construction market to submit bids. The competitors must meet specific requirements in order to bid. Based on award criteria the winning bid is determined. The European procedure must be used for works with a value greater than 5.382.000 euros.

6. European restricted:

   The European restricted procedure is applied with exactly the same conditions as the National restricted procedure with the exception of the project value. This value should be greater than the European threshold of 5.382.0000 Euro.

   A variant of the European restricted procedure is the competitive dialogue. With this variant, up to 5 contractors are selected after registration. But prior to submitting the actual bid all competitors together with the client define what solutions meet the client's demands. After the project requirements are clear to all competitors they are asked to submit a bid.

7. Framework- and concession agreement

   Other frequently used procedures are framework- and concession agreements. Both agreements concern work that is partly undefined, for example, maintaining roads and waterways in specific areas. The type of work is known but it is unsure where the work will take place. The difference is that a concession agreement is entered into with a single party while a framework agreement usually is concluded with multiple parties. Because of the partial uncertainty at the basis of these agreements, they are not included in this research.

Competition is present in the open- and restricted tender procedures (procedure number 3 - 6). The competition especially experiences influence from the decision to tender national or European. Deciding to open the tender for European parties expands the number of potential contractors and possibly increases the competition within the procurement. Whether a project is procured with an open or restricted procedure does not really affect competition. In both cases, the competitors will need to meet specific requirements. To determine the number of competitors in restricted procedures one should look at the number of registrations while with an open procedure, the number of bids represents the competition.

Award criteria

The award criteria are set up by the client and describe how the contractor bids are scored. Depending on the award criteria and incentives contractors could be lured or spooked to join the tender. The client specifies the game and the rules in the awarding process.

The simplest form, lowest-price, scores the bids on value and awards the project to the cheapest contractor. Other common award criteria focus on the best price-quality ratio or best cost-effectiveness. With the application of these two, the client scores what solution will provide the best quality or how the cost-effectiveness is calculated. This is exemplary of how the client is able to influence what solution is applied by contractors and how this solution is executed. (PIANNOo, 2022)

With a fourth type of award criteria, the client offers virtual discounts on the bid price when certain sustainability-, speed-, or emission goals are met. When contractors earn these discounts they could virtually become cheaper than competitors and win the tender. This type of scoring is known as the "Economically Most Advantageous Bid" (EMVI) methodology and is a combination of the award criteria methods mentioned above. (PIANNOo, 2022)

Contract type

The contract type depends on the chosen construction organization form which divides the tasks between the parties involved in a project. Roughly three organization forms exist, the traditional-, the integrated-, and the "Bouwteam" form (Chao-Duivis et al., 2018). Within the traditional form the client solely procures the construction of a project, he will supply the contractor with a design. The integrated form will make the contractor responsible for both the design and construction of a project. In the alliance form both the client and contractor are responsible for the entire project or parts of a project.

Following the decision on the construction organization generally a standard type of contract is used that is derived from uniform administrative conditions specific to the construction organization form. In the Dutch construction sector, traditional contracts are derived from the UAC 2012, integrated contracts from UAC-IC 2005, and "Bouwteam" contracts use specific "Bouwteam" conditions. Following, a list of commonly used contracts is supplied together with their implications and administrative basis (PIANOo, 2022b):

1. C-contract

   A construction contract is a traditional contract based on the UAC-2012 conditions. With a construction contract, just the construction of a project is procured. The client is responsible for the design.

2. D&C-contract

   A design and construct contract is an integrated contract based on the UAC-IC 2005 conditions. Both the design and construction of a project are outsourced to a contractor.

3. E&C-contract

   An engineer and construct contract is an integrated contract based on the UAC-IC 2005 conditions. This contract is especially applicable for projects in which little design work is present. Therefore, just the engineering and construction by itself are sufficient to complete a project. The contractor is responsible for both.

4. DBM-contract

   A design, build, and maintain contract is also an integrated contract based on the UAC-IC 2005 conditions. Within this contract, the contractor is responsible for the design, construction, and maintenance of a work.

5. DBFM-contract

   A design, build, finance, and maintain contract is similar to a DBM contract but additionally gives a contractor the responsibility to finance a project in advance. Specific for the finance part is that contractors start to get paid after completion, during the agreed maintenance time. The contract is based on the UAC-IC 2005 conditions.

6. DBFMO-contract

   A design, build, finance, maintain, and operate contract is an integrated contract based on the UAC-IC 2005 conditions. With respect to DBFM contracts, it gives the contractor the extra task to operate the work during its lifetime. Just as with the DBFM contract the contractor is paid after the work is delivered and functional.

7. Two-phase-contract

   A two-phase contract is an integrated contract based on the UAC-IC 2005 conditions. The core meaning of the contract is that the design and execution phases are separated. The price of the execution phase is negotiated after the design is finished.

8. ”Bouwteam”-contract

   A ”bouwteam” contract describes a contract form in which the client and contractor(s) start to cooperate from the design phase of a project. As a result, the knowledge of contractors is used in the early stage of a project. Within the ”bouwteam” form risks and responsibilities are allocated based on the knowledge and prospected tasks of each party involved. Commonly, the contractors involved in the design phase will execute the work as well. But when the client and contractors can’t get to an agreement, on for example risks and prices, the client could decide to execute the work with different contractors. Specific ”bouwteam” conditions are generally used in this contract form.

Depending on the contract type different risks and liabilities are allocated among the involved parties. These have an impact on which part of the construction market could execute the work and thus the possible competition. Especially contracts involving financial responsibility impact how many contractors are able to take the risk. Also, individual contractor preferences for specific contracts could limit the number of competitors.

2.2. Environmental factors

This chapter describes a list of environmental factors that affect the number of competitors, namely the: economic situation, order-book of the market, availability of building materials, legislation, and contractor relationships. Although the factors are not of direct interest to the main research question they are added to the research for completeness. Every factor possibly affecting the number of competitors should be included in order to make realistic calculations.

Economic situation

The economic situation refers to the amount of money that circulates in the economy. In times of recession, less money is available for investments, also in the construction industry. In economically good times generally more capital is available for investments. Economic favorable circumstances usually result in higher market revenue and low government bond yields. It is found that during an economic crisis, fewer construction projects are procured, and therefore competition increases (Gugler et al., 2015).

Market order-book

The order book represents the amount of already awarded work that the market as a whole has yet to complete. The order book indicates what the need for work is in the construction market. This makes that the order book is linked to how many workers and how much machinery is available for new projects. As a result, the number of competitors could be affected by the order-book value of the market. A well-filled order-book reduces the number of contractors that need work and need to compete in tenders.

Availability of building materials

The availability of building materials could influence the number of competitors. In times of high uncertainty on material availability and prices it could be beneficial for the market to dodge the risk and wait with bidding. For example, recent developments with the war in Ukraine had a large impact on the building material availability. As a result contractors experienced higher costs that heavily affected their profit. This new risk could stop parts of the construction market to tender.

Legislation

Legislation could have a major impact on the number of competitors in construction tenders. A recent example is the regulation on nitrogen emissions in the Netherlands, as a result, many construction projects are suspended or stopped completely. The effect on competition is indirect, due to a decreasing number of projects it is plausible number of competitors will increase. Legislation could also have a decreasing effect. For example, recent intentions from the Dutch government to build more houses could increase the number of projects and decrease competition.

Contractor to contractor relationships

Contractor relationships, in the form of consortia, especially concern the possible number of competitors within large projects. Due to the size of these projects, just a limited part of the construction market is able to compete. When multiple contractors unite within a consortium this decreases the number of potential competitors.

2.3. Conclusion

To answer the sub-question ”What is known about competition within the tender phase of construction projects and factors affecting this competition?”, an extensive overview of factors is provided in this chapter. It is acknowledged that the competition is the sum of contractor bids and that a variety of factors could impact this decision. Within this chapter, the factors were divided into two groups, client factors, and environmental factors. The client factors can be influenced by the client, while the environmental factors are outside of the client’s power. It can be concluded that the client factors that affect competition are: project location, type of work, project duration, project size, tender procedure, award criteria, and contract type. The environmental factors that impact competition are; economic situation, market order book, availability of building materials, legislation, and contractor relationships. The factors above are to be considered when predicting the competition in the construction industry.

3 View on the contractors' bid-decision

This chapter is written to answer the sub-question ”What drives the contractors’ bid decision within tenders?”. Since the competition in a tender is the sum of all contractor bid-decisions (Chua & Li, 2000), it is important to look into this subject. Compared to the subject competition, more research has been conducted on the individual bid-decision. The bid is widely considered the contractors’ most important decision (El-Mashaleh, 2010; Shash, 1993; Wanous et al., 2000). Deciding to but losing the bid leaves a contractor with the loss of tender costs and work, while possibly damaging its reputation (Leśniak et al., 2018; Ravanshadnia et al., 2011). On the other hand, deciding not to bid results in losing the opportunity to make a profit, build a relationship with the client and secure a position in the construction sector (Leśniak et al., 2018; Leśniak & Radziejowska, 2017). Lowe and Parvar take it a step further: ”the decision to bid, as with that of determining the project mark-up, is very important as success or failure of a contractor’s business lies in the outcome derived from those decisions.” (Lowe & Parvar, 2004).

Besides being very important, the bid decision is considered highly complex due to the many affecting factors (Al-Humaidi, 2016) and the inter-relatedness of those factors (Egemen & Mohamed, 2008). The fact that contractors often must choose among several announced tenders and need to prepare their bids in a limited time further complicates the process (Leśniak & Radziejowska, 2017). As already mentioned in the quote of Lowe and Parvar, of equal importance, next to the decision to bid, is defining the bid price (Cheng et al., 2011; Lowe & Parvar, 2004). In the case of winning a contract, this bid price will determine whether the contractor is able to make a profit.

Research in the field of bid-decisions can roughly be categorized into two groups, statistical analyses on bid mark-ups and the chance of winning, and qualitative research on factors considered in a bid-decision. Both research areas tend to come up with a model that could be used by contractors as support in their bid decision.

The statistical research tries to optimize a bid by calculating, based on historical data, the chance of winning a bid given a certain mark-up. Section 3.1 will elaborate on the scientific attempts made in this field. It will be shown that a two-dimensional characterization of projects forms the basis of the calculations and that no distinction is made between other important factors like the type of work, client, or project size. Also, in order to use the models in real life often figures on the bids of competitors are necessary, which are hard to get. Following these insights, it is concluded the statistical models are very hard to use in the actual bidding process.

The qualitative research tries to support contractors in their subjective judgment on tenders of construction projects. Generalized, it is attempted to score the judgment of contractors with a model and affirm or deny this contractor’s decision. A more in-depth view of these models is provided in section 3.2. It will be shown that the bid models are built on simplified, limited, hard-to-get, subjective, or self-fabricated data. Also, the models at best affirm individual contractor preferences. This makes that all models are of very limited use and their utilization rate is low (Cheng et al., 2011; Kalan & Ozbek, 2020). Still, the models disclose what factors are considered by contractors in a bid decision, these factors are presented in section 3.3. It is shown that the factors mentioned in chapter 2 are of interest. Also, it is confirmed that the internal processes of contractors, mainly; the need for work, available workers, and available machinery, influence the bid decision. Additionally, section 3.4 will give a brief overview of contractor considerations out of the scope of the research displayed in section 3.1, section 3.2, and 3.3.

3.1. Statistical research

The statistical research has created models that in one way or another use historical tender data to calculate: the optimal bid mark-up with the optimal chance of winning or just the chance of winning the bid. The sections below will go into a bit more detail on how the calculations are made and what problems influence the results.

Optimal bid considering the mark-up

The pioneers in the bid supporting models, Friedman and Gates, aimed to create a model that would calculate the optimum bid in a closed bidding auction (Friedman, 1956; Gates, 1967).

Friedman uses two types of historic bidding data. Firstly, historic opposition bid patterns are used to calculate the probability of winning against a given opponent. Secondly, historic profit results are used to calculate the probability of making a profit. Those two probabilities can be combined to calculate the probability of winning for a given mark-up, the method aims to optimize for the expected value. In other words, the bid is calculated in which the sum of the probability of winning and the expected markup has the highest value. In order to apply Friedmans' method, extensive data must be available on bidding results and especially competitors bidding behavior. (Friedman, 1956)

Gates used the same principle as Friedman, maximizing the expected value. Also, extensive competitor bidding data and historic bidding results are used to calculate the optimal bid. The difference with Friedmans' method lies in the way the bid is calculated. Instead of using the probability of bidding less than the competitors/winning, Gates finds the probability of beating a known number of competitors. (Gates, 1967)

After the publications of Friedman and Gates controversy arose because they are both the representation of the same problem but generate different results. Multiple papers have been published on whether Friedmans' or Gates' model is correct on the representation of the probability of winning a bid with a given mark-up (Benjamin & Meador, 1979; Crowley, 2000; Skitmore et al., 2007). Benjamin and Meador pointed out that Friedmans' model always finds a smaller optimal markup, and therefore wins bids more often, but it also needs twice the amount of work to obtain Gates' profit (Benjamin & Meador, 1979). Christodoulou explained this by stating Friedman assumes stochastic independence and Gates assumes stochastic dependence on the interaction between bidders (Christodoulou, 2004). In other words, only Gates takes into account that the number of competitors affects the probability of winning. Crowley stated that it seemed neither of the models is correct (Crowley, 2000).

Independently of which model best represents the probability of winning, both models are built on just two parameters, adding to the controversy (Christodoulou, 2004). The question arises if it is even possible to simplify a project into two parameters. It is suggested that this is one of the reasons the models are rarely used in real-world problems and remain in academic circles (Wanous et al., 2000). On top of that, the data on competitors' bid behavior is hard to obtain (Gates, 1967).

Variants of the studies mentioned above exist. For example, Skitmore and Pemberton implemented the possibility to choose between different strategies in the bid calculations. They propose the introduction of three bid strategies, the optimal, no-loss, and break-even strategy. Then for each bidding strategy, it is calculated what strategic markup value should be used in the bid. But this does not solve the problems mentioned above. The same data and similar methods are used as in Friedmans' and Gates' models, and projects thus still are simplified into just two parameters. (Skitmore & Pemberton, 1994)

Chance of winning

Another variant of the bid model was created by Ballesteros-Perez et al. (2015) who step away from using the bid markup in their attempt to calculate the probability of becoming the bid winner. Since the winning bids are not solely selected based on value, Ballesteros-Perez et al. suggest that calculations should not be based on markups from historic bids. They propose that historic bid-winning performances of contractors are the best representative of one's competitiveness and should be used to determine the probability of winning a bid. This new approach still faces the problems mentioned earlier, extensive tender data is needed and the projects still get simplified into very few parameters. On top of that, it should be doubted whether it is meaningful to calculate the possibility of winning without taking a single project characteristic into consideration. (Ballesteros-Pérez et al., 2015)

3.2. Qualitative research

As mentioned in section 3.1 the statistical bid models are rarely used in real-world problems (Kalan & Ozbek, 2020; Wanous et al., 2000). The main reason is that the models do not address the practical needs of contractors (Cheng et al., 2011). Due to this, there was a demand for models that solve this problem.

Following, models were created based on qualitative research, so-called bid decision support models. In general, these models use the judgment of a contractor on a certain contract to come up with a bid/no-bid decision for this certain contract. The models process the judgment of a contractor either directly or with predetermined weight factors on project factors/characteristics. We can roughly divide the models into three categories, models using a multi-criteria approach, data envelopment analysis approach, or neural network approach. The data envelopment analysis approach processes the judgment of the contractors directly, while the other two approaches need predetermined weight factors. These weight factors follow either from interviews/surveys with contractors, as with the multi-criteria approach, or from computations on historical contractor judgments, as with the neural network approach. All three categories will be considered subsequently below.

Multi-criteria approach

With the multi-criteria approach contractors are asked to directly score certain project characteristics of a contract, like; profit (Paranka, 1971), need for work (Ahmad, 1990), market condition (Dawood, 1996), project size (Wanous et al., 2000). These project scores are then processed with predetermined project characteristic weight factors into a resulting project score. Based on a threshold project score, derived from historic bid decisions, a bid/no-bid decision is given.

How these weight factors are determined varies. Some studies directly ask contractors to assign weights to project characteristics (Egemen & Mohamed, 2008; Leśniak & Radziejowska, 2017; Lowe & Parvar, 2004; Paranka, 1971). While others determine the weights based on an Analytic Hierarchy Process (AHP) (Ahmad, 1990; Chua & Li, 2000; Kalan & Ozbek, 2020). In this analytic hierarchy process, the weight factors are determined, by contractors, with a pairwise comparison between project characteristics. Project characteristics that ”win” the most over other characteristics are considered the most important. Another variant in the weight factor determination uses fuzzy logic (Araújo et al., 2022; Cheng et al., 2011; Leśniak et al., 2018; Lin & Chen, 2004; Marzouk & Mohamed, 2018; Ravanshadnia et al., 2011). Fuzzy logic is used to express scores in subjective linguistic terms (Lin & Chen, 2004), this allows contractors to score the importance of certain project characteristics in words. In the calculation of the weight factors, these linguistic terms are processed into numbers.

Despite the different approaches in determining the weight factors of project characteristics in the bid/no-bid decision, they all do the same. Transform the subjective assessment of contractors on project characteristics into numbers that can be used in a support model. Then this model is used to process contractors’ subjective scores on project characteristics. It is very questionable if these models are of any use, they mimic the bidding behavior of the interviewed contractors and do not account for the subjective basis of the research.

Data envelopment analysis approach

The Data Envelopment Analysis approach differs from the multi-criteria approach because no weight factors have to be determined upfront, eliminating a subjective step. Instead, contractors' scores on project characteristics are directly used in a visualized bid/no-bid advice (El-Mashaleh, 2010, 2013; Polat & Bingol, 2017). Still, the results are subjective, but solely based on the contractor that uses the data envelopment analysis approach.

The method first pares negative and positive project characteristics and asks contractors to score each characteristic. This will lead to two values for each pair, a negative score, and a positive score. The ratios between the positive/negative score are then placed on a multi-dimensional plane with historical bids. When the "new contract" finds itself within the boundaries of previous bids, the so-called "Favorable Frontier", the advice to bid is given (El-Mashaleh, 2010). When the contract finds itself outside of the boundaries, the advice not to bid is given. If the contractor still decides to bid the "Favorable Frontier" is replaced due to the new bid decision. The data envelopment analysis helps individual contractors to bid in a consistent way.

The application of the data envelopment analysis is rather limited. It is only of use within a contractor's own organization and only shows if a new bid is according to previous bids. Changing perceptions of project characteristics could make the use of historic bid scores worthless. On top of that, the model will probably just affirm the contractors' decision to bid.

Neural network approach

The neural network approach is very similar to the multi-criteria approach, apart from the determination of the weight factors for each project characteristic. The weights are calculated by the network based on historic bid data (Dias & Weerasinghe, 1996). But the input for the model is the same, contractors will have to supply the scores on project characteristics from a certain contract (Shi, 2012; Shi et al., 2016; Sonmez & Sözgen, 2017; Wanous et al., 2003). Al-Humaidi allowed fuzzy input in the neural network, but this is still a subjective source (Al-Humaidi, 2016). Because of the statistical capabilities and the many nodes in a Neural Network, the model is able to mimic complex behavior.

Still, the application of a neural network, or any other bid-supporting model, on subjective data raises an important question. What is the use of a model when it tells exactly the same as a contractor? The model is trained to follow the gut feelings of the contractors that supplied the training data. So, the model will affirm each decision. The model of Shi even states it reached an alarming 100 percent accuracy on the training set (Shi, 2012). The models only tell us something about the contractors it has cloned. To what extent the project characteristics generally affect a bid decision is still unknown.

Usability of bid decision support models

It is shown in the parts above that all models in one way or another depend on the subjective input from contractors. As a result, the models are not representative of the market and at best just clone the bidding behavior of a few individual contractors. Besides, one could question whether the models are of any use when they in fact just affirm the judgment of the contractor that makes a decision. Although subjective data forms the basis of the studies, the models that use weight factors on project factors do give an insight into the deliberations of contractors. The weight factors of the project factors indicate which project factors are considered important and should be accounted for in this thesis. The project factors that are distilled from the qualitative research are presented in section 3.3.

3.3. Factors considered in qualitative research

As stated in section 3.2 the bid support models are not that useful. But parts of the research in the development of the bid support models help to gain insight into the deliberations of contractors in the bid decision. All bid support model studies attempt to score project characteristics or bid decision factors so they can process contractors' scores on individual projects. The results on which factors are considered by contractors are of use in this research. It was found that 10 factors could best describe the contractors' deliberations, namely: Project type, Project value, Project duration, Project location, Client, Contract type, Market conditions, Need for work, Tender method & Competition. Interestingly award criteria are not explicitly mentioned by contractors. It was also found that in past research no unambiguous view exists on which factor is most important. How articles rank the decision factors on importance varies heavily.

Factors mentioned in articles

Kalan & Ozbek reviewed seven articles on bid decision factors. Out of 100 considered factors, a list of the 14 best-describing and most-used factors was determined (Kalan & Ozbek, 2020). Ravanshadnia et al. reviewed 15 articles on bid decision factors. After which they summarized a list of 103 factors into a list of 25 bid decision factors (Ravanshadnia et al., 2011). Cheng et al. reviewed 10 articles to identify the most mentioned factors of interest. This resulted in a list of 44 found factors (Cheng et al., 2011). Mehrabani et al. also reviewed 10 articles and created a list of 72 mentioned bid decision factors (Mehrabani et al., 2020). It must be noted that the four articles often reviewed the same articles or authors. For example (Chua & Li, 2000), (Shash, 1993), and (Wanous et al., 2000), are cited in all four articles. Therefore, those articles heavily impact the view on factors considered most important.

The main difference between the articles lies in the detail in which factors are described. Where Ravanshadnia et al. describe the factors "Site condition", "Suitable climate and weather and geographical conditions", and "stabling laws, standards, and requirements", Kalan & Ozbek describe effectively the same factors more concisely as "Project location". In fact, many mentioned factors mean the same. Another example, Mehrabani et al. mention the different factors "Availability of the needed human resources", "Availability of required equipment", "Availability of required staff", and "Need for work in company" while they are all interrelated (Mehrabani et al., 2020). Because, when there is an availability of human resources, staff, and equipment there is a need for work. Therefore, the factor "Need for work" covers the factors more concisely.

When we consider most factors could be described more concisely, it is found that the list of Kalan & Ozbek shown in table 3.1 is rather complete, and a good summary of the other articles. But even the 14 mentioned factors could be summarized more. For example, factor numbers 1, 3, and 5 could all be classified as "Market conditions", mentioned in the research of both Mehrabani et al. and Cheng et al (Cheng et al., 2011; Mehrabani et al., 2020). Also, "Time to tender", mentioned in the research of Ravanshadnia et al. and Cheng et al, should be added as "Tender method" to the final list (Cheng et al., 2011; Ravanshadnia et al., 2011). As a result, the factors of interest for contractors could be best described in 10 factors: Project type, Project value, Project duration, Project location, Client, Contract type, Market conditions, Need for work, Tender method, and Competition. The resulting list of bid decision factors is displayed in table 3.2.

Table 3.1: Bid decision factors from Kalan & Ozbek (Kalan & Ozbek, 2020)

Factor number:	Article factor description:
1	Current workload
2	Experience with similar projects
3	Availability of equipment, materials, and human resources
4	Financial ability
5	Need for work
6	Technical know-how
7	Compliance with business plan
8	Project size
9	Project duration
10	Project location
11	Project type
12	Contract conditions and type
13	Owner identity
14	Competition

Table 3.2: Summary of representative bid decision factors

Factor number:	Factor description:
1	Project type
2	Project value
3	Project duration
4	Project location
5	Client
6	Contract type
7	Market conditions
8	Need for work
9	Tender method
10	Competition

3.4. Non rational factors

It is also found that the bid decision is sometimes made without reasonable basis and in a subjective manner (Egemen & Mohamed, 2008). Ahmad affirms this, he states it is common practice to make the decision based on a mixture of gut feelings, experience, and guesses (Ahmad, 1990). Due to the need to anticipate on the process of the project, it is logical that gut feelings, experience, and guesses play a role in the bid decision. But the articles mentioned also give the insight that irrational behavior could influence the competition within certain tenders. This is important to note because this thesis approaches the topic with a rational view.

One could imagine a few reasons for a contractor to bid irrationally. Firstly, a project could be a prestige project, building skyline changing structures, or water defense works known by everyone is a good advertisement. Secondly, a contractor could bid just to get in the waters of main competitors. In the case it is known a main competitor desires a specific contract, contractors could even decide to win the project at all costs. A third option is, in the case no work is needed, that a contractor just joins tenders to supply work to the tender department. The fourth and ultimate irrational option is a contractor just decides to bid.

3.5. Conclusion

This chapter is written to answer the sub-question ”What drives the contractors’ bid decision within tenders?”. It is found in the literature that there are ten different factors that are considered essential by contractors to decide to bid on a project. The factors are: Project type, project value, project duration, project location, client, contract type, market conditions, need for work, tender method, and competition. These factors affirm the client and environmental factors mentioned in chapter 2, except for the award criteria. Taking all factors found until this point pretty much describes all that could be known of a project prior to awarding. Besides these findings, it must also be noted that contractors do not always bid rationally and irrational motives could be decisive in the bid decision as well.

4 Research method

This chapter is written to answer the sub-question ”Which method could be used to investigate what connection exists between client decisions and tender competition?”. It is already disclosed in chapter 1 that regression analyses will be performed on the available project data. The main goal of the regressions is to predict the number of competitors for a project based on its characteristics, i.e. the factors covered in chapter 2. Because the number of competitors is a count-value, specific regression models must be utilized. It is found that the Poisson- and Negative Binomial regression models are most commonly applied to cope with this type of data. Depending on the performance of the models the resulting regression coefficients calculated in the analyses disclose the individual effects of variables on the predicted number of competitors. It is attempted to include a variable for each factor considered important, either because the variable is connected with a client decision or an environmental factor.

Two steps are added to the research to validate the performance of the models and the quality of the available data. The performance of the models is validated with cross-validation, a method that solves the chance of reporting a coincidental good model performance. The quality of the data is validated with an analysis of internal mutual information and entropy. This will disclose how much ”chaos” is present in the data and whether the information available in the data is able to unravel this ”chaos”.

All analysis steps are explained in this chapter. Section 4.1 will display a general overview of the steps taken in the analysis. Subsequently section 4.2 will explain more in-depth what the Poisson- and Negative Binomial models are and how they operate. After this, section 4.3 will disclose how the results of the regression analyses are further validated. Following, section 4.4 explains how the quality of the data is validated. Finally, section 4.5 concludes on the information provided in this chapter.

4.1. Steps within a regression

A regression is applied to identify the relationship between a dependent variable, in the case of this thesis the number of competitors, and independent variables, the decision factors, or project characteristics. The relation is expressed in regression coefficients, these can be used to predict the number of competitors based on the independent variables of a given project. How well these predictions match the actual numbers depends on the data, model, and predictability of the number of competitors. This section focuses solely on the general process of a regression analysis. More in-depth insights into the model are explained in section 4.2. Views on the available data and its predictability are shared in respectively chapter 5 and chapter 6.

Data preparation

Prior to a regression analysis the data must be cleaned. This means that for all events, in the case of the projects in this thesis, both the dependent variable and matching independent variable must be matched. Events that miss a value for the dependent variable must be dropped. Events that miss certain values for independent variables can be kept in the data set, only when the specific independent variables are used the events are automatically left out of the regression.

For example, in the preparation independent variables must also be converted in a way that the regression is able to cope with the data. For example, categorical variables must be split into dummies. For example, a column with a variable either ”Large”, ”Medium” or ”Small”, is split into three columns assigned with ones and zeros. An event with a original variable ”Large”, will be assigned a one in the new ”Large” column and a zero in the new ”Medium” and ”Small” columns. This way categorical data is converted into numerical data.

When the data is cleaned and prepared and only a list of usable events is left, the data must be split into different sets. These different sets, the training-, validation-, and testing set are needed in order to determine the model performance. Regression coefficients are calculated with the training set, subsequently, predictions are made on the validation set. When the performance on the validation data is according to desire the model performance is calculated on the test data set. It is important that the split labels are assigned to the events at random. In the labeling a 80-10-10 split is used, resulting in 80% of the data being training data. Other splits exist as well but the application of an 80% training split is common (Rácz et al., 2021).

Regression process

The regression process, disregarding the used model, subsequently follows the steps below:

1. The input data is selected. This could be the entire cleaned data set or a selection based on certain values/variables.
2. It is chosen which independent variables are included in the regression. Based on this decision events in the data could be dropped if they miss certain independent variable values. It must be noted here that the training-events/variables ratio must not be less than 10/1, this is a rule of thumb meant to prevent over-fitting (Pavlou et al., 2015). This phenomenon results in an ineffective model, with good performance on the training data but poor performance on the validation- and test data.
3. A regression is executed on the training data with the chosen variables. The resulting regression coefficients for all variables are stored.
4. The regression coefficients are used to predict the dependent variable on the validation data. Following it is determined how well the predictions match the actual values. The accuracy of the predictions represents the performance of the model, variables, and data combination.
5. Steps 1-4 are applied in an iterative way until a set-up is found of the model and data with satisfactory accuracy. When this setup is found once more the accuracy is calculated but now on the test data. The accuracy of the test data is reported as set-up performance. The regression coefficients are reported to supply insight into the relation between the dependent and independent variables.

Connection with thesis

The dependent variable in this thesis is the number of competitors. The independent variables are values that represent decision factors and environmental factors. In the iterative regression process every time a selection of the data and variables are used. Subsequently, the accuracy of the regression is calculated on the validation data.

The accuracy is the percentage of projects at which the number of competitors is predicted correctly. A prediction is considered correct when the actual competition value is within the range of 1 count of the rounded prediction. In other words, when the rounded prediction differs by a max of 1 count from the actual competition the prediction is considered good enough.

4.2. Count data regression models

As already mentioned the dependent variable in this research is the number of competitors, meaning the values are non-negative integers. Due to the characteristics of count data, performing a regression analysis meets some specific challenges. The regression must be able to cope with sample data that are concentrated on a few small-number discrete values and the predictions must be positively valued.

The most commonly utilized count data regression model for this occasion is the Poisson model. An advantage of the Poisson model is that it is considered the simplest count data regression model, the disadvantage is that the model is not able to cope with over-dispersion. This phenomenon, describing the variance of the used data is larger than its mean, is often found in count data. As a result, an alternative model must be used, a model that is commonly applied to cope with over-dispersion is the Negative Binomial model. Instead of assuming equi-dispersion, meaning the mean is equal to the variance, it allows the variance and mean to be different. How much this second model allows a difference depends on the results of the first model, which is used as the foundation for the calculations. (Cameron & Trivedi, 2013)

Poisson model

At the basis of the Poisson model lies the Poisson distribution. This is a single parameter distribution that remains positive and is able to deal with small numbers of samples. As shown in figure 4.1 the shape of the distribution changes depending on the inserted parameter λ. What could also be seen is that the lambda represents the mean of the distribution, i.e. the location of the peak of the distribution. Another property of the distribution is that the mean is equal to the variance. This is something that also showed in figure 4.1, when the λ increases the width of that distribution increases accordingly. What must be known to understand the Poisson model is that in the training operation of this thesis, the model will fit this distribution to predict the number of competitors of each individual training project. In other words, the model will determine a λ that represents the number of competitors for each training project. This λ is the direct result of the multiplication of independent variable values of a project and the generated regression coefficients. Due to the model characteristics, the model will notify a higher score when the λ is close to the actual number of competitors and therefore know its predictions improve.

The model does this with a large formula that contains all the training projects information. This formula calculates the joint probability that all predictions of all projects match the actual number of competitors. This joint probability will generate a high value when the predictions are close to the actual counts and a lower value when the predictions are far off. The model influences the predictions with the regression coefficients in an iterative process. The model will try different regression coefficients until it reaches the maximum achievable joint probability. The final regression coefficients will generate the best possible predictions. These resulting regression coefficients are the output of the model and they are used to make predictions on validation and test data.

The calculations of the Poisson model are executed by a Statsmodel package in the programming language Python. The calculations are too extensive to do by hand, especially due to the necessary iterations.

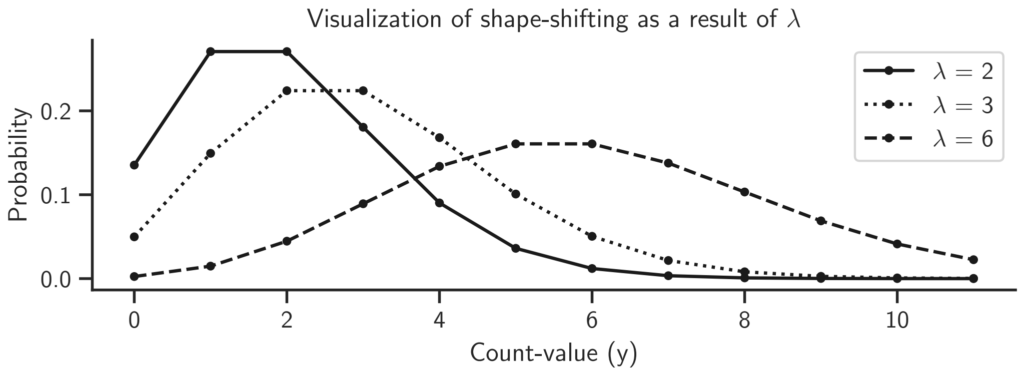

Figure 4.1: Visualization of Poisson distribution

Negative Binomial model

The Negative Binomial applies the same principle as the Poisson model. A distribution is fitted on all individual training projects and the regression coefficients are optimized in the iterative process. The final regression coefficients are then used to make predictions. The difference lies in the distribution that is fitted. The Negative Binomial model uses a distribution of the same shape as the Poisson distribution but without the equi-dispersion property. Instead, an altered Poisson distribution is utilized that allows the distribution to be wider or narrower. The alteration is a result of the standard coefficient α. Meaning an equally more wide or narrow Poisson distribution is used on all training projects. This α coefficient is calculated based on the Poisson regression results, it is determined with what alpha the Poisson predictions accuracy would have increased. The best-performing alpha is used in the Negative Binomial model.

Just as with the Poisson model the calculations are executed by Statsmodel packages, both the α calculations and Negative Binomial model calculations.

4.3. Validating model performance

The results of a single regression are reported in the form of regression coefficients and a prediction accuracy. To verify how good the accuracy of the model is some additional research must be conducted. The cross-validation paragraph below explains how the method is used to redetermine the model accuracy. The constant guess paragraph shows, with the help of the constant guess accuracy, how well the regression model performs.

Cross-validation

As explained in section 4.1 the initial results of a single regression are obtained from 10% of the data. Due to chance, it is a possibility the model accidentally performs above average well. To counter this phenomenon k-fold cross-validation is applied to the regression set-ups of interest. With this method the regression is redone multiple times on different ”folds” of the data, the reported accuracy is the average of these regressions.

In general, the data is split into 10 folds (Govindarajan & Chandrasekaran, 2010), which is why in this research also 10 folds are used. In the case of ten folds ten regressions are executed. Every time a new fold is used as test-data and the other nine folds as training data. Figure 4.2 below gives an impression of the k-fold cross-validation. The result of the cross-validation is an average model accuracy of all ten regressions.

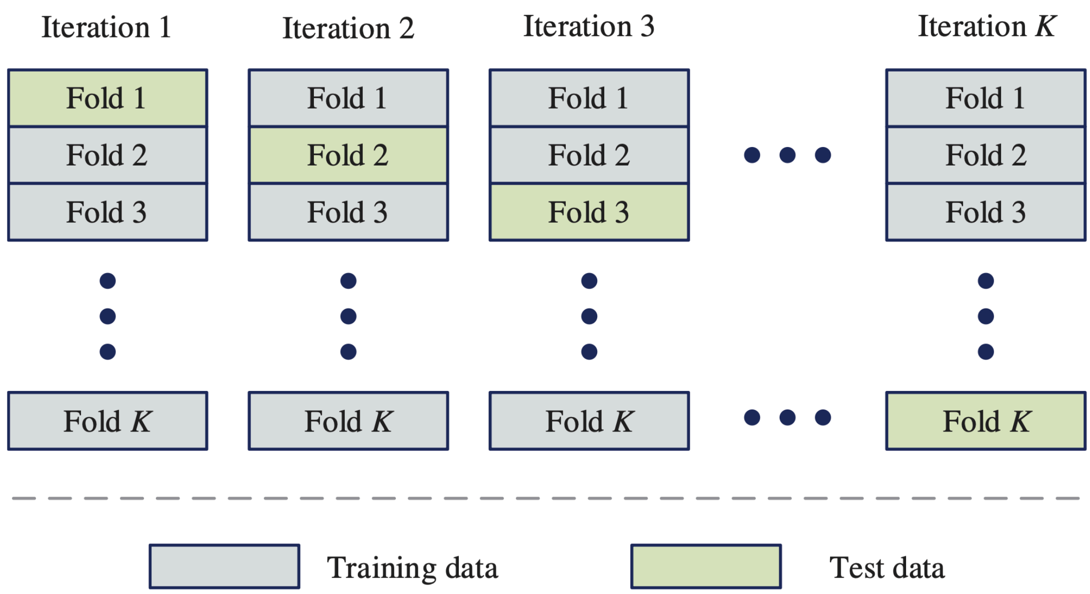

Figure 4.2: Visualization of K-fold cross validation, retrieved from (Ren et al., 2019)

Constant guess performance

In order to find out whether the regression performance is any good, the model accuracy is compared with the performance of a constant guess. The mode of the dependent variable of the specific data used, i.e. the most occurring number of competitors, is taken as the guess value. The accuracy of the guess is calculated in the same way as the prediction accuracy mentioned in section 4.1. The same range of 1 count is allowed for this constant guess. With the gained information it can be calculated how much percent on average the model performance differs from a constant guess performance. The constant guess performance is also calculated in the cross-validation so the average performance of the model can be compared with the average guess performance.

4.4. Validating data quality

The regression models will show a certain prediction accuracy. When this accuracy is considered poor the additional question arises if this is due to the chosen model or the utilized data. A measure for the quality of the data is the entropy and mutual information. These values help to indicate how hard it is to make predictions with the available data. (Cover & Thomas, 1991)

The entropy is a measure of the ”unpredictability” of a random variable. The higher the entropy value the more ”chaos” is present within that variable and the harder it is to make predictions on this variable. Within this research, we are interested in the entropy of the dependent variable number of competitors. The mutual information is a value calculated between two variables and indicates how much one variable tells about the other variable. In this research, the mutual information is used to investigate how much each independent variable tells about the dependent variable number of competitors. The unit of mutual information is comparable with the entropy. This makes that it could be compared if all independent variables’ mutual information combined add up to the value of entropy. Meaning, is all information available in the data, assuming the variables are completely independent, able to solve the chaos of the dependent variable. When the mutual information combined does not come close to the entropy value, the dependent variable with the given data is very likely unpredictable.

Both values are calculated within the programming language Python by utilizing Statsmodel packages.

4.5. Conclusion

This chapter is written to answer the question ”Which method could be used to investigate what connection exists between client decisions and tender competition?”. It is explained that both the Poisson and Negative Binomial are suitable to investigate relations between the competition in projects and available independent variables. But it is also explained why supplemental research is added to investigate model performance and data quality. In order to verify the performance of the model, cross-validation is applied, compensating for the chance of accidentally reporting extraordinary results. In order to verify whether the model performances are according to expectations, the entropy and mutual information within the data are calculated. These values will disclose whether the information between the number of competitors and its independent variables is able to account for the ”chaos” within the number of competitors distribution.

5 Available tender data

This chapter will answer the sub-question ”What data is available to utilize in the research?”. In order to investigate whether it is possible to predict the number of competitors in the tender phase of construction projects a variety of historical data is desired; a list of projects, corresponding project characteristics, and corresponding values for the actual number of competitors. The data that is used in this research is gained from four sources; the Dutch tender platform TenderNed, the Dutch economic research institute for construction EIB, the Dutch central bureau for statistics CBS, and a financial market monitoring website called Investing.com. The TenderNed data forms the basis of the research since the list of investigated projects, together with matching competition and project characteristics values, is distilled from this source. The other three sources are used to enrich the project list with extra information on the market conditions during the tender phase of each project. Section 5.1 will describe how the project list is distilled from TenderNed. Section 5.2 displays all information that is gathered from TenderNed while section 5.3 explains what additional market condition information is added to each project. In this research, it is attempted to include the found decision- and environmental factors distilled in chapter 2. How it is attempted to link the found tender data to these bid factors is explained in section 5.4. A more complete overview in appendix B displays the information in the completed tender data set in key numbers and graphs.

5.1. Creation of the project list

As mentioned in the introduction the TenderNed data is the main source of the data used in this research (TenderNed, 2022). TenderNed is an online platform used by mainly Dutch government bodies to both guide and publish the procurement process of deliveries, services, and works. The bodies are obliged to use TenderNed when the value of the to-be-procured goods exceeds certain thresholds which are explained in chapter 2. This threshold varies depending on what is procured, for works, this threshold is 5.382.000 Euros. In practice, many goods below this threshold are published on TenderNed as well because clients choose to use the platform to guide the procurement process. These clients that use TenderNed are almost without exception of governmental origin, generally, the client is a ministry, central governmental executing body, province, waterboard, or municipality. Other clients of indirect governmental origin are for example the Dutch rail company ProRail, the Port of Rotterdam, Schiphol Airport, and energy supplier Nuon.

When clients use TenderNed for their procurement process, the platform will post information regarding the tender on its website in the form of notifications. The notifications cover the procurement processes from the initial announcement to the final award. The detail of the information available on each project grows with each notification. The information present in the notifications stretch from specific delivery dates to the tender description and winning bidder.

TenderNed makes all the notification data available in large tables shared via Excel files, in total the files cover more than 225.000 notifications. Within this research, we are only interested in tenders that fall in the category ”Works”, represent a project, consist of both an announcement and award notification, and show the number of competitors. The flowchart of figure 5.3 shows which criteria led to a ”notification list” with only announcement and award notifications of ”Works” projects. Figure 5.1 below gives an impression of the output of the flowchart. It is shown that a list is created that contains only announcement and award notices, in this example two fictional projects are displayed that can be identified by their TenderNed label.

Notification number	TenderNed label	Date publication	Type of notification	#	#	#	#	#
1	88358	01-01-2023	Award of project					
2								
3	88358	01-01-2022	Announcement of project					
4								
5								
6	88472	11-12-2021	Award of project					
7								
8								
9								
10	88472	05-06-2021	Announcement of project					

Figure 5.1: Notification list output of flowchart

The notification list in its turn is processed to become a "project list" in which each project, represented by a unique TenderNed label, appears once. This project list is created by combining the information of all project announcements and awards. The most important information from the initial announcement is the publication date, which will disclose the tender period of the specific project. The award notification, being the most complete in terms of information, will reveal which party won the tender and how many parties were interested in the project. An overview of all other information that is gathered from TenderNed will be given in section 5.2. The goal of this section is to show how the TenderNed data is transformed into the project list. Figure 5.2 shows what the transformation of figure 5.3 into a "project list" looks like.

Project number	TenderNed label	Initial publication date	Award publication date	#	#	#	#	#
1	88358	01-01-2022	01-01-2023	[redacted]				
2	88472	05-06-2021	11-12-2021	[redacted]				
3	[redacted]							
4	[redacted]							
5	[redacted]							

Figure 5.2: Impression of resulting project list

It was previously mentioned that the original data set displays over 225.000 TenderNed notifications. Around 28.000 out of these notifications were related to "Works" projects in which almost 12.000 unique tenders were registered. In the end, a little over 2600 projects have been selected to be used in the research. The fact that more than 9.000 projects are "lost" in the selection process has four potential causes. The first and most important one is that of only around 6000 projects an award notification is present in the data. It is likely the other project awards have not been published because it was not obliged due to the project value or the projects were not awarded at all. The second most important cause is that the majority of the projects do not show the number of competitors, essential for this research. The third cause is that a part of the projects has no official initial announcement. Possibly because it is not obliged for the specific projects. The fourth and final cause is that it is possible the projects are not registered properly within TenderNed.

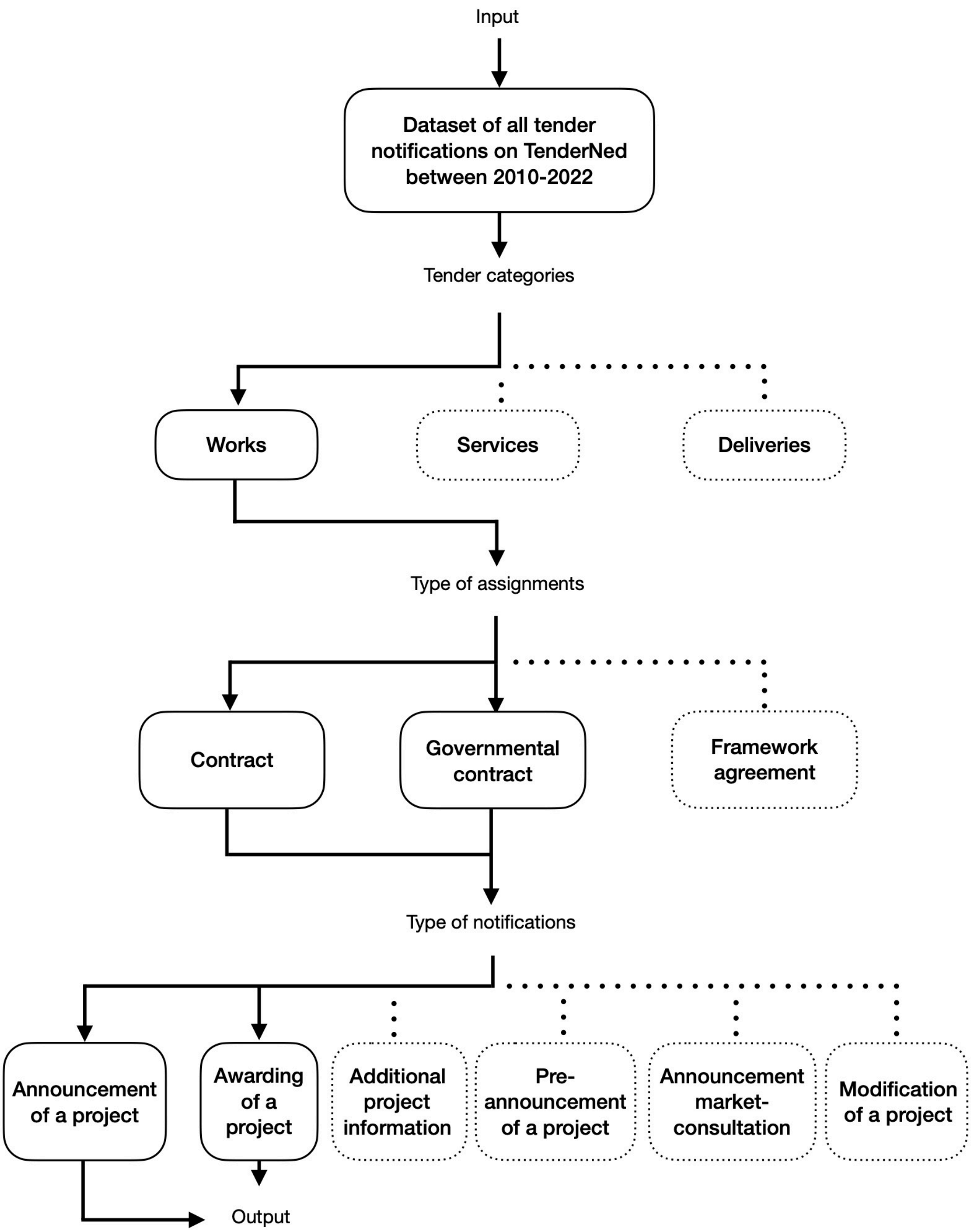

Figure 5.3: Flowchart to describe the selection process of relevant TenderNed notifications

5.2. Collected TenderNed information

Of the 2600 selected projects TenderNed displays a variety of information. Below a concise overview is supplied of all the information that is extracted, including some background details if necessary. The most important value is the number of competitors for each project. The other information will be included in the research as independent variables.

Competition and Tender procedure

The number of competitors arises from either the number of bids on a given project or the number of registrations for a given project. TenderNed gathers this information depending on the used procedure. Three different procedures are present in the used data set, namely:

- Public procedure
- Restricted procedure
- Competitive-dialogue procedure

In the case of a public procedure, the number of bids is used as the number of competitors. When either the restricted- or competitive-dialogue procedure is applied the number of registrations is used as the number of competitors. This is done because, with the application of restricted or competitive-dialogue tender procedures, the number of bids is a result of who is invited to bid. Prior to this invitation contractors explicitly have to state whether they would like to be invited to bid by registering as a candidate. The number of registrations is more representative to use as a value for competitors than the actual number of bids.

With regard to the tender procedure, two variables are present in the data set, firstly whether a National or European procedure is applicable, and secondly if a public, restricted, or competitive-dialogue procedure is applied. Which procedures are chosen by the client is directly connected with the contract value. Due to the obligation to apply the Dutch Proportionality Guide (Heurkens et al., 2022), the thresholds shown in figure 5.4 prescribe the to-be-used procedure for tenders of works. In the data set, only projects with a restricted or competitive-dialogue procedure are included that also show a value for the number of registrations. The largest value present of either the number of bids or the number of candidates is taken as competition value for the number of competitors.

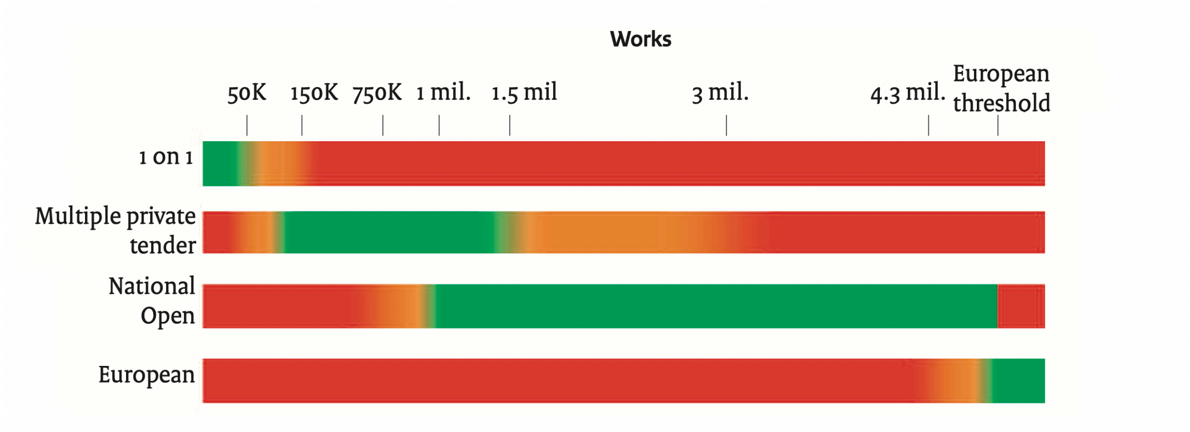

Figure 5.4: Thresholds guiding the decision on tender procedures, retrieved from (Heurkens et al., 2022)

Project size/value

Instead of the contract value a project size label is assigned to each project. The project size is a label based on certain ranges in contract value. The determined ranges are explained below, six labels are used: Very small, Small, Medium, Large, Very large, King size. The contract value is based on either the reported, estimated value, final value, or lowest offer, in that order and dependent on availability. Because the project size uses a value range, the differences between estimated and actual contract values are less of an influence in further analysis. Also, in the regression analysis projects with a very high value weigh heavily on resulting factors while not being very common in the data set. Therefore the classification might fit better with the analysis.

The classification is partly based on thresholds in the procurement legislation, partly on articles, and partly on gut feeling. Because according to the Dutch procurement guidelines it is considered proportional to use a one-on-one award on works with a value below 150.000 Euro. Therefore, those projects are not labeled. The projects with a value above 150.000 Euro are labeled according to the following ranges:

- Very-small: 150.000 - 1.500.000 Euro
- Small: 1.500.000 - 5.382.000 Euro
- Medium: 5.382.000 - 10.000.000 Euro
- Large: 10.000.000 - 35.000.000 Euro
- Very-large: 35.000.000 - 100.000.000 Euro
- King-size: 100.000.000 and up

The thresholds of the Very-small projects fall in a range in which a private award is considered proportional. Small projects represent a range in which National procedures are most commonly used. The bottom value of Medium projects represent the European tender procedure threshold. The boundaries of the Large projects are set based on gut feeling and to create two extra steps between Medium and King-size projects. The King-size projects bottom boundary is determined based on news articles in which contractors explain that they will not bid on project from 100 million and up, mentioned in chapter 1. Projects that do not contain any a contract value or show a value lower than 150.000 are not labeled with a project size. (Heurkens et al., 2022)

Contract type

TenderNed does not disclose directly which contract is applied within each project. But it does display whether the project is a Construct only or Design and Construct project. It is very likely but not certain that respectively a UAV or UAV-GC contract is applied with the two options mentioned above. It must also be noted that the projects in the resulting data set from the year 2016 onward rarely suggest what type of contract is used. Therefore, when this variable is used in a regression it eliminates most of the projects after 2016.

Tender period

The tender period is calculated as the number of days between the initial announcement date on TenderNed and the displayed closing date for the specific tender. This figure is highly connected with the chosen procedure since minimal tender periods are prescribed per tender procedure. With the closing date of each tender also the year of tendering is disclosed.

Project duration

The project duration is calculated as the number of days between the starting and end date of each project. This number is the prognosis made when the tender is initiated.

Client type

Five types of clients are present in the data set, being mostly of governmental origin. The labels as assigned represent the sphere of influence of each client. The following labels are used:

- Federal (Ministry/Central-government)
- Regional (Province/Water-board)
- Local (Municipality)
- Public (Education organization/Hospital)
- Special (Port/Rail operator/Energy company/Garbage company)

Type of work

The type of work executed in a project is described by a CPV code (Common Procurement Vocabulary). This is a code that is used in the procurement of goods, services, and works, to describe what is procured. The description of a CPV code varies from very detailed (”Construction of a mobile telephone base station”, CPV 45232340-7) to very general (”Construction works”, CPV 45000000-7). Also, many CPV codes differ while they describe the same work. For example, more than ten different CPV codes describe the construction of a road. As a result, the need arose to regroup CPV codes into groups with similar work. Most categories are self-explanatory, utility work describes buildings not meant for living while civilian work represents the construction of housing. Note that the group general work represents rather unspecified projects. The 10 less detailed categories of work that have been used are listed below, it is possible that the original CPV codes match multiple categories:

- Installation
- Bridge
- Tunnel
- Utility
- Civilian
- Rail
- Water
- Road
- Ground
- General

Project location

The nuts code labels the economic area in which a project is executed, a good representation of the project location. Multiple levels of detail exist, level 3 (municipality level, most detailed), level 2 (province level), and level 1 (country level, least detailed). The level 1 notation used in this analysis contains 4 areas, also shown in figure 5.5: NL1 for North-, NL2 for East-, NL3 for West- & NL4 for South-Netherlands. When no clear area is specified the unspecified nuts code NL is often assigned to the project.

Figure 5.5: Nuts codes that represent the Dutch economic zones, retrieved from (Commision, 2020)

5.3. Additional market condition information

Since no values for market conditions are included in the TenderNed data set additional data had to be collected. The additional market information is gathered from three sources: EIB, CBS, and investing.com (CBS, 2023; EIB, 2022; investing.com, 2023). The EIB reports the revenue of the construction market and the revenue of the GWW sector (de Lange & Visser, 2021; Visser, 2014, 2015, 2016a, 2017a, 2020a; Visser & Nicolas, 2018, 2019; Vrolijk, 2013), shown in figure 5.7. Next to that, the EIB publishes the reported order book of the construction market as a whole (Kok & Visser, 2019a, 2019b; Straatmeyer, 2015; Visser, 2016b, 2017b, 2017c, 2018a, 2018b, 2020b, 2020c, 2021a, 2021b, 2022a, 2022b; Visser & Straatmeyer, 2016; Vrolijk, 2014, 2015), displayed in figure 5.8. The CBS reports the number of open vacancies in the construction sector (CBS, 2023), shown in figure 5.9. Figure 5.6 shows the monthly average bond rate on NL 10 year bonds collected from investing.com (investing.com, 2023). All figures mentioned above show a similar trend, an unfavorable market between 2012-2016 and a favorable market from 2016 and on. In the figure 5.6 below this is expressed in a relatively low bond yield from 2016 to 2022. The other figures all show a rising trend from 2016 on. The difference between the figures lies in the interval in which the data is collected. The EIB revenue values are collected annually, the EIB order book values semi-annually, the CBS vacancy data each quarter, and the bond rate is daily. Therefore all variables are included as project characteristics. The latest available market data on the closing date of each individual tender is assigned to that specific project.

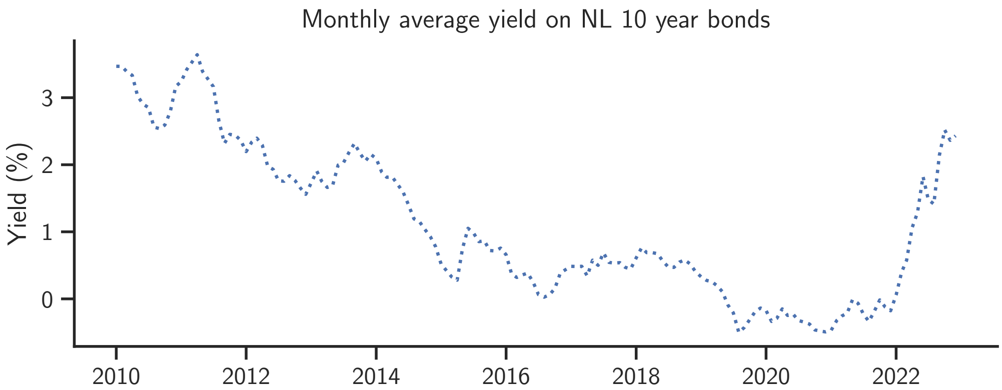

Figure 5.6: Bond yield on Dutch 10 year bonds over time

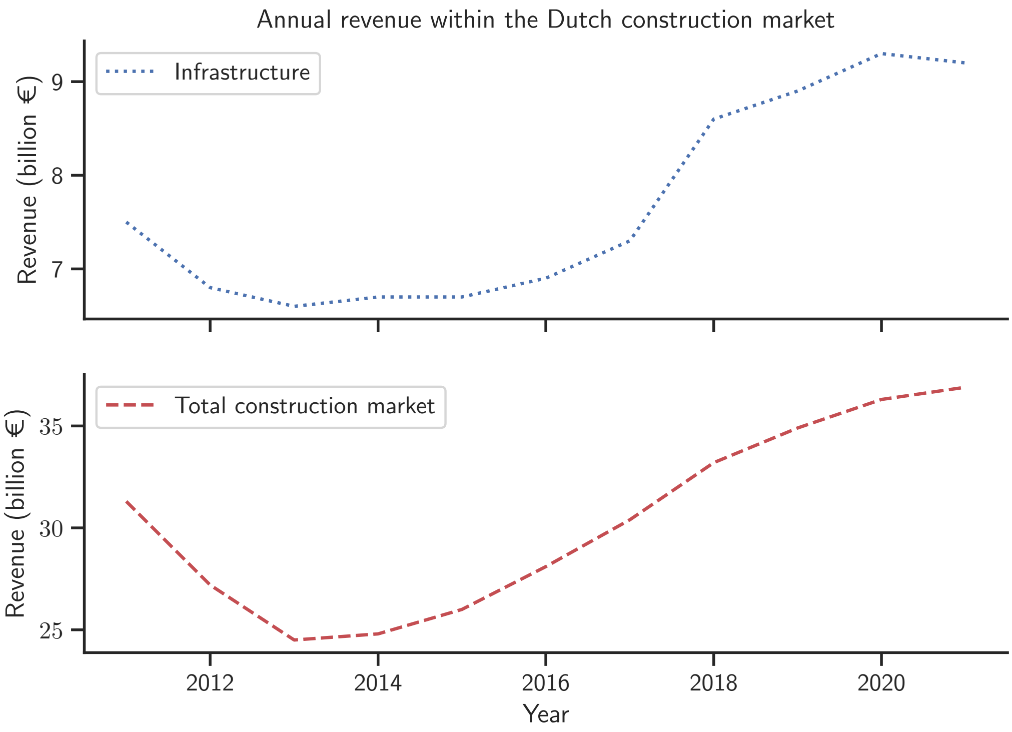

Figure 5.7: Visualization of the revenue in the Dutch construction sector over time

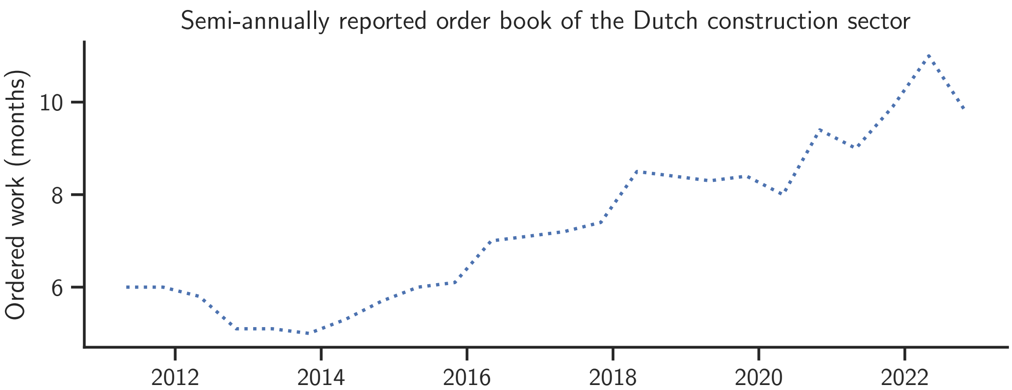

Figure 5.8: Reported order book by the Dutch construction sector over time

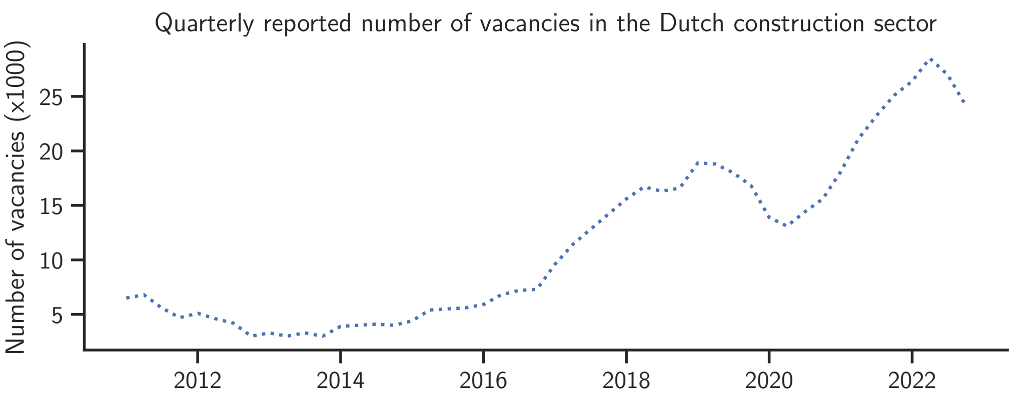

Figure 5.9: Number of vacancies over time in the Dutch construction sector

5.4. Conclusion

This chapter answers the question ”What data is available to utilize in the research?”. An extensive list of available data has been presented in this chapter. A connection is made with chapter 3 which explains what factors potentially impact the number of competitors. The ten found factors are summed up below. It is attempted to match each factor to the available data shown in this chapter.

- Project type
- Project value
- Project duration
- Project location
- Client
- Contract type
- Market conditions
- Need for work
- Tender method
- Competition

Some connections between the actual variable and the factor are obvious, like the project type, project value, and tender method. Other connections are more far-fetched because the connection between the variable and factor is indirect, like the contract type and need for work. For a few factors, multiple variables are used in the research. It must be noted here that some variables are unable to be grasped in the research. No value has been found to represent the award criteria. Also, it was not possible to add a value that represents legislation on for example nitrogen emission regulations.

6 Results of three important model set-ups

The results that are generated should help in answering the sub-question ”What results does the model show and which relations are identified between client decisions and the number of competitors?”. As mentioned in chapter 4 a variety of analyses are executed in order to investigate the relation between the number of competitors and client factors possibly affecting this number. For completeness, other environmental factors must be included in the analyses as well, but the focus in this chapter will be on factors that are influenced by clients. Three types of analyses are performed, of which two regression analyses on each model set-up and one entropy analysis on the entire data set. The two regression analyses include an initial regression that indicates which count data regression model performs best on the data and allows to harvest the regression coefficients, and a cross-validation that reports the accuracy of the chosen count data regression model.

A wide variety of model set-ups, i.e. combinations of input data, chosen variables, and regression models, were analyzed. Three model set-ups are considered important, namely (1) the model that uses all available variables, (2) the model that utilizes the highest number of projects, and (3) the model that generated the highest prediction accuracy. It is determined that these models are worth mentioning because the first allows us to compare all variables with each other, the second gains the best ’view’ on the entire market, and the third had the best performance on a niche of road projects in the data. From now on we will be referring to these model set-ups with respectively ”All”-model, ”Most”-model, and ”Best”-model.

The results for the initial regression and cross-validation will be laid out for each model in subsequently, section 6.1, 6.2, and 6.3. The found effects of each variable for each model are compared in section 6.4. Additionally, in section 6.6 the quality of the data in terms of predictability is investigated. It is shown how the values for the Entropy within the number of competitors distribution itself and the Mutual Information values between this value and the independent variables are determined. The calculations disclose the ”chaos” within the dependent variable and whether the available data is able to solve this ’chaos’.

6.1. Analysis with all variables

This section displays the analysis that uses all possible variables in the regression analyses. Due to the need to convert some client and environmental factors into dummies, the factors are represented in 53 variables. The list of all 53 variables plus the intercept are visible in table C.1, which compares the variable effects of the three different models covered in this chapter. As a result of the utilization of all variables, the analyses must drop a part of the data because of missing variable values. Due to the variable selection a total of 571 projects were available. Especially the application of the "contract type" variable, i.e. construct- or design & construct contract, heavily limits the number of available projects. This is because only projects up to 2017 have a value for the "contract type".

In this section, the results of the analyses on the "All"-model are presented. Subsection 6.1.1 reflects on the initial regression results, subsection 6.1.2 shows how a stable model accuracy is calculated with cross-validation, and subsection 6.1.3 shows what variable effects have been found in the initial regression.

6.1.1. Initial regression

The initial regression analysis consists of a general train-validate-test operation on the data with the two chosen count data regression models. The results of these operations are displayed in table 6.1. Indicated by the Pearson χ^2 value exceeding the allowed Pearson χ^2, it is shown the Poisson model has a poor fit. With respect to the NB model, the Pearson χ^2 test indicates a better fit. The difference between the deviance of the models affirms this. As a result, only the NB model is used in further calculations. To illustrate the instability of the model's predictions the accuracy values are displayed in the table as well. The NB testing and validation accuracy differ more than 7%. In order to generate a representative accuracy value, the entire data set is used in a cross-validation analysis in section 6.1.2.

Table 6.1: Initial regression results on projects that contain all variables

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	40.6	41.0	35.5	41.3	42.6	35.5
Best-guess accuracy (%)	33.5	27.9	30.6	33.5	27.9	30.6
Pearson χ^2	540.6	NaN	NaN	394.1	NaN	NaN
Allowed Pearson χ^2	457.2	NaN	NaN	457.2	NaN	NaN
Deviance	543.7	NaN	NaN	398.6	NaN	NaN

The test data set consists of 62 projects. The distribution of their actual number of competitors and predicted competition are shown in the histogram of figure 6.1. What stands out is that the distribution of the predictions is way more concentrated, the peak is higher and the distribution is more narrow than the distribution of the actual number of competitors This tells us that the NB model has a hard time making predictions far from the distribution center. The scatter plot next to the histogram also confirms this image, the majority of the dots lie off the "ideal prediction" line and especially actual competition values below 5 are predicted wrong. As we know from table 6.1 only 35.5% of the predictions lie within the range of one count from the actual number of competitors.

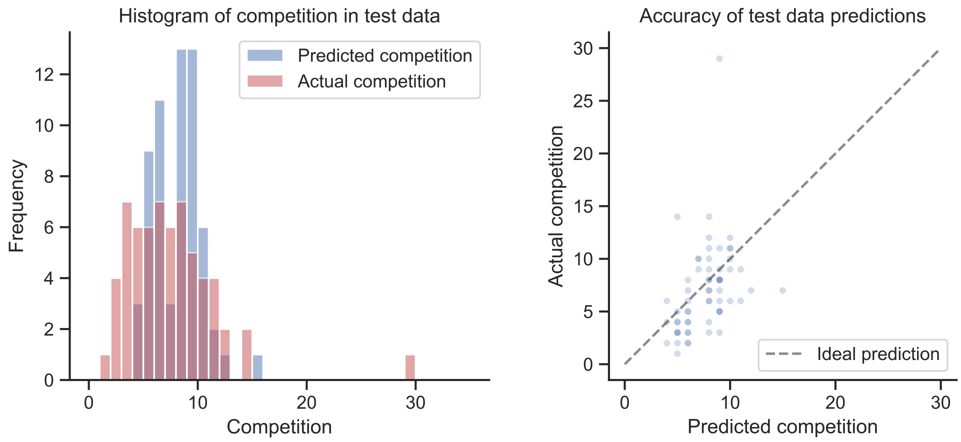

Figure 6.1: Visualization of Negative Binomial predictions on the test data that contains all variables

6.1.2. Cross validation

To establish a representative model accuracy cross-validation is applied to the entire data set of 571 projects. The results of the cross-validation are presented in figure 6.2. The accuracy on the ten folds shows big differences, the constant guess even outperforms the NB model with folds 1, 2, and 5. Still, the average NB model accuracy of 39.4% outperforms the average guess accuracy of 32.5% by 6.9%. To put it in perspective of the around 55 test projects about 22 projects receive an accurate prediction.

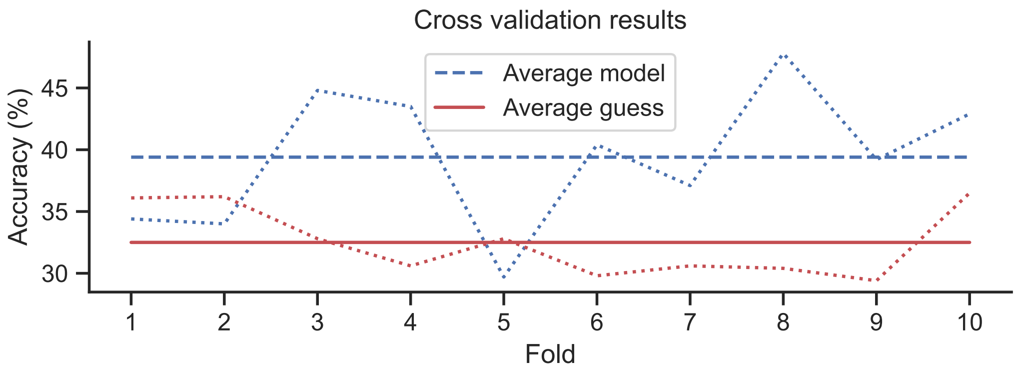

Figure 6.2: Negative Binomial Cross-validation results on projects that contain all variables

6.1.3. Found variable effects in initial regression

The found variable effects with the "All"-model range from 1.74 to -1.10. A positive value results in an increasing effect on the number of competitors while a negative value generates a decreasing effect. The list of all distilled variable effects is presented in table C.1.

The 5 variables that in an absolute sense show the largest effect on the number of competitors are: Revenue of the infrastructure market (1.74), revenue of the construction market (0.55), tunnel construction work (0.48), market order book (-0.91), and civilian construction work (-1.10). Interestingly, 3 of the 5 variables are out of the client's influence.

The 5 client decision variables with the largest effect are; Tunnel construction work (0.48), Restricted tender procedure (0.37), Hydraulic construction work (0.27), Rail construction work (-0.33), and Civilian construction work (-1.10). The other client decision factors and their effects are displayed and compared in section 6.4. It stands out that 4 of the 5 most influential variables consider a type of work.

6.2. Analysis with most projects

When the variables ”project duration” and ”contract type” are left out of the analyses a total of 2179 projects are available for the research. Because only two variables had to be left out the ”Most”-model has the best number of variables/number of projects ratio. As a result, this regression analyzes the construction market most completely. The main difference with the ”All”-model is that projects from 2017 up to 2022 are included in the set.

In this section, the results of the analyses on the ”Most”-model are presented. Subsection 6.2.1 reflects on the initial regression results, subsection 6.2.2 shows how a stable model accuracy is calculated with cross-validation, and subsection 6.2.3 shows what variable effects have been found in the initial regression.

6.2.1. Initial regression

The results of the general train-validate-test operations are presented in table 6.2. It can be seen from the Pearson χ^2 values that the Poisson model has a poor fit while the NB model scores below the threshold. The deviance of the NB model also indicates the model performs better than the Poisson model. Resulting from this observation it is chosen to continue with only the NB model in following calculations.

Table 6.2: Initial regression results on data containing the most projects

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	33.9	35.9	35.3	33.8	35.0	34.5
Best-guess accuracy (%)	31.2	29.6	29.3	31.2	29.6	29.3
Pearson χ^2	3034.3	NaN	NaN	1637.0	NaN	NaN
Allowed Pearson χ^2	1778.5	NaN	NaN	1778.5	NaN	NaN
Deviance	2927.5	NaN	NaN	1586.6	NaN	NaN

Table 6.2 displays an NB model accuracy of 34.5% on the test data, this data set contains 232 projects. The predictions on these projects are displayed in figure 6.3. The histogram shows that the distribution of the predictions is more narrow and has a higher peak than the distribution of the actual number of competitors. This has already been seen in the results of section 6.1. It is an indication that the model has a hard time predicting counts far from the center of the distribution. The scatter plot also clearly shows many predictions miss the actual number of competitors. A better accuracy would have led to the predictions centering around the ”ideal prediction” line.

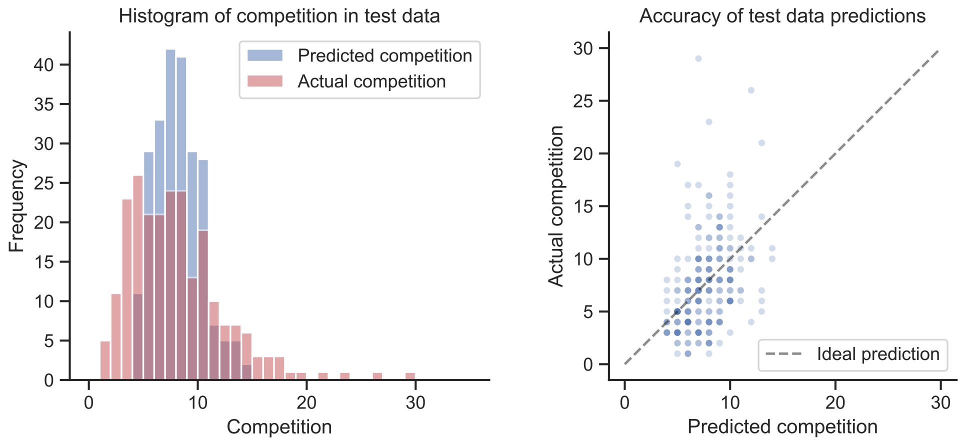

Figure 6.3: Visualization of Negative Binomial predictions on the test data containing the most variables

6.2.2. Cross validation results

The cross-validation on the "Most"-model data resulted in the graph presented in figure 6.4. It is clearly visible the model performs worse than the guess with some folds. But on average the model performs better than the average guess. It is found the model reached an average accuracy of 34.1% while the average guess accuracy performs 3.3% worse with 30.8%. In other words, the model is able to predict the number of competitors well on around 75 of the 220 analyzed projects.

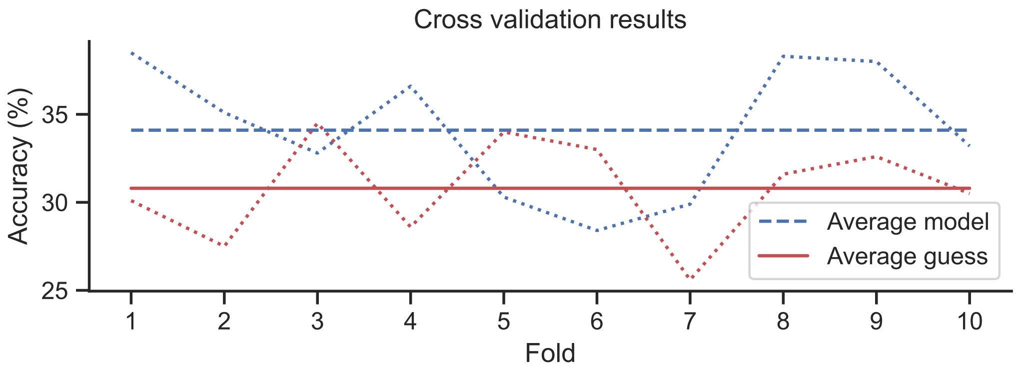

Figure 6.4: Negative Binomial cross-validation results on data containing the most projects

6.2.3. Found variable effects in initial regression

The found variable effects with the "Most"-model range from 1.31 to -0.46. A positive value results in an increasing effect on the number of competitors while a negative value generates a decreasing effect. The list of all distilled variable effects is presented in table C.2.

The 5 variables with in an absolute sense the largest effect on the number of competitors are: Revenue of the infrastructure market (1.31), 2015 year of tender (0.36), Project size very small (0.32), 2022 year of tender (-0.34), and Market order book (-0.46). It stands out that just 1 of the 5 variables is client related.

The 5 client decision variables with the largest effect are; Project size very small (0.32), Restricted tender procedure (0.31), Civilian construction work (0.25), Utility construction work (-0.24), and Project size king-size (-0.29). The other client decision factors and their effects are displayed and compared in section 6.4. Within this analysis, it appears that especially the project size and the type of work are of influence.

6.3. Best performing analysis

The best-performing model set-up was created with two measures; only road projects were selected, plus the "project duration" and "contract type" variables were dropped. As a result, a data set of 345 projects was available for analysis. An implication of the decisions made in the data selection is that the results are not representative of the construction market as a whole. But the results could give an insight into how the factors affect the number of competitors in road projects.

In this section, the results of the analyses on the "Best"-model are presented. Subsection 6.3.1 reflects on the initial regression results, subsection 6.3.2 shows how a stable model accuracy is calculated with cross-validation, and subsection 6.3.3 shows what variable effects have been found in the initial regression.

6.3.1. Initial regression

Table 6.5 displays the initial regression results on the "Best"-model. It can be seen from the goodness-of-fit statistics both the Poisson and NB model perform within χ^2 boundaries. Looking at the model accuracy's it can also be seen the two utilized models perform identically. The difference lies in the deviance, which gives the NB its advantage. Therefore it is chosen to continue the further calculations with the NB model.

Figure 6.5: Initial regression results only taking road projects into account

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	44.2	28.6	30.8	44.6	28.6	30.8
Best-guess accuracy (%)	37.4	32.1	28.2	37.4	32.1	28.2
Pearson χ^2	276.9	NaN	NaN	273.3	NaN	NaN
Allowed Pearson χ^2	284.7	NaN	NaN	284.7	NaN	NaN
Deviance	269.7	NaN	NaN	266.0	NaN	NaN

The test data consists of 39 projects, and the NB model reached a prediction accuracy of 30.8% on this data. The predictions are displayed in figure 6.6. Looking at the histogram, at first sight, it seems the distribution of the prediction matches the distribution of the actual competition quite well. But the peak of the predicted competition reveals the distribution is narrower than the distribution of the actual counts. The scatter plot shows especially projects with a high actual number of competitors are predicted wrong. While the predictions below the value of ten seem to follow the "ideal prediction" line.

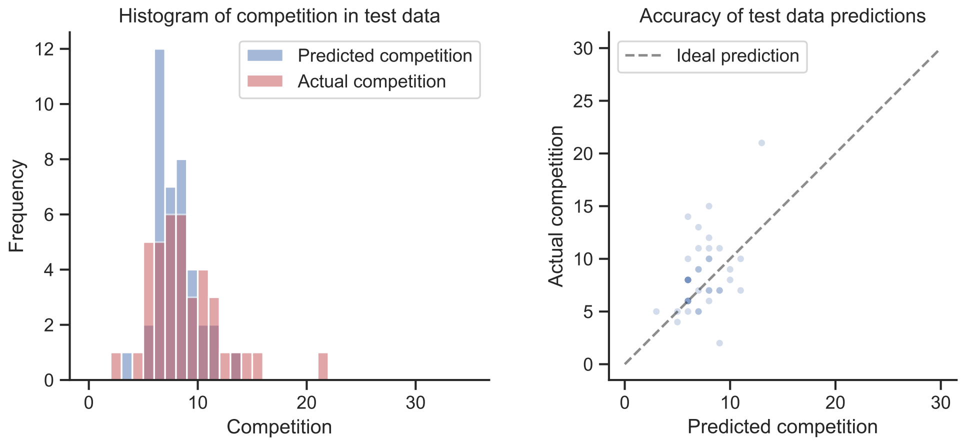

Figure 6.6: Visualization of Negative Binomial prediction results only taking road projects into account

6.3.2. Cross validation results

With the cross-validation on the "Best"-model, it was found that the NB model generated an accuracy of 44.9% on average. While the constant guess resulted in an average accuracy of 36.2%. This makes that the "Best"-model on average has an 8.7% increased accuracy compared to the guess. The average value and the individual fold results are shown in figure 6.7. To put it in perspective, around 17 of the 40 projects are predicted correctly.

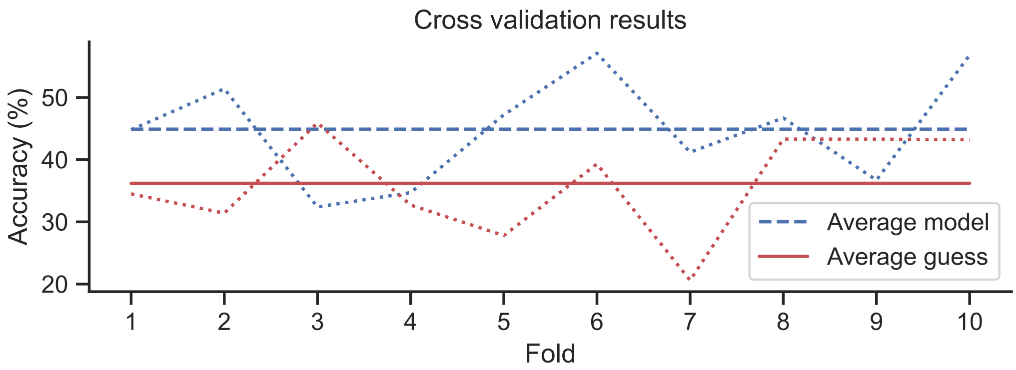

Figure 6.7: Negative Binomial Cross-validation results only taking road projects into account

6.3.3. Found variable effects in initial regression

The found variable effect ranged from 1.00 to -0.45. As already stated in the sections above a positive value has an increasing effect on the number of competitors while a negative value has a decreasing effect. A list with all calculated variable effects is shown in table C.3.

The 5 variables that have the largest impact on the number of competitors in an absolute sense are: Revenue infrastructure market (1.00), Revenue of construction market (0.88), Market order book (0.36), 2013 year of tender (-0.36), and Open vacancies in the sector (-0.45). It is very clear no client-related variable belongs to the top five.

The 5 client decision factors that mostly affect the number of competitors are; Restricted tender procedure (0.31), Nutscode NL1 (0.24), National tender scope (0.10), Public tender procedure (-0.13), and Nutscode NL (-0.18). Interestingly within the road projects the location and tender procedure seem most important.

6.4. Compare coefficient results

To generate realistic models, as many variables possibly affecting the number of competitors as possible have been included in the research. But to answer the main research question of this thesis we will take a look at the variables connected with client decisions. An overview of all resulting variable effects is listed in the tables from appendix C.

In table 6.3 below the variable effects, connected with the client decisions, of the three covered model set-ups are presented. The variable effects are translated into the plus and minus signs as a visualization of their effect on the competition. An increasing effect is labeled with a plus and a decreasing effect is labeled with a minus. For example, the ”Nutscode NL” variable is labeled positive to the ”Most”-model, while having a negative effect within the ”All”- and ”Best”-model. Effect values that are in an absolute sense larger than 0.1 receive a double sign, when the absolute value exceeds 0.3 a triple sign is applied. These thresholds are chosen based on gut feelings and determined to supply insight into the variable effect.

To support the answer to the main research question we would like to look at analyses that considered the entire market. Respectively, the ”All”-model and ”Most”-model. The analyses included variables for; project location, project size, tender procedure, tender scope, and type of work. The ”All”-model also included; project duration, and contract. Within these variables, the client has certain options, as shown in table 6.3 for example project size shows six options. On each project only one of these options is applicable.

The variable effects indicate client decisions impact the competition in a variety of ways. The decision on project location for example shows that NL2, NL3, and NL4 locations increase competition. The effect of choosing an NL or NL1 location is not clear due to the differences in model results. With respect to the project size, the ”King-size” and ”Very-large” project sizes decrease the competition, while the other project sizes increase the competition. Another clear distinction is visible within the options on procedures. The restricted tender procedure is considered highly competition-increasing while the public procedure results in decreasing competition. The effect of utilizing the competitive dialogue procedure is not unambiguous. Whether a national or European scope is applied doesn’t really have a distinctive effect on the competition. Considering the type of work, variables also show clear effects from a few variables. For example, both rail- and installation work decrease the competition in the models. While general-, road-, bridge-, tunnel-, ground-, and hydraulic work increases the competition. The effects of civilian- and utility work on the competition are not unanimous among the models. The ”All”-model also indicates that applying a construct contract increases the competition while choosing a design and construct contract has no effect. It was found that the project duration was assigned a very marginal effect in this model (-0.01) and therefore does not really influence the competition.

Interestingly the ”Best”-model affirms the ”All”-model and ”Most”-model in many variables of the tender procedure, tender scope, and project location but differ in the project size effect. Contrary to the other models the ”Best”-model assigns a decreasing effect to ”Small” projects while an increasing effect is assigned to ”Very-large” and ”King-size” projects. The project size effect on road projects is therefore considered the opposite of that of the entire market.

Table 6.3: Translation of the variable effects on the competition from the three covered models

Variable:	**"All":**	**"Most":**	**"Best":**
Nutscode NL	-	+	- -
Nutscode NL1	0	++	++
Nutscode NL2	++	++	++
Nutscode NL3	+	+	-
Nutscode NL4	+	+	+
Project size very small	++	+++	+
Project size small	+	++	-
Project size medium	++	++	+
Project size large	++	++	+
Project size very large	- -	-	+
Project size king-size	- -	- -	+
Procedure restricted	+++	+++	+++
Procedure public	- -	- -	- -
Procedure competitive	- -	++	-
Tender scope national	+	++	++
Tender scope European	+	++	+
Contract construct	+		
Contract design and construct	0		
General work	++	++	
Installation work	- -	- -	
Road work	++	+	
Bridge work	+	++	
Tunnel work	+++	+	
Ground work	++	++	
Utility work	++	- -	
Civilian work	- - -	++	
Hydraulic work	++	++	
Rail work	- - -	- -	
Project duration	-		

6.5. Additional reflection on coefficient results

Based on the results of all variable effects, as listed in the table of appendix D, some interesting things can be identified. In line with the statements that the number of competitors on tenders in the Dutch construction industry is decreasing, the same trend is visible in the table. Considering the ”Most”-model and ”Best”-model from the year 2015 until 2022 a decreasing trend is visible.

Another interesting phenomenon in the results is that without an exemption the variables connected to market conditions were assigned the largest effect on the number of competitors.

6.6. Entropy & Mutual Information results

The entropy and mutual information help to identify the predictability of a dependent variable given the available independent variables. Both values are gathered in an objective way. The entropy is calculated based on the distribution of the number of competitors, the dependent variable. This value expresses the amount of ”chaos” present in the variable, the higher the chaos the harder it is to predict the variable. The mutual information is a value for the information an independent variable contains about the dependent variable. In other words how much ”chaos” potentially can be solved with the independent variable? The mutual information origins from the joint probability of the dependent variable and each independent variable distribution. When the sum of all mutual information values is taken an expression is found for how much ”chaos” potentially could be solved by the independent variables. Note that this assumes all variables are completely independent.

It is calculated, as shown in table E.1, that the entropy within the number of competitors variable has a value of 3.96. The combined mutual information is determined to be 1.24, as displayed in table E.2. This means the available data is able to ’solve’ around 30% of the ’chaos’ present within the dependent variable. Therefore, the results show that the combined mutual information does not meet the entropy threshold. This indicates it is difficult to make predictions on the dependent variable with the available independent variables.

6.7. Reflection on other model set-ups

There are two ways to impact the model set-ups, by choosing a certain selection of variables and by changing the input data. It is attempted to choose a set-up that generates a model capable of making good predictions. The first threshold was set by the ”All”-model, which has an accuracy of 39.4% which is 6.9% better compared to the guess accuracy. A wide variety of set-ups was analyzed, but no set-up except the ”Best”-model was able to increase the prediction accuracy. The enumeration below shows which set-up attempts, except from the three already covered models, were undertaken:

1. Only restricted tender procedure projects
2. Only public tender procedure projects
3. Without projects labeled Nutscode NL or General type of work
4. Only very small project size projects
5. Only small project size projects
6. Only nutscode NL projects
7. Only nutscode NL3 projects
8. Only project with a local client
9. Only project with a national tender scope
10. Only projects with a European tender scope
11. Only general type of work projects
12. Only ground work projects

The analysis results on the set-ups are added in appendix F. One might notice not every variable is used as a selector for the input data. This is due to the fact that using those variables resulted in a data set with too few projects to analyze. As mentioned in chapter 4 a 1/10 ratio must be kept between projects and variables. Therefore, in some of the enumerated analyses, a part of the variables had to be dropped prior analysis to comply with the demanded ratio.

6.8. Conclusion

This chapter is written to answer the sub-question ”What results does the model show and which relations are identified between client decisions and the number of competitors?”.The results show without exemption the Negative Binomial (NB) model produces the best predictions. With the NB model, it was found that the ”Best”-model set-up predicts the number of competitors well for 44.9% of the test projects, an accuracy of 8.7% better than the accuracy of a constant guess. The ”All”-model generates an accuracy of 39.4%, a 6.9% improvement compared to the constant guess, and the ”Most”-model displays an accuracy of 34.1%, just a 3.3% improvement. The performance shows that the models were only able to gain a marginal advantage compared to the performance of a standard guess.

Although the accuracy of the models is low, the coefficients found still account for certain relations. The resulting regression coefficients from the three models indicate that market-wide the very-large and king-size projects have a decreasing impact on the number of competitors compared to the other project sizes. Also, the restricted tender procedure has an increasing effect while the public tender procedure has a decreasing effect. Another insight is that installation- and rail-works both have a decreasing impact on the number of competitors compared to the other types of work. It must be noted here that the effects of individual variables sometimes alter dependent on the input data, for example with the road projects in the ”Best”-model the small project size has a decreasing effect while large- and king-size projects have an increasing effect on competition. It must be noted here, although the focus lies on client decision, that in each analysis the market conditions appeared to have the largest impact on the predicted number of competitors.

The calculations regarding the Entropy disclose the ’chaos’ within the dependent variable is expressed with an Entropy of 3.96. While the sum of the Mutual Information values of the independent variables shows a value of approximately 1.24. In other words, it is estimated the variables are able to solve a little over 30% of the ”chaos”. This indicates that it is hard to make predictions on the number of competitors with the available data.

7 Discussion

The results in chapter 6 show that, considering this research, it is hard to quantify the effect of client decisions on the number of competitors. This fact is a direct result of unsatisfactory prediction accuracies. The errors in the predictions could have a few causes. Either no relation exists between the number of competitors and the data, essential data is missing, the data is not of sufficient quality, the utilized model is not suitable, or a combination of those. The considerations on these topics are given in the sections below.

7.1. Number of competitors predictability

The predictions in this research are built on the assumption that a relation exists between the number of competitors and the available data. Especially considering the results this assumption is debatable. The lack of quantified relations could be the result of a fundamental lack of uniform and rational bidding.

Taking the highly complex environment of the construction industry into account, it is very likely that the construction industry as a whole does not bid in a uniform way. Not every contractor will have the same project preferences and the same risk-taking behavior. When there is no uniform behavior there is also no possibility to quantify uniform relations. Nevertheless, it would still be interesting to identify the absence of relations. But because still many improvements in the research can be made this is yet impossible.

7.2. Available variables

The results on the Entropy and Mutual Information show that the available data is not able to solve the ”chaos” within the distribution of the number of competitors. This is an indication that information is missing, assuming a relation can be found between the data and the number of competitors. In this research, it was not possible to include information on award criteria, world events like Covid and the war in Ukraine, and abrupt legislation changes on for example nitrogen emission. One could argue that the application of the time factor in this research (the year of tender variable) compensates for the missing information on world events and legislation changes, but the information on award criteria is still missing. It is unlikely that accounting for the missing data would result in solving the Entropy problem, but it would probably help in improving the model performances.

Additional information that is also desired and currently unavailable, is data on which parties competed in the tender. Now, only the winning contractor along with its offer is disclosed. The missing competitor information makes it impossible to investigate which part of the construction market competed. More importantly, missing competition information makes it impossible to gain insights into which parties did not bid. Overlooking the available data, the client side is rather transparent, while information from the contractors is often lacking. One could argue that since the client and winning contractor are transparent about their deal, information about the losing competitors should be disclosed as well.

7.3. Data quality

The data that is utilized in this research is of varying quality. Where some projects in the data set are described very precisely, others are described in undetailed and general terms. The lack of detail on some projects and especially the varying in data quality makes it harder to quantify the relation between the number of competitors and the available data. Variables considered important and sometimes of insufficient quality are the project location, contract type, and type of work.

Considering the project location many projects are just labeled as projects located in the Netherlands. While the nuts codes have the possibility to define project locations up to municipality level. Labeling a project as located in the Netherlands is useless since it is a property shared among the projects displayed on TenderNed.

The contract type only shows two contract types in the data: (1) construct contracts and (2) design and construct contracts. However, a much larger variety of contracts are applied in the Netherlands. For example, no DBFM contract is found in the data while it is known this contract type is applied multiple times in the Netherlands. Also, a value for the contract type is only registered in the data up to 2017. Given the fact that a contract type has a big impact on the contractor, it is a great loss for the analysis that just a few contract types are registered for such a limited time.

With regard to the type of work, the same problem is registered as with the project location. Many projects are labeled as general construction work, leaving the actual work up to the imagination. This is especially a shame because it was found that, when registered correctly, the type of work could have a large impact on prediction accuracies. Note that the model set-up with the best performance was based on only road projects. Also, the description of the work does not clearly indicate how the work should be executed. For example, a project labeled as tunnel work does not disclose what type of tunnel must be constructed. When prescribed that the tunnel must be submerged or bored completely changes the project and which contractors are able to complete the work. These inaccuracies in the description of types of work are an crucial bias in the utilized data set.

7.4. Model choice

The results presented in chapter 6 also give food for thought on the model choice. It appears that the utilized count data regression models are not fit for the job. This is indicated by the prediction capabilities of the model and especially the different results of the three model set-ups. It is shown that different variables had an opposite effect on the road project from the ”Best”-model compared to the ”All”-model and ”Most”-model. While the results for road projects identified that very-large and king-size road projects increase the number of competitors while small road projects decrease the competition. The results of the other models identify that projects of bigger size have a decreasing effect on the number of competitors and smaller projects show an increasing effect.

This sparks the understanding that clusters in the data could have conflicting relations that cannot be caught in a single value. In other words, considering construction sector-wide data, one cannot quantify the effect of a variable in a single value when it increases the number of competitors for one project and reduces the number of competitors for another. Something that has been tried with the models in this research. Therefore, the models of this research are unsuitable to quantify sector-wide relations.

8 Conclusion

This conclusion has the goal to answer the main research question, ”To what extent do client decisions in the procurement of construction projects in the Dutch construction sector affect the number of competitors?”. Answering this question should contribute to solving the problem that the number of competitors decreases in the tenders of projects in the Dutch construction sector. To guide the research sub-questions were formulated, and the answers to these sub-questions are shared subsequently below.

- What is known about competition within the tender phase of construction projects, and factors affecting this competition?

It has been determined that a variety of factors potentially affect the number of competitors. The factors were divided into two groups, factors of influence of the client, the client factors, and factors out of influence of the client, the environmental factors. The identified client factors are project location, type of work, project duration, project size, tender procedure, award criteria, and contract type. The found environmental factors are economic situation, market order book, availability of building materials, legislation, and contractor relationships.

- What drives the contractors’ bid decision within tenders?

To verify whether the previously determined factors of interest are also considered by contractors a literature review has been conducted. It has been determined from the literature that ten factors could summarize which factors are considered by contractors. The factors are project location, project type, project duration, project value, tender method, client, contract type, market conditions, competition, and need for work. These factors match the previously determined factors well, and pretty much describe all a contractor could know during the procurement of projects.

- Which method could be used to investigate what connection exists between client decisions and tender competition?

It has been identified that two different count regression models are suitable to investigate how client decisions impact the number of competitors. Both the Poisson model and Negative Binomial model are considered up to the task. In order to compensate for the chance of accidentally reporting extraordinary results, cross-validation is applied to these models. Next to the regression analyses the quality of input data is also investigated. By calculating the Entropy and Mutual Information within the available data it is determined if the independent variables in the data are able to solve the ’chaos’ within the dependent variable. Thus, disclosing how well the number of competitors is predictable.

- What data is available to utilize in the research?

It was found that a little over 2600 projects could be included in the research. And that for every project, for almost each identified factor that influences the number of competitors, a representative value has been found. The utilized data represents the factors of project type, project value, project duration, project location, client, contract type, market conditions, need for work, and tender method. It was not possible to find values for the award criteria, sudden legislation on for example nitrogen emission, and world events like the war in Ukraine.

- What results does the model show and which relations are identified between client decisions and the number of competitors?

It has been determined that three model set-ups, i.e. combinations of input data, chosen variables, and the regression model, are important to mention. Namely, (1) the model that uses all available variables, (2) the model that utilizes the highest number of projects, and (3) the model that generated the highest prediction accuracy.

The two analyses that utilized construction sector-wide data, the analysis with all variables (”All”-model) and the analysis with the most projects (”Most”-model), report a prediction accuracy of respectively 39.4% and 34.1%. This means that, although the predictions outperform a constant guess by 6.9% and 3.3%, the majority of the predictions are still wrong. The reported performances indicate some relations are identified between the number of competitors and the input data. But the results also show the analyses set-ups are not able to fully grasp how the competition is driven. It can be concluded from this that with the used data and models it is not possible to quantify the effect of client decisions on the number of competitors.

Despite the inability to quantify the effects, the results still show a certain relation between the number of competitors and utilized variables. These results should be cautiously treated as indications of how the variables impact the number of competitors. In the two mentioned sector-wide analyses it was found that decisions on project size, tender procedure, contract type, and type of work, resulted in either a positive or negative effect on the competition. Considering the project size it is indicated that very-large or king-size projects decrease while other smaller project sizes increase the number of competitors. With regard to the tender procedure, contrary to the public procedure the restricted procedure was indicated to impact the competition in a positive way. The results on the contract type indicate a construction contract has an increasing effect while a design and construct contract has no impact on the number of competitors. Type of works that showed a decreasing effect on the competition are installation- and rail work. An increasing effect is found in general-, road-, bridge-, ground-, and hydraulic work.

Considering the third model with the highest prediction accuracy (”Best”-model), some interesting results were found. It is identified that contrary to the ”All”- and ”Most”-model, large projects increase the number of competitors while small projects decrease the number of competitors. This shows that project clusters within the data value project characteristics differently.

Also noteworthy, it has been identified with all three model set-ups that the environmental factors, i.e. market conditions, had a larger impact on the number of competitors than the client factors.

Next to the regression analyses the predictability of the number of competitors has been investigated. The calculations on Entropy and Mutual information disclosed that the independent variables within the data can approximately solve around 30% of the ’chaos’ in the distribution of the number of competitors. In other words, it is considered very hard to make predictions on the number of competitors with the available data. Something that is affirmed by the generated prediction results.

- **To what extent do client decisions in the procurement of construction projects in the Dutch construction sector affect the number of competitors?**

Summarizing the results on the sub-question a conclusion can be formulated on the main research question. It has been found that the predictive capabilities of the created model set-ups are unsatisfactory. Therefore, it is concluded that the effects of client decisions in the procurement of construction projects in the Dutch construction industry could not be quantified.

Although there is room for improvements in the methodology and utilized data, the results raise the question of whether even a relation exists between the identified factors and the number of competitors. Interestingly, the regression analyses accounted for almost all variables that were considered important, yet the prediction accuracies remained below par. Also, the marginal relations that were identified pointed at the market conditions as most important to the number of competitors. As a result, it seems that clients in the Dutch construction industry have a marginal effect on the industry and that they are just as dependent on the market conditions as the contractors are. This raises the question of whether clients are able to solve the decreasing number of competitors in the Dutch construction sector.

Prior to the conclusion can be made that no relation exists between project factors and the number of competitors, improvements must be made to the research.

9 Recommendations

Although this research failed to accurately predict the number of competitors a lot of knowledge has been gained on the topic. What has been learned is especially useful for future research, clients in the construction sector like Rijkswaterstaat, and the provider of the procurement platform TenderNed. This knowledge is translated into a set of recommendations. With respect to future research, the focus of the recommendations will lie on the data, the data quality, the to-be-utilized model, and supporting qualitative research. Considering the clients the recommendations will reflect on how the gained knowledge potentially influences the future procurement process. Finally, recommendations are given that should help TenderNed in improving the quality of its platform, something both beneficial to its users and for researchers in the entire field of procurement.

9.1. For further research

Before firm conclusions can be drawn on the existence and absence of certain relations within the data more research must be conducted. Especially because it is harder to prove the lack of a relationship than the existence of one. Therefore the available data, data quality, utilized model, and supporting qualitative research must be of high quality.

Available data

In order to improve future research additional information on projects and more projects should be collected. Additional information on the projects could be collected from TenderNed. Accompanied with each project notification a variety of documents is supplied on TenderNed. These documents generally supply more detailed information on the project plans. This gives the opportunity to gather additional information on award criteria and prescribed execution methodology, often disclosed in these documents. It is recommended to utilize pdf-readers to scrape this information and enrich the project data. Considering the number of projects, in this research some hard-to-clean projects were left out. For example, projects that were awarded in multiple plots did not make the final set. With these projects, it was unclear whether the number of competitors were competing on specific plots or on all the plots at once. Also, it was not clear if contractors were able to win all plots or just one. It is expected that the documents attached to each project solve these uncertainties. As a result, more projects could be added to the research, the more projects are included in the data the more completer the research becomes.

Data quality

It is also recommended to improve the data quality. This could be achieved by analyzing the already mentioned documents on TenderNed which are attached to each project. The documents generally supply more detailed information on the type of work, contract type, and project location. The information in the notifications regarding these variables is often less detailed. Where for example the project is labeled as general work, project describing documents could disclose the actual type of work. Because of the large number of construction projects on TenderNed, this job is considered very time-consuming. It is advised to use a pdf-reader to automate the scraping process.

Model choice

With respect to the model choice, it is recommended to choose another model and step away from the application of a regression. The Poisson and Negative Binomial models are the first count data regression models to go to. But the results indicate they are not appropriate for the data. A model should be chosen that is able to generate parameters based on project clusters within the data. That means, the model should be capable of in theory showing that large projects have a positive impact on the number of competitors for road projects but for example a negative impact on utility projects. The first type of model that comes to mind is a decision-tree model, that alters coefficients based on event characteristics. It is imaginable that for example, such a model is able to treat road projects in a different way than utility projects.

Qualitative research

It is recommended to support improved quantitative research with qualitative research. The qualitative research should focus on the questions: Whether contractors in the Dutch industry think project factors have an influence on the number of competitors? How much influence do they believe each project factor has on the number of competitors? Which factors are considered most important? Attempting to answer the same questions as in the quantitative research, using exactly the same factors in the same construction industry, will make the findings can be combined. It would be valuable to compare the found relations in future quantitative research with the relations identified in the qualitative research. When the results of both studies are merged this will make the future findings more robust. In the ”Multi-criteria approach” part of the literature review of section 3.2 a variety of methods discussed that rank factors of interest. These or similar methods that ask contractors to rank factors of interest could be utilized in qualitative research.

9.2. For clients in the Dutch construction sector

It seems from the results that if project factors even influence the number of competitors the environmental factors are dominant. This indicates that the influence of clients is marginal and they are unable to impact the number of competitors. Although more research has to be conducted, you could say that it appears that clients in the Dutch construction industry are just as dependent on the market conditions as contractors are. Therefore, it is recommended that clients acknowledge that their impact on the outcomes of tender processes is marginal. As a result, clients should reevaluate if attempts to solve the decreasing competition in the market, like the introduction of the two-phase contract by Rijkswaterstaat (Rijkswaterstaat, 2019), are useful and necessary. Maybe the key to increasing the number of competitors for an individual client is by gaining a reputation as a good and reliable client.

Besides the acknowledgment of their position and influence in the construction market, it is recommended that clients fulfill their role in supplying correct information on the procurement processes. In using TenderNed the clients are responsible for the displayed data, but many projects lack important information. For example, many projects do not show a contract value or fail to report which contract type is utilized. Given that public clients are obliged to disclose procurement data and that they benefit from research on the topic, clients should take the task to supply correct data seriously.

9.3. For TenderNed

TenderNed has the task of supplying clients and contractors with a procurement platform. Although they are not responsible for the quality of the data they receive from clients, they could positively impact the data quality with little adjustments to their process. This is both benefits the service they provide and the usability of the data for research. It is recommended that improvements are made in threefold.

Firstly, TenderNed should make it impossible for clients to publish awards with missing information. Given that it has legal consequences when clients fail to publish on TenderNed in time it is likely that data quality will improve with this measure. Additionally, when unrealistic information is supplied, like a contract value of 1 Euro or an unspecified nuts code, the announcement should be refused as well.

Secondly, it is recommended that TenderNed implements new data fields on currently missing information. A client should be able to specify which type of contract is used and what underlying uniform administrative conditions apply to this contract. Also, it should be able to report which award criteria are used and when applicable what discounts could be earned when specific criteria are met.

Thirdly, it is recommended that TenderNed uses a more simpler representation of the type of work next to the now common CPV code. The CPV code varies heavily in detail, from general work to the installation of rail power supply lines. Instead of using hundreds of CPV codes to describe the work, TenderNed should give the possibility to classify the work in just a few correct and sufficient labels. The amount of CPV codes makes it almost impossible for clients to select the proper code, they often refuse to search their type of work in the list and just fill in that the project concerns general construction work.

10 Reflection

10.1. Research process

Looking back at the research I would do some things differently. With regard to the research process, one of the major ”mistakes” that I made is that I looked too late into the usability of the data. This was partly because I, before allowing myself to start on the fun data science part, forced myself to focus on the literature review and thesis structure. And partly, because I figured I had to deal with the data despite its quality. As a result, valuable time was lost that could have been used to enhance the data quality. It was after the first regression results were collected that I looked into the predictability of the data (with Entropy and Mutual Information). The Entropy and Mutual Information analyses affirmed that it would be hard to make predictions with the available data. This analysis could have easily been executed at the beginning of the research and would have allowed to intervene early in the graduation process. Therefore, in future research, I would make sure to start with the data early on.

10.2. Personal process

With respect to my personal process, I would also reconsider the sequence in which I executed the research. The fact that I forced myself to first complete the literature review and thesis structure prior to data research also had another implication. Due to the lack of understanding of the data it was difficult to determine the writing perspective of the report. As a result, little has been written at the beginning of the graduation process. It was after the first results were collected that I found clarity on how I wanted the report to be written. I believe that the difficulties in writing are the main cause of the delay I experienced during my graduation. When I would have started with data research early on in the graduation I would probably have prevented parts of the delay. Although, I must also admit that, considering my education track record, a little delay belongs to my process. Let’s call it my fight against the rising retirement age.

References

Aanbestedingswet. (2012). https://wetten.overheid.nl/BWBR0032203/2022-03-02

Ahmad, I. (1990). Decision□support system for modeling bid/no□bid decision problem. *Journal of Construction Engineering and Management*, *116*, 595–608. https://doi.org/10.1061/(ASCE)0733-9364(1990)116:4(595)

Al-Humaidi, H. (2016). Construction projects bid or not bid approach using the fuzzy technique for order preference by similarity ftopsis method. *Journal of Construction Engineering and Management*, *142*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001180

Araújo, M., Alencar, L., & Mota, C. (2022). Classification model for bid/no□bid decision in construction projects. *International Transactions in Operational Research*, *29*, 1025–1047. https://doi.org/10.1111/itor.13037

Ballesteros-Pérez, P., del Campo-Hitschfeld, M., Mora-Melià, D., & Domínguez, D. (2015). Modeling bidding competitiveness and position performance in multi-attribute construction auctions. *Operations Research Perspectives*, *2*, 24–35. https://doi.org/10.1016/j.orp.2015.02.001

Benjamin. (2022). Rijkswaterstaat vreesde bankroet van grootste nederlandse bouwer bij project afsluitdijk. *NRC*. https://www.nrc.nl/nieuws/2022/06/14/rijkswaterstaat-vreesde-bankroet-van-grootste-nederlandse-bouwer-bij-project-afsluitdijk-a4133389

Benjamin & Meador, R. (1979). Comparison of friedman and gates competitive bidding models. *Journal of the Construction Division*, *105*, 25–40. https://doi.org/10.1061/JCCEAZ.0000821

Cameron, A., & Trivedi, P. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

CBS. (2023). *Cbs open data statline*. Retrieved January 1, 2023, from https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS

Chao-Duivis, M., Bruggeman, E., Koning, A., & Ubink, A. (2018). *A practical guide to dutch building contracts* (4th ed.). Instituut voor bouwrecht.

Cheng, M., Hsiang, C., Tsai, H., & Do, H. (2011). Bidding decision making for construction company using a multi-criteria prospect model. *Journal of Civil Engineering and Management*, *17*, 424–436. https://doi.org/10.3846/13923730.2011.598337

Christodoulou, S. (2004). Optimum bid markup calculation using neurofuzzy systems and multidimensional risk analysis algorithm. *Journal of Computing in Civil Engineering*, *18*, 322–330. https://doi.org/10.1061/(ASCE)0887-3801(2004)18:4(322)

Chua, D., & Li, D. (2000). Key factors in bid reasoning model. *Journal of Construction Engineering and Management*, *126*, 349–357. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:5(349)

Commision, E. (2020). *Statistical regions in the european union and partner countries*.

Cover, T., & Thomas, J. (1991). *Elements of information theory*. Wiley-Interscience.

Crowley, L. (2000). Friedman and gates—another look. *Journal of Construction Engineering and Management*, *126*, 306–312. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:4(306)

Dawood, N. (1996). A strategy of knowledge elicitation for developing an integrated bidding/production management expert system for the precast industry. *Advances in Engineering Software*, *25*, 225–234. https://doi.org/10.1016/0965-9978(95)00091-7

de Lange, M., & Visser, N. (2021). *Bedrijfseconomische kencijfers*. EIB.

Dias, W., & Weerasinghe, R. (1996). Artificial neural networks for construction bid decisions. *Civil Engineering Systems*, *13*, 239–253. https://doi.org/10.1080/02630259608970200

Dudkin, G., & Välilä, T. (2006). Transaction costs in public-private partnerships: A first look at the evidence. *Competition and Regulation in Network Industries*, *1*, 307–330. https://doi.org/10.1177/178359170600100209

Egemen, M., & Mohamed, A. (2008). Scbmd: A knowledge-based system software for strategically correct bid/no bid and mark-up size decisions. *Automation in Construction*, *17*, 864–872. https://doi.org/10.1016/j.autcon.2008.02.013

EIB. (2022). *Publicaties*. https://www.eib.nl/publicaties/

El-Mashaleh, M. (2010). Decision to bid or not to bid: A data envelopment analysis approach. *Canadian Journal of Civil Engineering*, *37*, 37–44. https://doi.org/10.1139/L09-119

El-Mashaleh, M. (2013). Empirical framework for making the bid/no-bid decision. *Journal of Management in Engineering*, *29*, 200–205. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000147

Feng, C., Li, L., & Sadeghpour, A. (2020). A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology*, *20*. https://doi.org/10.1186/s12874-020-01055-2

Friedman, L. (1956). A competitive-bidding strategy. *Operations Research*, *4*, 104–112. https://doi.org/10.1287/opre.4.1.104

Gates, M. (1967). Bidding strategies and probabilities. *Journal of the Construction Division*, *93*, 75–110. https://doi.org/10.1061/JCCEAZ.0000192

Govindarajan, M., & Chandrasekaran, R. (2010). Evaluation of k-nearest neighbor classifier performance for direct marketing. *Expert Systems with Applications*, *37*, 253–258. https://doi.org/10.1016/j.eswa.2009.04.055

Gugler, K., Weichselbaumer, M., & Zulehner, C. (2015). Competition in the economic crisis: Analysis of procurement auctions. *European Economic Review*, *73*, 35–57. https://doi.org/10.1016/j.euroecorev.2014.10.007

Heurkens, J., Keulen, J., de Koning-van Rutte, M., Stuijts, M., & Hebly, J. (2022). *Gids proportionaliteit 2012 3e herziening*. Ministry of Economic affairs. Retrieved January 1, 2023, from https://www.pianoo.nl/sites/default/files/media/documents/2021-10/gids_proportionaliteit_3e_herziening-januari2022.pdf

investing.com. (2023). *Netherlands 10-year bond yield*. Retrieved January 1, 2023, from https://www.investing.com/rates-bonds/netherlands-10-year-bond-yield-historical-data

Kalan, D., & Ozbek, M. (2020). Development of a construction project bidding decision-making tool. *Practice Periodical on Structural Design and Construction*, *25*. https://doi.org/10.1061/(asce)sc.1943-5576.0000457

Klumpenaar, S. (2022). Miljardenklus? de grote bouwers bedanken ervoor. *NRC*. https://www.nrc.nl/nieuws/2022/02/21/miljardenklus-de-grote-bouwers-bedanken-ervoor-a4092493

Kok, J., & Visser, N. (2019a). *Monitor bouwketen*. EIB.

Kok, J., & Visser, N. (2019b). *Monitor bouwketen*. EIB.

Lee, C., Famoye, F., & Akinsete, A. (2021). Generalized count data regression models and their applications to health care data. *Annals of Data Science*, *8*, 367–386. https://doi.org/10.1007/s40745-019-00221-8

Lemonte, A., Moreno-Arenas, G., & Castellares, F. (2020). Zero-inflated bell regression models for count data. *Journal of Applied Statistics*, *47*, 265–286. https://doi.org/10.1080/02664763.2019.1636940

Leśniak, A., Kubek, D., Plebankiewicz, E., Zima, K., & Belniak, S. (2018). Fuzzy ahp application for supporting contractors' bidding decision. *Symmetry*, *10*, 642. https://doi.org/10.3390/sym10110642

Leśniak, A., & Radziejowska, A. (2017). Supporting bidding decision using multi-criteria analysis methods. *Procedia Engineering*, *208*, 76–81. https://doi.org/10.1016/j.proeng.2017.11.023

Lin, C., & Chen, Y. (2004). Bid/no-bid decision-making – a fuzzy linguistic approach. *International Journal of Project Management*, *22*, 585–593. https://doi.org/10.1016/j.ijproman.2004.01.005

Lowe, D., & Parvar, J. (2004). A logistic regression approach to modelling the contractor's decision to bid. *Construction Management and Economics*, *22*, 643–653. https://doi.org/10.1080/01446190310001649056

Marzouk, M., & Mohamed, E. (2018). Modeling bid/no bid decisions using fuzzy fault tree. *Construction Innovation*, *18*, 90–108. https://doi.org/10.1108/CI-11-2016-0060

Mehrabani, M., Golafshani, E., & Ravanshadnia, M. (2020). Scoring of tenders in construction projects using group method of data handling. *KSCE Journal of Civil Engineering*, *24*, 1996–2008. https://doi.org/10.1007/s12205-020-1537-5

Paranka, S. (1971). Competitive bidding strategy. *Business Horizons*, *14*, 39–43. https://doi.org/10.1016/0007-6813(71)90115-7

Pavlou, M., Ambler, G., Seaman, S., Guttmann, O., Elliott, P., King, M., & Omar, R. (2015). How to develop a more accurate risk prediction model when there are few events. *BMJ (Online)*, *351*. https://doi.org/10.1136/bmj.h3868

PIANNOo. (2022). *Gunningscriterium*. Retrieved September 1, 2022, from https://www.pianoo.nl/nl/inkoopproces/fase-1-voorbereiden/keuze-gunningscriterium-en-opstellen-subgunningscriteria#gunningscriterium

PIANOo. (2022a). *Aanbestedingsprocedures*. Retrieved September 1, 2022, from https://www.pianoo.nl/nl/inkoopproces/aanbestedingsprocedures

PIANOo. (2022b). *Bouworganisatievormen*. Retrieved September 1, 2022, from https://www.pianoo.nl/nl/sectoren/gww/inkopen-gww/bouworganisatievormen-gww

PIANOo. (2022c). *Europese standaardprocedures*. Retrieved September 1, 2022, from https://www.pianoo.nl/nl/inkoopproces/aanbestedingsprocedures/europese-standaardprocedures

PIANOo. (2022d). *Nationale procedures*. Retrieved September 1, 2022, from https://www.pianoo.nl/nl/inkoopproces/aanbestedingsprocedures/nationale-procedures

Polat, G., & Bingol, B. (2017). Data envelopment analysis (dea) approach for making the bid/no bid decision: A case study in a turkish construction contracting company. *Scientia Iranica*, *24*, 497–511. https://doi.org/10.24200/sci.2017.2413

Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification. *Molecules*, *26*. https://doi.org/10.3390/molecules26041111

Ravanshadnia, M., Rajaie, H., & Abbasian, R. (2011). A comprehensive bid/no-bid decision making framework for construction companies. *Iranian journal of science and technology*, *35*, 95–103.

Ren, Q., Li, M., & Han, S. (2019). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: A comparative study from multiple perspectives. *Big Earth Data*, *3*, 8–25. https://doi.org/10.1080/20964471.2019.1572452

Rijksoverheid. (2020). Aanbestedingsregelement werken 2016. *Staatscourant*. Retrieved January 1, 2023, from https://www.pianoo.nl/sites/default/files/media/documents/ARW%5C%202016%5C%20zoals%5C%20gepubliceerd%5C%20in%5C%20Staatscourant%5C%202020.pdf

Rijkswaterstaat. (2019). *Toekomstige opgave rijkswaterstaat*. Rijkswaterstaat.

Shash, A. (1993). Factors considered in tendering decisions by top uk contractors. *Construction Management and Economics*, *11*, 111–118. https://doi.org/10.1080/01446199300000004

Shi, H. (2012). Aco trained ann-based bid/no-bid decision-making. *International Journal of Modelling, Identification and Control*, *15*, 290. https://doi.org/10.1504/IJMIC.2012.046408

Shi, H., Yin, H., & Wei, L. (2016). A dynamic novel approach for bid/no-bid decision-making. *SpringerPlus*, *5*, 1589. https://doi.org/10.1186/s40064-016-3230-1

Skitmore, M., & Pemberton, J. (1994). A multivariate approach to construction contract bidding mark-up strategies. *Journal of the Operational Research Society*, *45*, 1263–1272. https://doi.org/10.1057/jors.1994.199

Skitmore, M., Pettitt, A., & McVinish, R. (2007). Gates' bidding model. *Journal of Construction Engineering and Management*, *133*, 855–863. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:11(855)

Sonmez, R., & Sözgen, B. (2017). A support vector machine method for bid/no bid decision making. *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, *23*, 641–649. https://doi.org/10.3846/13923730.2017.1281836

Straatmeyer, J. (2015). *Monitor bouwketen*. EIB.

TenderNed. (2022). *Datasets aanbestedingen*. Retrieved August 1, 2022, from https://www.tenderned.nl/cms/aanbesteden-in-cijfers/datasets-aanbestedingen

Valle, D., Toh, K., Laporta, G., & Zhao, Q. (2019). Ordinal regression models for zero-inflated and/or over-dispersed count data. *Scientific Reports*, *9*. https://doi.org/10.1038/s41598-019-39377-x

van Nieuwenhuizen, C. (2018a). *Parliament letter 35000 a nr. 89*. https://zoek.officielebekendmakingen.nl/kst-35000-A-89.pdf

van Nieuwenhuizen, C. (2018b). *Parliament letter 29385 nr. 99*. https://zoek.officielebekendmakingen.nl/kst-29385-99.pdf

van Nieuwenhuizen, C. (2018c). *Parliament letter 29385 nr. 100*. https://zoek.officielebekendmakingen.nl/kst-29385-100.pdf

Ver-Hoef, J., & Boveng, P. (2007). *Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data?* (11).

Visser, N. (2014). *Bedrijfseconomische kencijfers*. EIB.

Visser, N. (2015). *Bedrijfseconomische kencijfers*. EIB.

Visser, N. (2016a). *Bedrijfseconomische kencijfers*. EIB.

Visser, N. (2016b). *Monitor bouwketen*. EIB.
Visser, N. (2017a). *Bedrijfseconomische kencijfers*. EIB.
Visser, N. (2017b). *Monitor bouwketen*. EIB.
Visser, N. (2017c). *Monitor bouwketen*. EIB.
Visser, N. (2018a). *Monitor bouwketen*. EIB.
Visser, N. (2018b). *Monitor bouwketen*. EIB.
Visser, N. (2020a). *Bedrijfseconomische kencijfers*. EIB.
Visser, N. (2020b). *Monitor bouwketen*. EIB.
Visser, N. (2020c). *Monitor bouwketen*. EIB.
Visser, N. (2021a). *Monitor bouwketen*. EIB.
Visser, N. (2021b). *Monitor bouwketen*. EIB.
Visser, N. (2022a). *Monitor bouwketen*. EIB.
Visser, N. (2022b). *Monitor bouwketen*. EIB.
Visser, N., & Nicolas, R. (2018). *Bedrijfseconomische kencijfers*. EIB.
Visser, N., & Nicolas, R. (2019). *Bedrijfseconomische kencijfers*. EIB.
Visser, N., & Straatmeyer, J. (2016). *Monitor bouwketen*. EIB.
Vrolijk, M. (2014). *Monitor bouwketen*. EIB.
Vrolijk, M. (2015). *Monitor bouwketen*. EIB.
Vrolijk, M. (2013). *Bedrijfseconomische kencijfers*. EIB.
Wanous, M., Boussabaine, A., & Lewis, J. (2000). To bid or not to bid: A parametric solution. *Construction Management and Economics*, *18*, 457–466. https://doi.org/10.1080/01446190050024879
Wanous, M., Boussabaine, A., & Lewis, J. (2003). A neural network bid/no bid model: The case for contractors in syria. *Construction Management and Economics*, *21*, 737–744. https://doi.org/10.1080/0144619032000093323

A

Tables of bid decision factors most mentioned in the literature

Table A.1: Bid decision factors from Kalan & Ozbek (Kalan & Ozbek, 2020)

Factor number:	Article factor description:
1	Current workload
2	Experience with similar projects
3	Availability of equipment, materials, and human resources
4	Financial ability
5	Need for work
6	Technical know-how
7	Compliance with business plan
8	Project size
9	Project duration
10	Project location
11	Project type
12	Contract conditions and type
13	Owner identity
14	Competition

Table A.2: Bid decision factors from Ravanshadnia et al. (Ravanshadnia et al., 2011)

Factor number:	Article factor description:
1	Company workload and need for work
2	Project prestige, public exposure and its strategic fitness to policy
3	Project client, supervisor, and other stakeholders characteristics
4	Competitor's approach and the probability of winning the bid
5	Availability of time and human resource for tendering
6	Project complexity and company's familiarity with this kind of work
7	Duration of the project and its schedule feasibility
8	Site conditions (Accessibility and Space for Work)
9	The availability of the needed material, equipment, sub-contractors, and suppliers
10	Constructability of the work method and technical documents
11	Having no resource price fluctuation and general inflation effects
12	Disbursing payment without delays
13	Suitable climate and weather and geographical conditions
14	Stabling laws, standards, and requirements
15	Lack of probable claims and their effects
16	Expected benefit/loss and its rate of return
17	Project cash flow and payment conditions
18	Tendering bond size, bidding document price
19	Client's financial capability and its payment policy
20	The value of project advanced payment and its maximum required cash
21	Project distance from existing projects and facilities of the company
22	The similarity of the project type and size to other company projects
23	Project resource similarity and its influence on existing project performance
24	The similarity of client or supervisor to the existing projects
25	Project cash flow interrelation with existing projects

Table A.3: Bid decision factors from Cheng et al. (Cheng et al., 2011)

Factor rank:	How often mentioned:	Article factor description:
1	7	Nature of project
2	6	Reputation of client
3	6	Relationship with client
4	6	Project location
5	6	Experience for similar project
6	6	Time available for tender preparation
7	6	Contractual conditions
8	6	Availability of other projects
9	5	Project size
10	5	Project complexity
11	5	Availability of qualified/experienced staff
12	5	Current workload
13	5	Number of competitors
14	5	Expected profitability
15	5	Market conditions
16	5	Expected risk
17	4	Project period
18	4	Adequacy of tender information
19	4	Financial situation
20	3	Financial capability of the client
21	3	Proportions to be subcontracted
22	3	Relationship with other consultants
23	3	Cost of bidding
24	3	Tender conditions
25	3	Tendering method
26	3	Type of contract
27	3	Degree of competition
28	3	Expected cash flow
29	2	Client requirements
30	2	Reputation of other consultants
31	2	Professional demands of the contract
32	2	Compliance with business strategy
33	2	Promoting reputation
34	2	Operational capacity
35	2	Competence of the expected competitors
36	2	Perceived chances of being successful
37	2	Client budget
38	2	Confidence in the cost estimate
39	2	Local customs
40	1	Fostering good relationship with regular clients
41	1	Physical resources necessary to carry out project
42	1	Financial resources necessary to carry out project
43	1	Current workload in bid preparation
44	1	Project break-even point for the contract

Table A.4: Part 1 bid decision factors from Mehrabani et al. (Mehrabani et al., 2020)

Factor rank:	**How often mentioned:**	**Article factor description:**
1	10	The availability of the needed human resources
2	10	Uncertainty of costs
3	9	Availability of required equipment
4	9	Availability of required staff
5	9	Client/owner identity
6	9	Current workload in company
7	9	Project complexity and related technology
8	9	Reliability of subcontractors
9	8	Workshop conditions
10	8	Competition
11	8	Availability to subcontractors
12	8	Required cash
13	8	Confidence in workforce
14	7	Project location
15	7	Need for work in company
16	7	Invest return rate (Rate of return)
17	7	Project supervisor (Consultant Engineers)
18	7	Project duration
19	7	Relationship or having previous experience with owner
20	6	The availability of the needed material, equipment, subcontractors and suppliers
21	6	Complete documentation and information
22	6	Experience similar projects
23	6	Investment risk
24	6	Cash flow
25	6	Public exposure
26	6	Contract type
27	6	Governmental requirements/Government regulations
28	6	Experience of Consulting Engineers
29	6	Size of project
30	5	General overhead
31	5	Profit history of similar projects
32	5	Client financial capability and its payment policy
33	5	Time allowed for submitting bids
34	5	Prequalification requirements
35	5	Bidding method (open/close)
36	5	Company's ability in design, involvement, innovation
37	5	Company strength in the industry
38	5	Economic conditions of the company
39	4	Project risk rate
40	4	Project start time
41	4	Bid Bond size and validity
42	4	Bidding document price

Table A.5: Part 2 bid decision factors from Mehrabani et al. (Mehrabani et al., 2020)

43	4	Tax laws
44	4	Type of project
45	4	Design quality
46	4	Designer
47	4	Special conditions of contract
48	3	Bidding time
49	3	Possible events
50	3	Contractor involvement in the design phase
51	3	Market conditions
52	2	The availability of other projects
53	2	Suitable climate and weather geographical conditions
54	2	Quality of available labor
55	2	Degree of difficulty in obtaining bank loan
56	2	Possession of Subcontractors
57	2	Conformity of project with company's speciality
58	2	Public exposure of project
59	1	Possession of qualified labor
60	1	Possession of qualified staff
61	1	The status of sanctions and its impact on economic conditions
62	1	Exchange rate fluctuations and impact on economic conditions
63	1	Familiarity and communication with the suppliers
64	1	Project distance to existing projects of the company
65	1	The similarity of the project type and size to other company's projects
66	1	Project resource similarity and its influence on existing projects performance
67	1	Stabling laws standards and requirements
68	1	The similarity of client or supervisor to the existing projects
69	1	Possession of required equipment
70	1	Insurance premium
71	1	Public objection
72	1	Being the main or the sub-contractor

B

Overview data

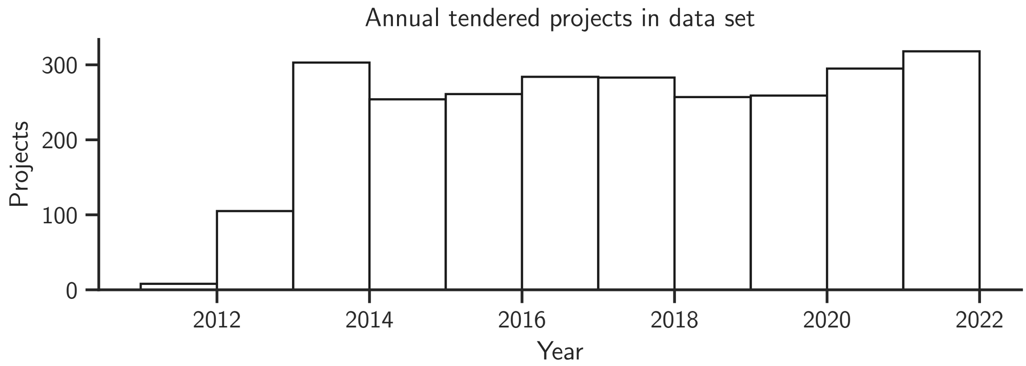

Figure B.1: Overview of the number of projects per year present in the data

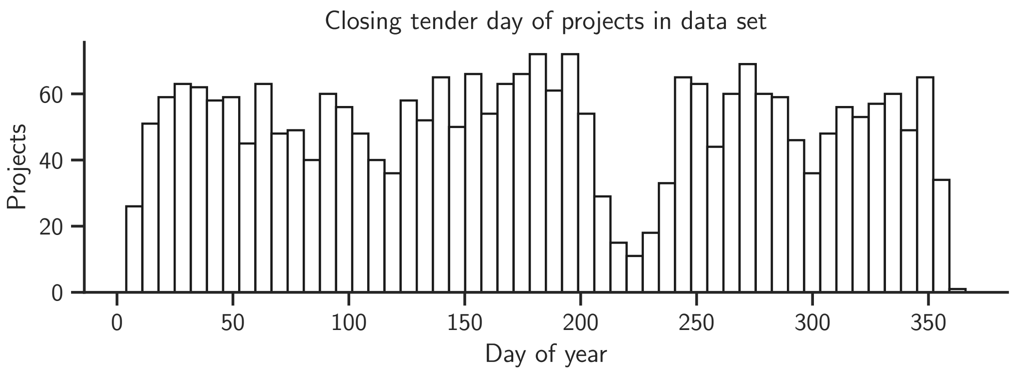

Figure B.2: Overview of the occurring tender closing days in the data

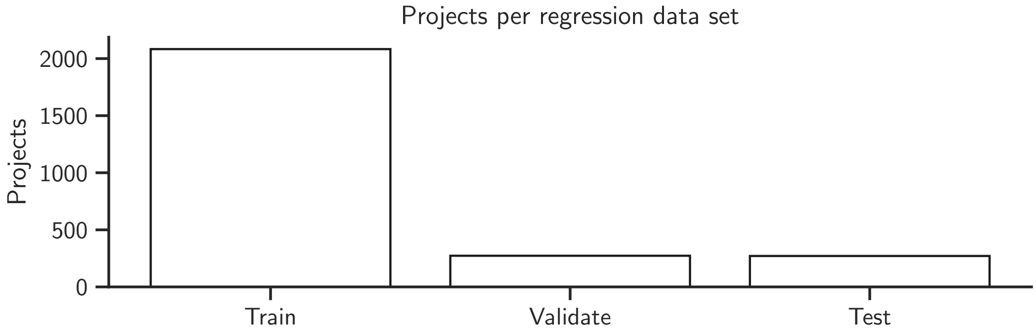

Figure B.3: Visualization of data splitting prior analysis

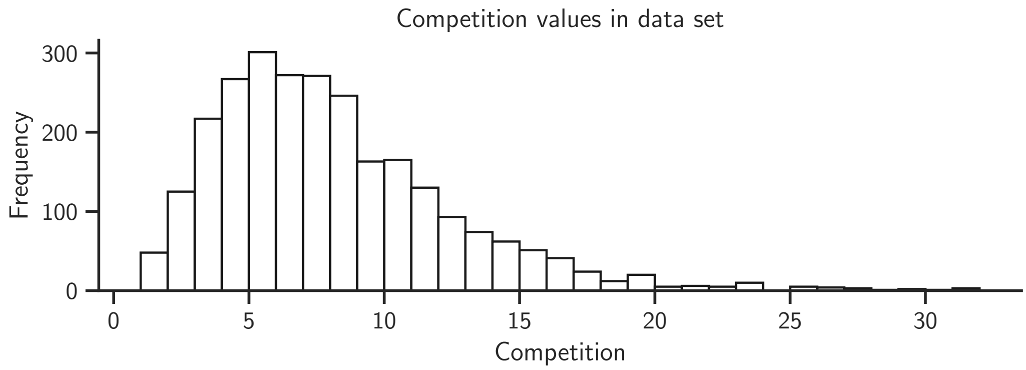

Figure B.4: Frequency of competition values within the data

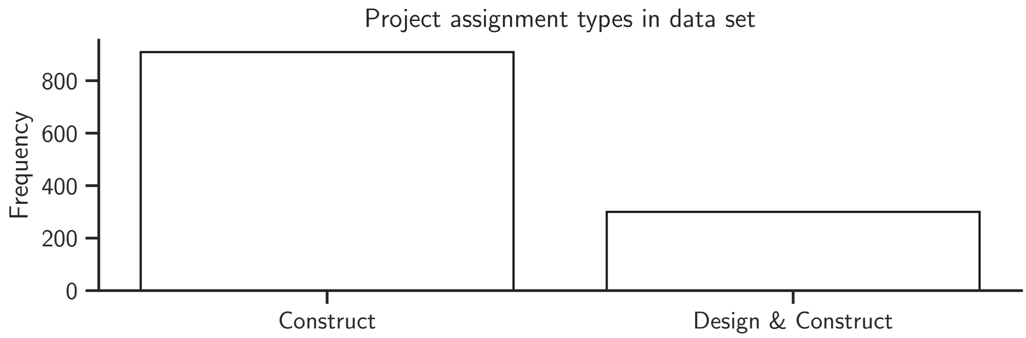

Figure B.5: Frequency of different assignment types within data

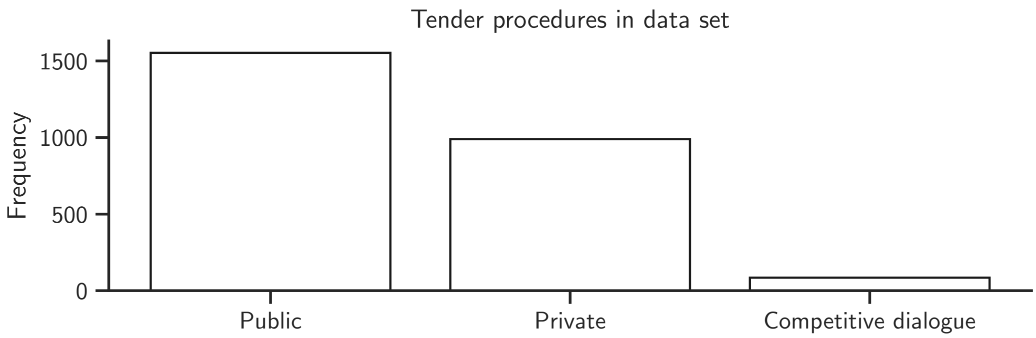

Figure B.6: Tender procedures in the data

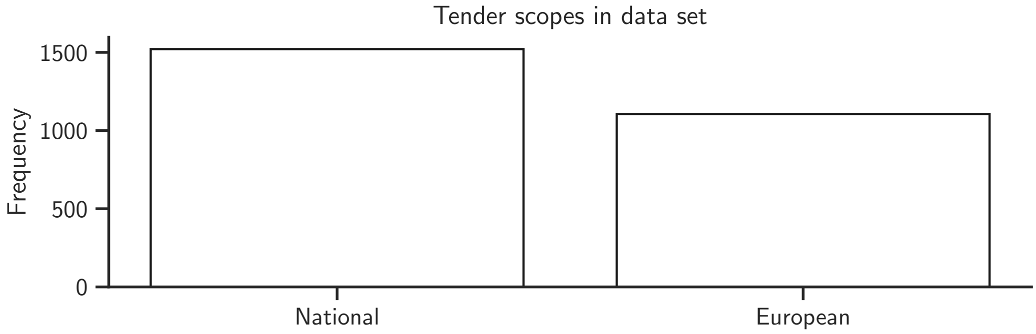

Figure B.7: Visualization of tender scope in the data

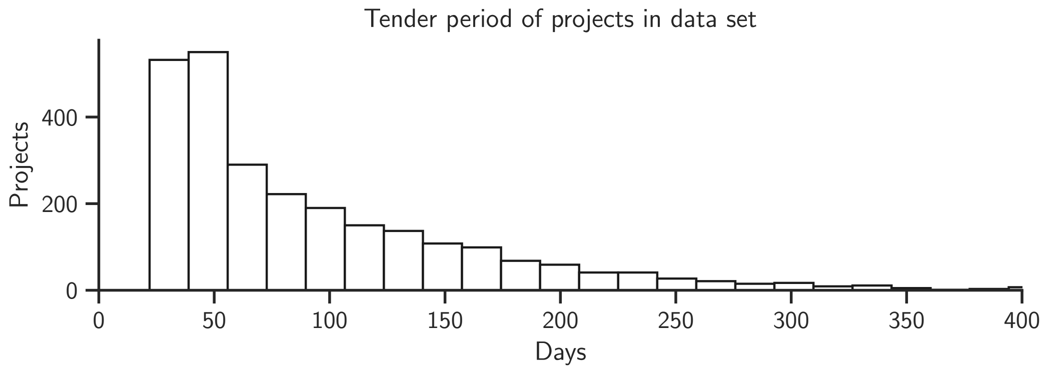

Figure B.8: Frequency of tender periods within data set

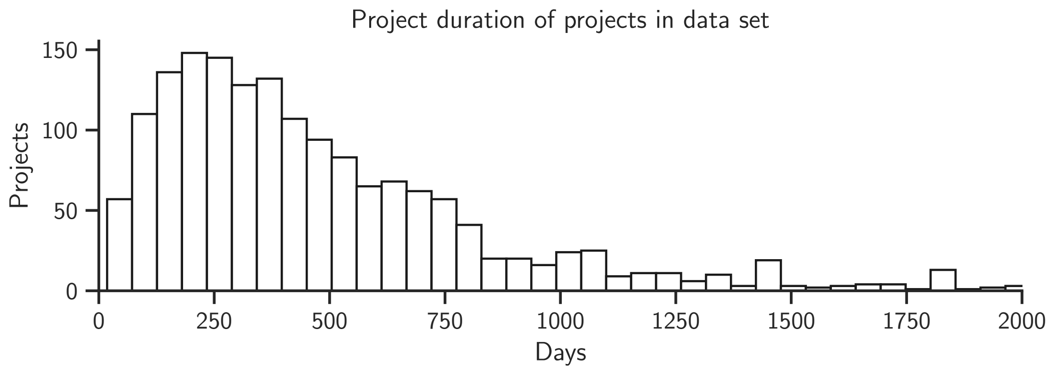

Figure B.9: Visualization of project duration's present in the data

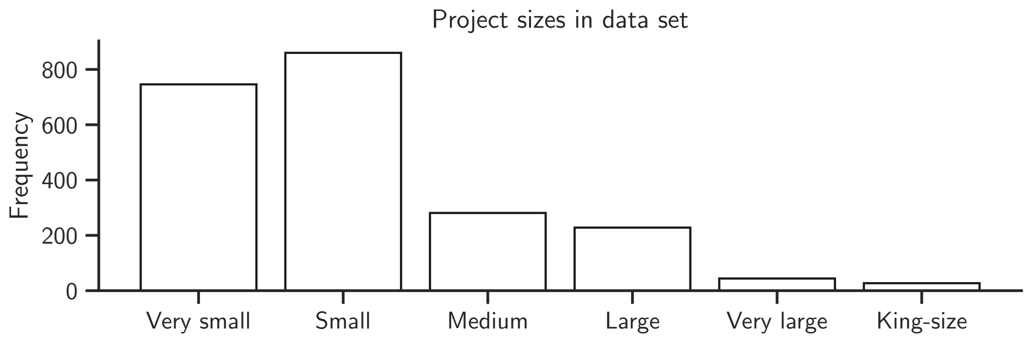

Figure B.10: Frequency of different project sizes within the data

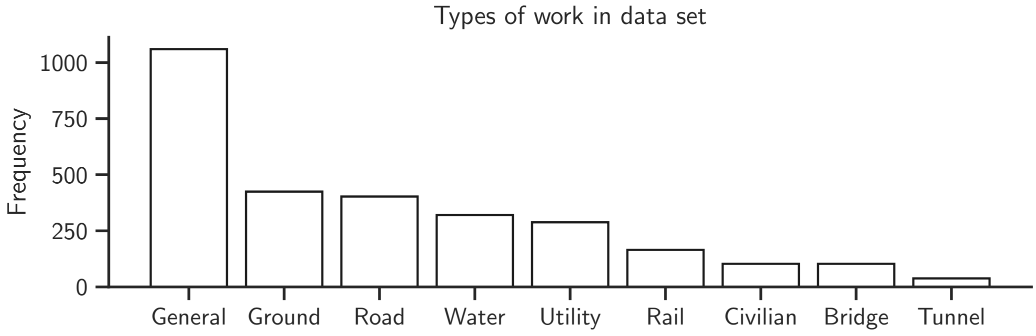

Figure B.11: Types of work and their frequency present within the data

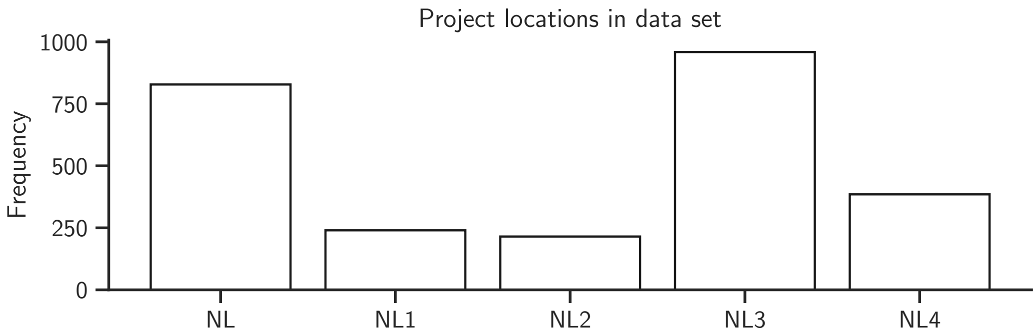

Figure B.12: Project locations of projects in the data

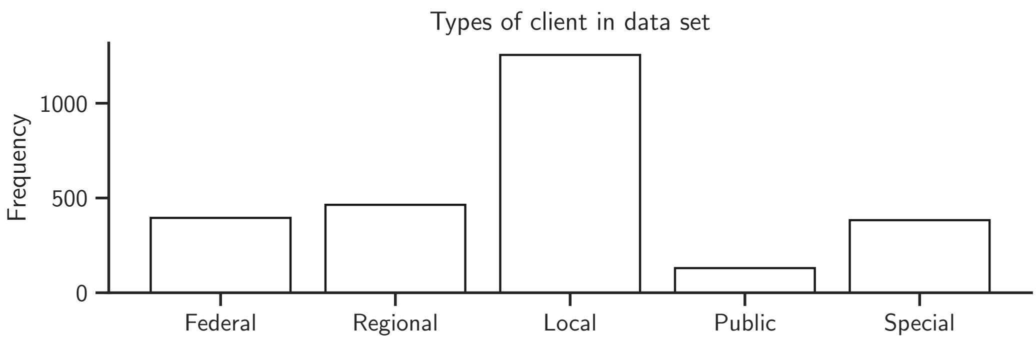

Figure B.13: Frequency of client types on project in the data

C
Variable effects

Table C.1: Variable effects on competition based on regression coefficients and average variables value from "All"-model

Variable:	Regression coefficient:	Average value:	Variable effect:
Revenue of infrastructure market	0.2565	6.8	1.74
Revenue of construction market	0.0213	26.0	0.55
Tunnel construction work	0.4787	1.0	0.48
Local government client	0.4111	1.0	0.41
Restricted tender procedure	0.3737	1.0	0.37
2017 year of tender	0.3676	1.0	0.37
2016 year of tender	0.3477	1.0	0.35
Hydraulic construction work	0.2688	1.0	0.27
Ground construction work	0.2325	1.0	0.23
General construction work	0.2073	1.0	0.21
Project size very small	0.2042	1.0	0.20
Regional government client	0.1957	1.0	0.20
2015 year of tender	0.1885	1.0	0.19
Utility construction work	0.1884	1.0	0.19
Road construction work	0.1232	1.0	0.12
Nutscode NL2	0.1177	1.0	0.12
Project size large	0.113	1.0	0.11
Project size medium	0.0948	1.0	0.10
Bridge construction work	0.0881	1.0	0.09
Intercept	0.0866	1.0	0.09
Construct contract	0.085	1.0	0.09
European tender scope	0.0736	1.0	0.07
NL 10 year bond yield	0.0575	1.2	0.09
Day of year tender ends	0.0002	187.1	0.04
Project size small	0.0345	1.0	0.04
Nutscode NL4	0.0219	1.0	0.02
Nutscode NL3	0.0136	1.0	0.01
National tender scope	0.013	1.0	0.01
Design and construct contract	0.0016	1.0	0.00
2011 year of tender	0.0	1.0	0.00
2019 year of tender	0.0	1.0	0.00
2018 year of tender	0.0	1.0	0.00
2020 year of tender	0.0	1.0	0.00
2021 year of tender	0.0	1.0	0.00
2022 year of tender	0.0	1.0	0.00
Nutscode NL1	-0.003	1.0	0.00
Length of tender period	-0.0001	83.5	-0.01
Prospected project duration	-0.00002	521.1	-0.01
2014 year of tender	-0.0342	1.0	-0.03
Nutscode NL	-0.0636	1.0	-0.06
Open vacancies in sector	-0.0196	4.6	-0.09
Public tender procedure	-0.0957	1.0	-0.10
Project size very large	-0.0988	1.0	-0.10
Central government client	-0.1186	1.0	-0.12
Installation work	-0.1191	1.0	-0.12
Special sector client	-0.1449	1.0	-0.15
Competitive dialogue procedure	-0.1914	1.0	-0.19
Public entity client	-0.2567	1.0	-0.26
Project size king-size	-0.2611	1.0	-0.26
2013 year of tender	-0.3226	1.0	-0.32
Rail construction work	-0.3256	1.0	-0.33
2012 year of tender	-0.4603	1.0	-0.46
Market orderbook	-0.1594	5.7	-0.91
Civilian construction work	-1.0952	1.0	-1.10

Table C.2: Variable effects on competition based on regression coefficients and average variables value from "Most"-model

Variable:	Regression coefficient:	Average value:	Variable effect:
Revenue of infrastructure market	0.1721	7.6	1.31
Intercept	0.365	1.0	0.37
2015 year of tender	0.3547	1.0	0.36
Project size very small	0.3237	1.0	0.32
Restricted tender procedure	0.3064	1.0	0.31
Local government client	0.3021	1.0	0.30
2014 year of tender	0.2952	1.0	0.30
2016 year of tender	0.2898	1.0	0.29
Civillian construction work	0.2505	1.0	0.25
2017 year of tender	0.2319	1.0	0.23
Revenue of construction market	0.0078	29.7	0.23
Ground construction work	0.212	1.0	0.21
National tender scope	0.2004	1.0	0.20
Competitive dialogue procedure	0.193	1.0	0.19
Project size small	0.1663	1.0	0.17
European tender scope	0.1646	1.0	0.17
Hydraulic construction work	0.1641	1.0	0.16
Nutscode NL2	0.163	1.0	0.16
Project size large	0.1287	1.0	0.13
Regional government client	0.1124	1.0	0.11
General construction work	0.1105	1.0	0.11
Bridge construction work	0.1019	1.0	0.10
Project size medium	0.0961	1.0	0.10
Nutscode NL1	0.0959	1.0	0.10
Road construction work	0.0939	1.0	0.09
2013 year of tender	0.087	1.0	0.09
Nutscode NL4	0.0814	1.0	0.08
2018 year of tender	0.0799	1.0	0.08
Public entity client	0.07	1.0	0.07
2012 year of tender	0.0327	1.0	0.03
Tunnel construction work	0.0248	1.0	0.03
Nutscode NL	0.0201	1.0	0.02
Nutscode NL3	0.0046	1.0	0.01
2011 year of tender	0.0	1.0	0.00
Central government client	-0.0039	1.0	0.00
Length of tender period	-0.0001	101.4	-0.01
Day of year tender ends	-0.0001	182.1	-0.02
NL 10 year bond yield	-0.0838	0.6	-0.05
Project size very large	-0.0586	1.0	-0.06
Open vacancies in sector	-0.009	10.8	-0.10
Special sector client	-0.1158	1.0	-0.12
Public tender procedure	-0.1345	1.0	-0.13
2019 year of tender	-0.1661	1.0	-0.17
Rail construction work	-0.2045	1.0	-0.20
Installation work	-0.2184	1.0	-0.22
2020 year of tender	-0.2325	1.0	-0.23
Utility construction work	-0.2375	1.0	-0.24
2021 year of tender	-0.2657	1.0	-0.27
Project size king-size	-0.2912	1.0	-0.29
2022 year of tender	-0.342	1.0	-0.34
Market orderbook	-0.0643	7.2	-0.46

Table C.3: The variable effects distilled from the "Best"-model

Variable:	Regression coefficient:	Average value:	Variable effect:
Revenue of infrastructure market	0.1353	7.4	1.00
Revenue of construction market	0.0302	29.3	0.88
Market orderbook	0.052	7.0	0.36
Restricted tender procedure	0.3092	1.0	0.31
2015 year of tender	0.2659	1.0	0.27
Nutscode NL1	0.2444	1.0	0.24
Local government client	0.1811	1.0	0.18
2014 year of tender	0.1767	1.0	0.18
2017 year of tender	0.1763	1.0	0.18
Intercept	0.1149	1.0	0.12
National tender scope	0.0995	1.0	0.10
Nutscode NL2	0.0945	1.0	0.10
2018 year of tender	0.0798	1.0	0.08
Day of year tender ends	0.0004	181.8	0.07
Project size very small	0.0629	1.0	0.06
Project size king-size	0.0369	1.0	0.04
Project size medium	0.0338	1.0	0.03
2016 year of tender	0.0336	1.0	0.03
Nutscode NL4	0.0311	1.0	0.03
Project size very large	0.0277	1.0	0.03
2012 year of tender	0.0179	1.0	0.02
European tender scope	0.0154	1.0	0.02
Project size large	0.0154	1.0	0.02
Public entity client	-0.0	1.0	0.00
Regional government client	-0.0002	1.0	0.00
2011 year of tender	-0.0	1.0	0.00
Special sector client	-0.0203	1.0	-0.02
NL 10 year bond yield	-0.0359	0.7	-0.03
2019 year of tender	-0.0258	1.0	-0.03
Central government client	-0.0456	1.0	-0.05
Project size small	-0.0618	1.0	-0.06
2020 year of tender	-0.0633	1.0	-0.06
Competitive dialogue procedure	-0.0641	1.0	-0.06
2022 year of tender	-0.0714	1.0	-0.07
Nutscode NL3	-0.0751	1.0	-0.08
Length of tender period	-0.0011	88.4	-0.10
2021 year of tender	-0.12	1.0	-0.12
Public tender procedure	-0.1302	1.0	-0.13
Nutscode NL	-0.18	1.0	-0.18
2013 year of tender	-0.3548	1.0	-0.36
Open vacancies in sector	-0.0439	10.2	-0.45

D
Comparison variable effects

Table D.1: Comparison of the variable effects of the three covered models

Variable:	”All” effect:	”Most” effect:	”Best” effect:
Intercept	0.09	0.37	0.12
Project size very small	0.20	0.32	0.06
Project size small	0.04	0.17	-0.06
Project size medium	0.10	0.10	0.03
Project size large	0.11	0.13	0.02
Project size very large	-0.10	-0.06	0.03
Project size king-size	-0.26	-0.29	0.04
Nutscode NL	-0.06	0.02	-0.18
Nutscode NL1	0.00	0.10	0.24
Nutscode NL2	0.12	0.16	0.10
Nutscode NL3	0.01	0.01	-0.08
Nutscode NL4	0.02	0.08	0.03
Central government client	-0.12	0.00	-0.05
Local government client	0.41	0.30	0.18
Public entity client	-0.26	0.07	0.00
Regional government client	0.20	0.11	0.00
Special sector client	-0.15	-0.12	-0.02
European tender scope	0.07	0.17	0.02
National tender scope	0.01	0.20	0.10
Competitive dialogue tender procedure	-0.19	0.19	-0.06
Public tender procedure	-0.10	-0.13	-0.13
Restricted tender procedure	0.37	0.31	0.31
Installation work	-0.12	-0.22	
Bridge construction work	0.09	0.10	
Tunnel construction work	0.48	0.03	
Utility construction work	0.19	-0.24	
Civillian construction work	-1.10	0.25	
Rail construction work	-0.33	-0.20	
Hydraulic construction work	0.27	0.16	
Road construction work	0.12	0.09	
Ground construction work	0.23	0.21	
General construction work	0.21	0.11	
Length of tender period	-0.01	-0.01	-0.10
Day of year tender ends	0.04	-0.02	0.07
2011 year of tender	0.00	0.00	0.00
2012 year of tender	-0.46	0.03	0.02
2013 year of tender	-0.32	0.09	-0.36
2014 year of tender	-0.03	0.30	0.18
2015 year of tender	0.19	0.36	0.27
2016 year of tender	0.35	0.29	0.03
2017 year of tender	0.37	0.23	0.18
2018 year of tender	0.00	0.08	0.08
2019 year of tender	0.00	-0.17	-0.03
2020 year of tender	0.00	-0.23	-0.06
2021 year of tender	0.00	-0.27	-0.12
2022 year of tender	0.00	-0.34	-0.07
Construct contract	0.09		
Design and construct contract	0.00		
Months of work in market orderbook	-0.91	-0.46	0.36
Revenue of construction market	0.55	0.23	0.88
Revenue of infrastructure market	1.74	1.31	1.00
Open vacancies in construction sector	-0.09	-0.10	-0.45
NL 10 year bond yield	0.07	-0.05	-0.03
Prospected project duration	-0.01		

E

Entropy and Mutual information

Table E.1: Calculation of entropy of the competition distribution

Competition value:	**Probability:**	**Competition value entropy:**
1	0.0183	0.1056
2	0.0476	0.2091
3	0.0826	0.2972
4	0.1016	0.3352
5	0.1146	0.3582
6	0.1035	0.3387
7	0.1032	0.3381
8	0.0936	0.3199
9	0.0620	0.2487
10	0.0628	0.2508
11	0.0495	0.2147
12	0.0354	0.1706
13	0.0282	0.1452
14	0.0236	0.1276
15	0.0194	0.1103
16	0.0156	0.0936
17	0.0091	0.0617
18	0.0046	0.0357
19	0.0076	0.0535
20	0.0019	0.0172
21	0.0023	0.0202
22	0.0019	0.0172
23	0.0038	0.0306
24	0.0000	0.0000
25	0.0019	0.0172
26	0.0015	0.0141
27	0.0011	0.0108
28	0.0004	0.0045
29	0.0008	0.0082
30	0.0004	0.0045
31	0.0004	0.0045
Entropy:		**3.9634**

Table E.2: Calculation of mutual information between the compatition and the other variables

Variable:	Mutual Information
Project size	0.0453
Nutscode	0.0332
Client type	0.0416
Tender scope	0.0102
Tender procedure	0.0426
General construction work	0.0099
Bridge work	0.0047
Tunnel work	0.0035
Utility work	0.0150
Civilian work	0.0068
Rail work	0.0101
Hydraulic work	0.0094
Road work	0.0143
Ground work	0.0120
Installation work	0.0124
Tender period	0.1035
Year	0.0808
Tender closing day	0.1135
Contract type	0.0202
Orderbook market	0.0693
Revenue construction sector	0.0786
Revenue infrastructure sector	0.0303
Open vacancies	0.1305
NL 10 year bond yield	0.1430
Project duration	0.2007
Sum Mutual Information:	1.2414

F

Additional model set-up results

Only restricted tender procedure projects

Table F.1: Initial regression results on only restricted tender procedure projects

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	34.5	28.1	40.5	34.1	27.1	40.5
Best-guess accuracy (%)	29.5	34.4	30.4	29.5	34.4	30.4
Pearson χ^2	1059.6	NaN	NaN	639.9	NaN	NaN
Allowed Pearson χ^2	710.4	NaN	NaN	710.4	NaN	NaN
Deviance	1055.4	NaN	NaN	644.9	NaN	NaN

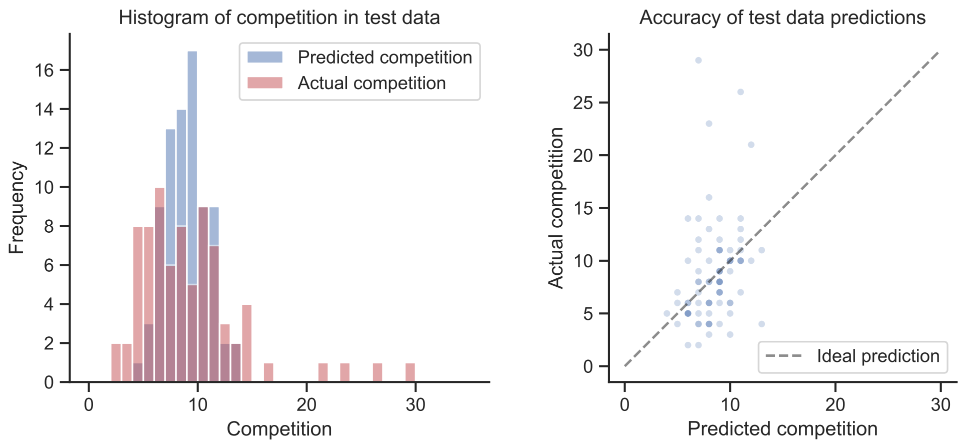

Figure F.1: Visualization of Negative Binomial predictions on the test data that contains only restricted tender procedure projects

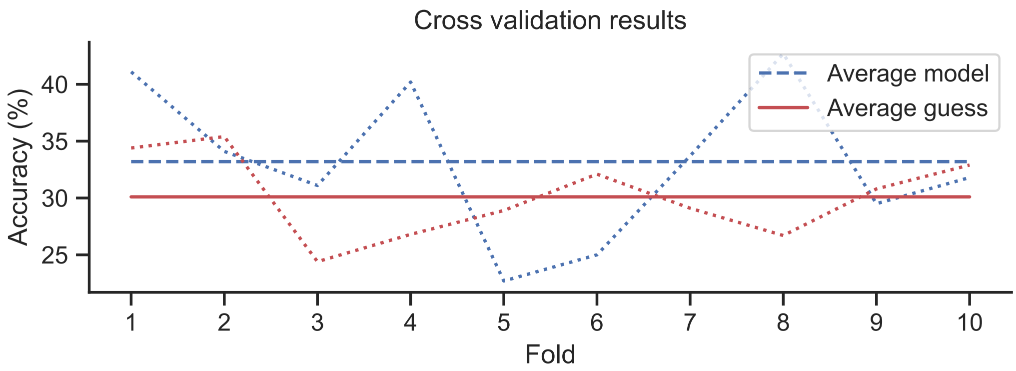

Figure F.2: Negative Binomial Cross validation results on projects that contain only restricted tender procedure projects

Table F.2: Negative Binomial Cross validation results on projects that contain only restricted tender procedure projects

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**average**
Model accuracy (%)	41.1	34.1	31.1	40.2	22.7	25.0	33.7	42.7	29.5	31.8	33.2
Guess accuracy (%)	34.4	35.4	24.4	26.8	28.9	32.1	29.1	26.7	30.8	32.9	30.1

Only public tender procedure projects

Table F.3: Initial regression results on only public tender procedure projects

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	37.2	30.6	30.3	36.7	31.5	32.4
Best-guess accuracy (%)	34.2	26.6	26.2	34.2	26.6	26.2
Pearson χ^2	1706.0	NaN	NaN	906.5	NaN	NaN
Allowed Pearson χ^2	997.9	NaN	NaN	997.9	NaN	NaN
Deviance	1621.7	NaN	NaN	862.9	NaN	NaN

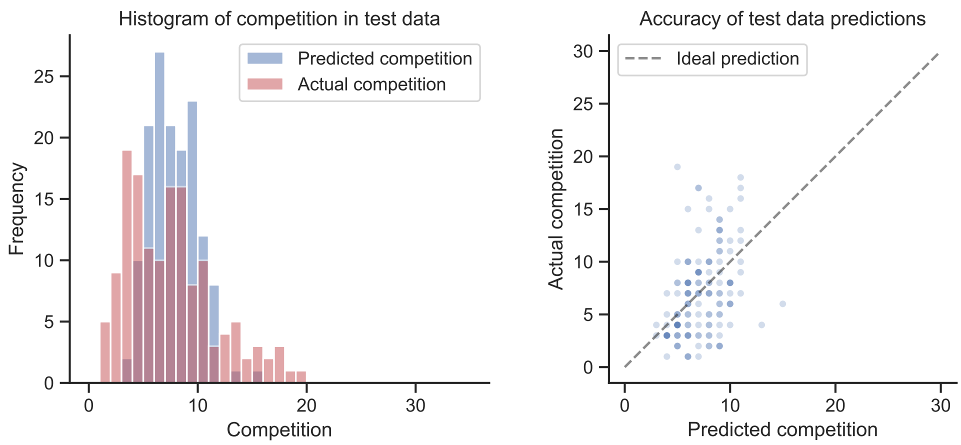

Figure F.3: Visualization of Negative Binomial predictions on the test data that contains only public tender procedure projects

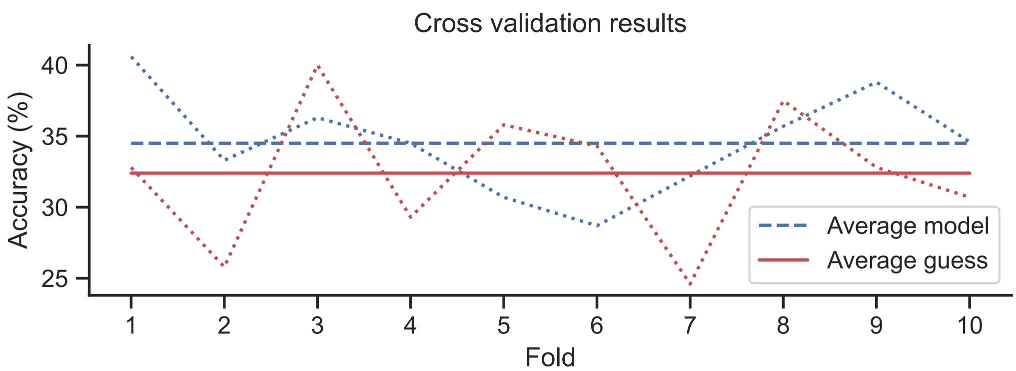

Figure F.4: Negative Binomial Cross validation results on projects that contain only public tender procedure projects

Table F.4: Negative Binomial Cross validation results on only public tender procedure projects

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	40.6	33.3	36.3	34.5	30.7	28.7	32.2	35.7	38.8	34.6	34.5
Guess accuracy (%)	32.8	25.8	40.0	29.3	35.8	34.3	24.6	37.5	32.8	30.7	32.4

Without projects labeled Nutscode NL or General type of work

Table F.5: Initial regression results on projects without labels Nutscode NL or General type of work

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	36.7	41.5	44.1	37.4	39.4	44.1
Best-guess accuracy (%)	33.2	33.0	28.4	33.2	33.0	28.4
Pearson χ^2	1092.8	NaN	NaN	659.6	NaN	NaN
Allowed Pearson χ^2	738.6	NaN	NaN	738.6	NaN	NaN
Deviance	1058.5	NaN	NaN	640.9	NaN	NaN

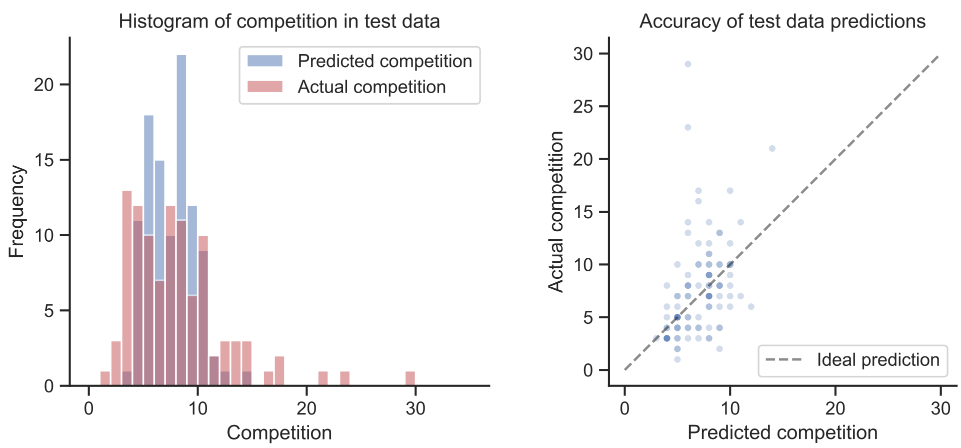

Figure F.5: Visualization of Negative Binomial predictions on the test data that contains only projects without labels Nutscode NL or General type of work

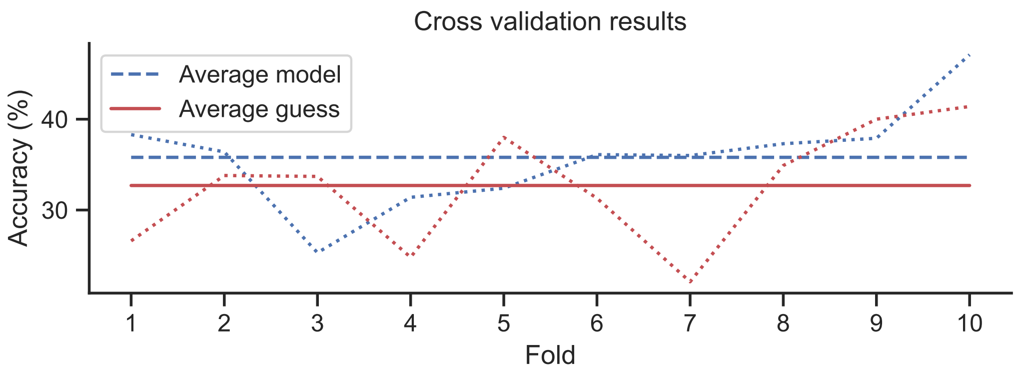

Figure F.6: Negative Binomial Cross validation results on projects without labels Nutscode NL or General type of work

Table F.6: Negative Binomial Cross validation results on projects without labels Nutscode NL or General type of work

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	38.3	36.4	25.3	31.4	32.4	36.1	36.0	37.3	37.9	47.1	35.8
Guess accuracy (%)	26.6	33.8	33.7	24.8	38.0	31.3	22.1	34.9	40.0	41.4	32.7

Only very small project size projects

Table F.7: Initial regression results on only projects labeled very-small

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	28.1	32.9	33.3	27.6	31.6	35.6
Best-guess accuracy (%)	25.5	27.8	32.2	25.5	27.8	32.2
Pearson χ^2	1218.7	NaN	NaN	568.0	NaN	NaN
Allowed Pearson χ^2	602.5	NaN	NaN	602.5	NaN	NaN
Deviance	1173.6	NaN	NaN	552.2	NaN	NaN

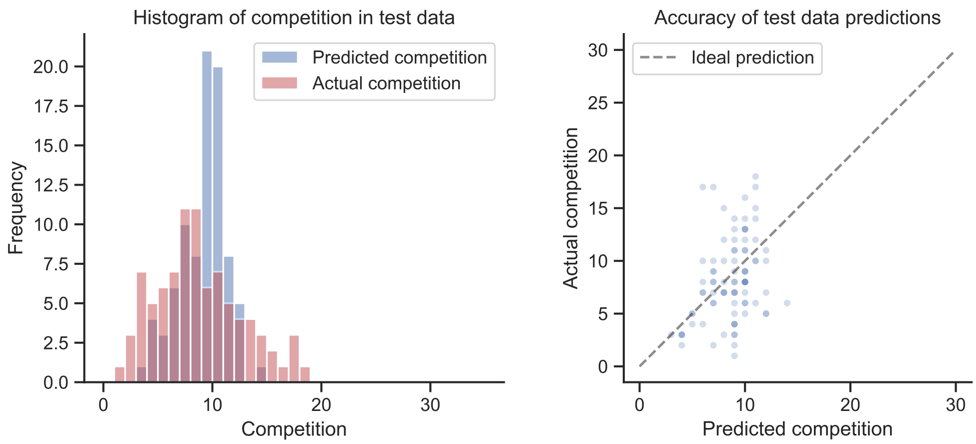

Figure F.7: Visualization of Negative Binomial predictions on the test data that contains only projects labeled very-small

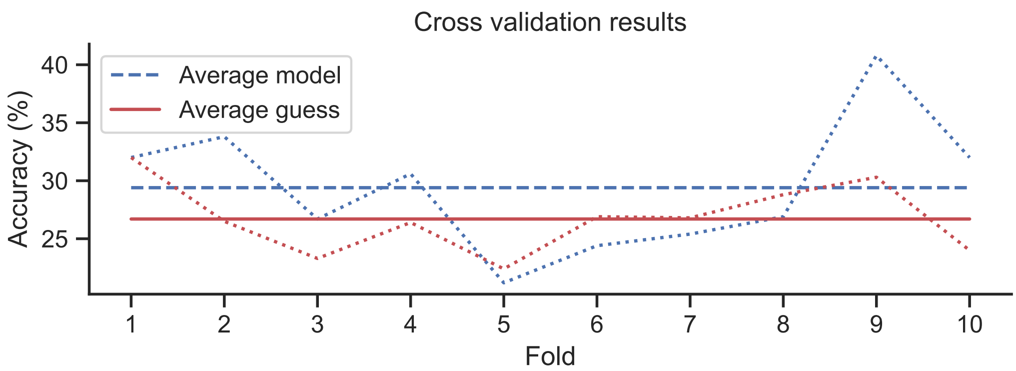

Figure F.8: Negative Binomial Cross validation results on projects labeled very-small

Table F.8: Negative Binomial Cross validation results on projects labeled very-small

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	32.0	33.8	26.7	30.6	21.2	24.4	25.4	26.9	40.8	32.0	29.4
Guess accuracy (%)	32.0	26.5	23.3	26.4	22.4	26.9	26.8	28.8	30.3	24.0	26.7

Only small project size projects

Table F.9: Initial regression results on only projects labeled small

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	34.3	39.5	32.9	35.3	38.4	35.3
Best-guess accuracy (%)	30.8	33.7	35.3	30.8	33.7	35.3
Pearson χ^2	1187.7	NaN	NaN	650.6	NaN	NaN
Allowed Pearson χ^2	719.8	NaN	NaN	719.8	NaN	NaN
Deviance	1141.7	NaN	NaN	626.8	NaN	NaN

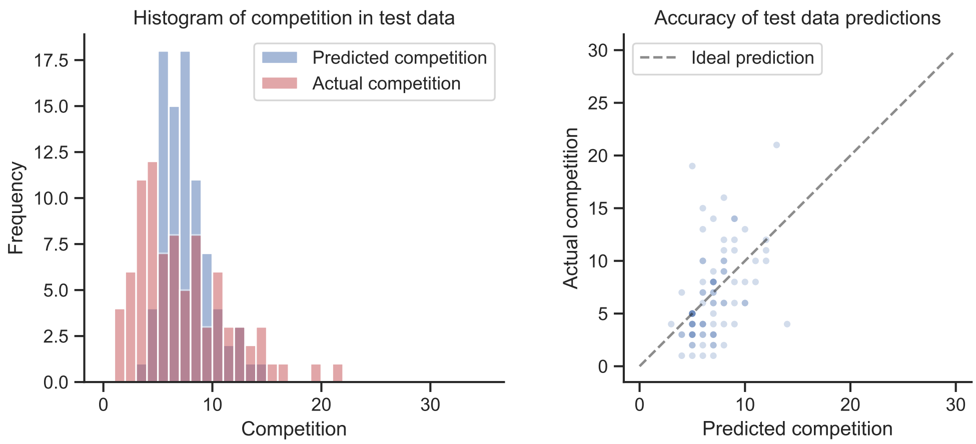

Figure F.9: Visualization of Negative Binomial predictions on the test data that contains only projects labeled small

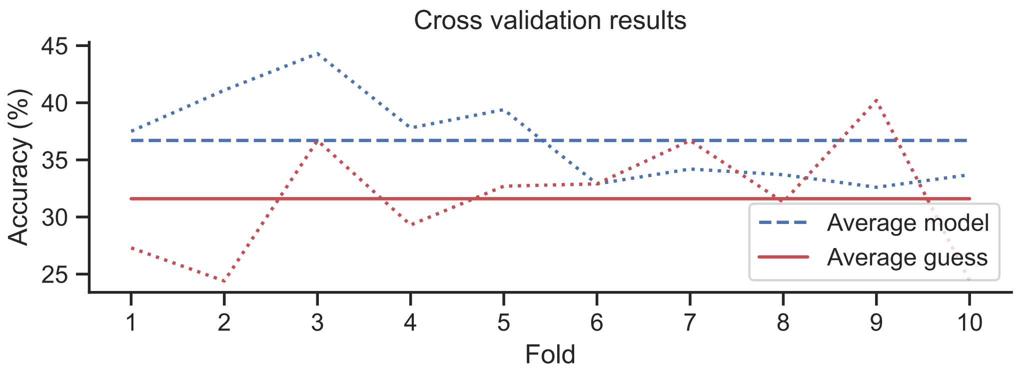

Figure F.10: Negative Binomial Cross validation results on projects labeled small

Table F.10: Negative Binomial Cross validation results on projects labeled small

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**average**
Model accuracy (%)	37.5	41.1	44.3	37.8	39.4	32.9	34.2	33.7	32.6	33.7	36.7
Guess accuracy (%)	27.3	24.4	36.7	29.3	32.7	32.9	36.7	31.3	40.2	24.4	31.6

Only nutscode NL projects

Table F.11: Initial regression results on only projects labeled with project location NL

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	35.4	36.2	38.0	36.6	34.8	39.4
Best-guess accuracy (%)	32.0	33.3	33.8	32.0	33.3	33.8
Pearson χ^2	929.0	NaN	NaN	554.9	NaN	NaN
Allowed Pearson χ^2	584.7	NaN	NaN	584.7	NaN	NaN
Deviance	897.8	NaN	NaN	536.2	NaN	NaN

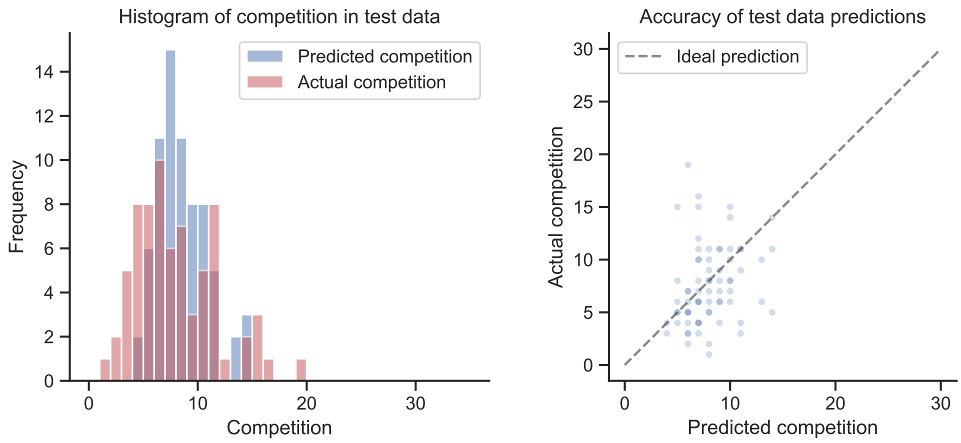

Figure F.11: Visualization of Negative Binomial predictions on the test data that contains only projects labeled with project location NL

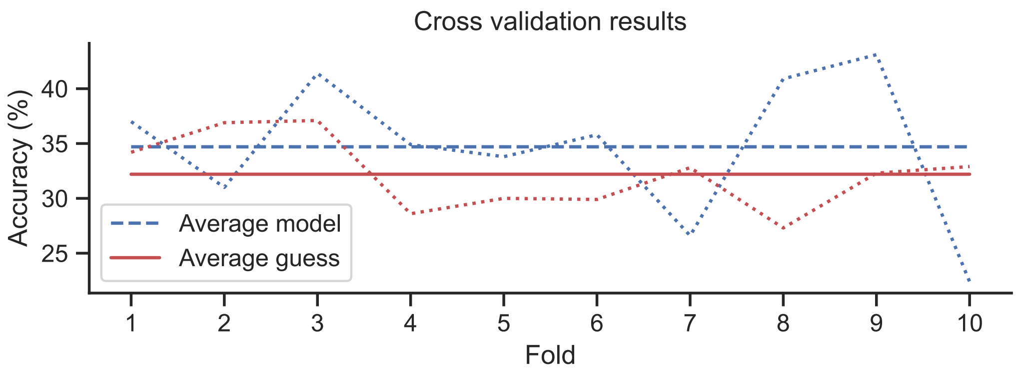

Figure F.12: Negative Binomial Cross validation results on projects labeled with project location NL

Table F.12: Negative Binomial Cross validation results on projects labeled with project location NL

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	37.0	31.0	41.4	34.9	33.8	35.8	26.6	40.9	43.1	22.4	34.7
Guess accuracy (%)	34.2	36.9	37.1	28.6	30.0	29.9	32.8	27.3	32.3	32.9	32.2

Only nutscode NL3 projects

Table F.13: Initial regression results on only projects labeled with project location NL3

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	34.0	32.4	32.6	34.3	33.8	32.6
Best-guess accuracy (%)	31.7	38.0	36.0	31.7	38.0	36.0
Pearson χ^2	1043.0	NaN	NaN	570.8	NaN	NaN
Allowed Pearson χ^2	649.7	NaN	NaN	649.7	NaN	NaN
Deviance	1061.0	NaN	NaN	594.9	NaN	NaN

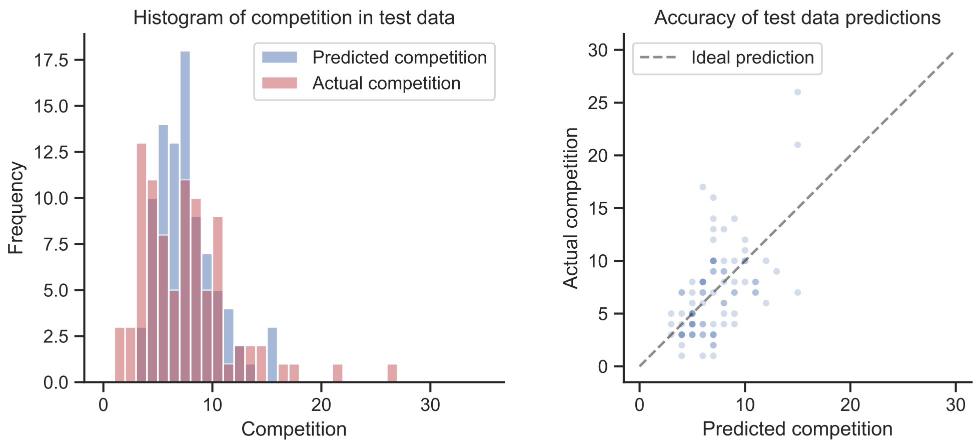

Figure F.13: Visualization of Negative Binomial predictions on the test data that contains only projects labeled with project location NL3

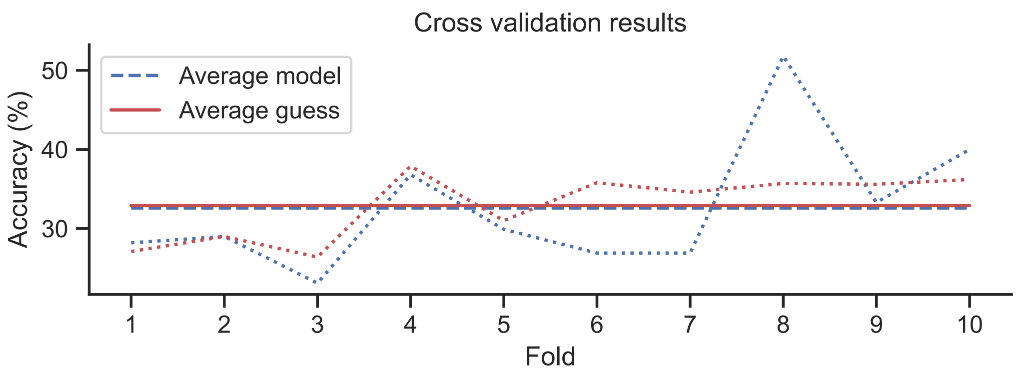

Figure F.14: Negative Binomial Cross validation results on projects labeled with project location NL3

Table F.14: Negative Binomial Cross validation results on projects labeled with project location NL3

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	28.2	29.0	23.1	36.8	29.9	26.9	26.9	51.8	33.3	40.0	32.6
Guess accuracy (%)	27.1	29.0	26.4	37.9	31.0	35.8	34.6	35.7	35.6	36.2	32.9

Only project with a local client

Table F.15: Initial regression results on only project with a local client

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	28.7	25.3	34.2	29.9	26.3	34.2
Best-guess accuracy (%)	26.0	27.3	34.2	26.0	27.3	34.2
Pearson χ^2	1532.4	NaN	NaN	724.2	NaN	NaN
Allowed Pearson χ^2	799.2	NaN	NaN	799.2	NaN	NaN
Deviance	1514.8	NaN	NaN	731.5	NaN	NaN

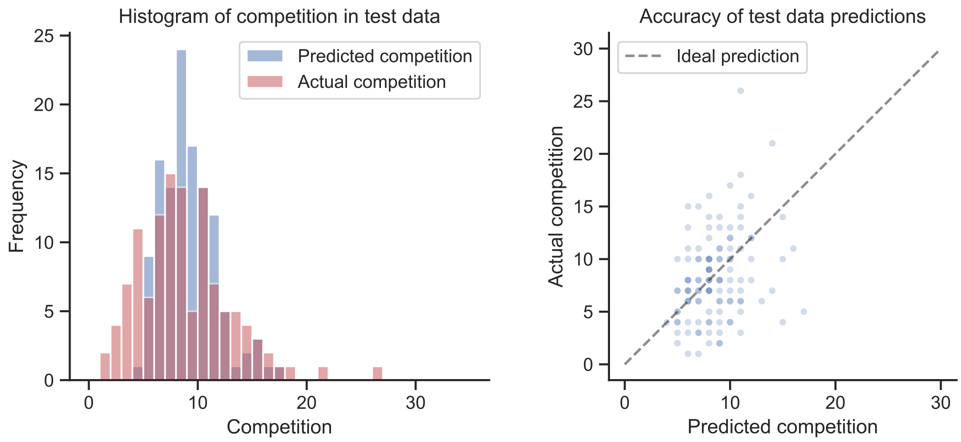

Figure F.15: Visualization of Negative Binomial predictions on the test data that contains only project with a local client

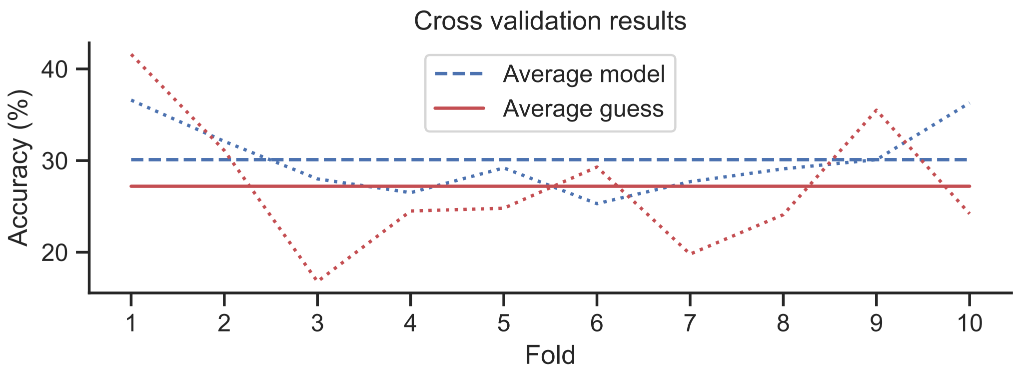

Figure F.16: Negative Binomial Cross validation results on project with a local client

Table F.16: Negative Binomial Cross validation results on project with a local client

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	36.6	32.1	28.0	26.5	29.2	25.3	27.7	29.1	30.1	36.3	30.1
Guess accuracy (%)	41.6	31.1	16.8	24.5	24.8	29.3	19.8	24.1	35.5	24.2	27.2

Only project with a national tender scope

Table F.17: Negative Binomial Cross validation results on project with a local client

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	31.3	32.8	33.8	31.5	31.2	33.1
Best-guess accuracy (%)	29.1	25.6	23.5	29.1	25.6	23.5
Pearson χ^2	1998.0	NaN	NaN	950.9	NaN	NaN
Allowed Pearson χ^2	1010.4	NaN	NaN	1010.4	NaN	NaN
Deviance	1903.3	NaN	NaN	907.7	NaN	NaN

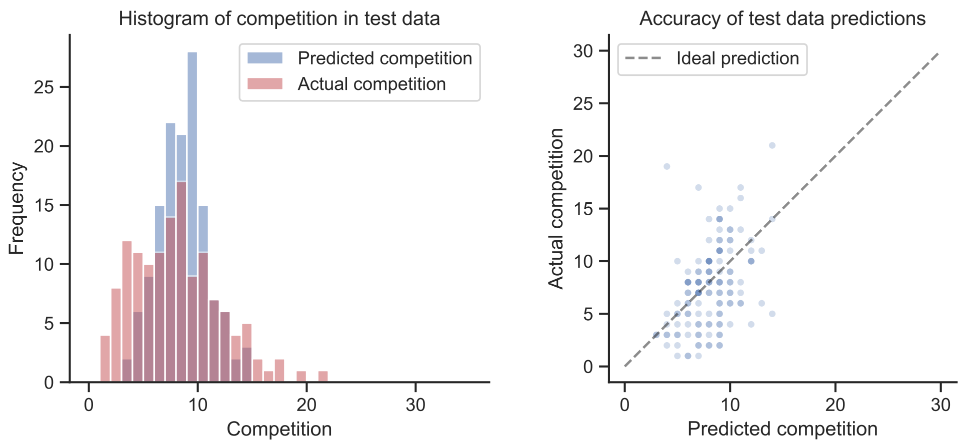

Figure F.17: Visualization of Negative Binomial predictions on the test data that contains only projects with a national tender scope

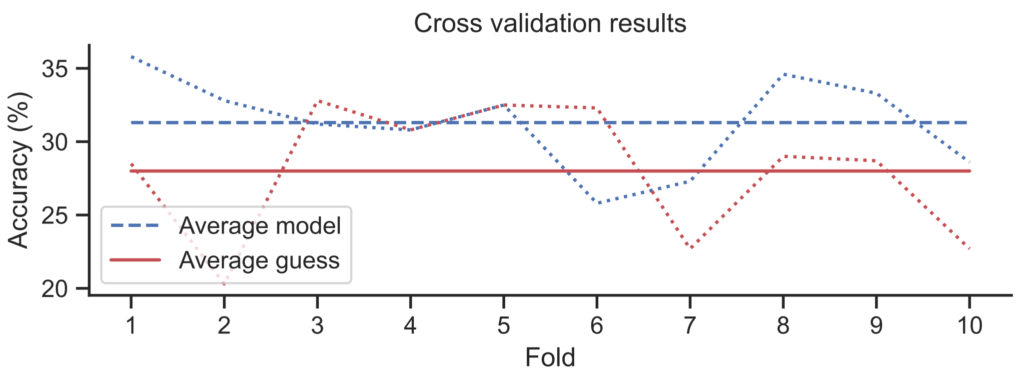

Figure F.18: Negative Binomial Cross validation results on projects that contain a national tender scope

Table F.18: Negative Binomial Cross validation results on projects that contain a national tender scope

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	35.8	32.8	31.2	30.8	32.5	25.8	27.3	34.6	33.3	28.6	31.3
Guess accuracy (%)	28.5	20.3	32.8	30.8	32.5	32.3	22.7	29.0	28.7	22.7	28.0

Only projects with a European tender scope

Table F.19: Initial regression results on projects with a European tender scope

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	41.0	36.7	34.4	40.2	35.7	33.3
Best-guess accuracy (%)	33.9	34.7	37.5	33.9	34.7	37.5
Pearson χ^2	908.1	NaN	NaN	699.8	NaN	NaN
Allowed Pearson χ^2	767.9	NaN	NaN	767.9	NaN	NaN
Deviance	911.1	NaN	NaN	704.7	NaN	NaN

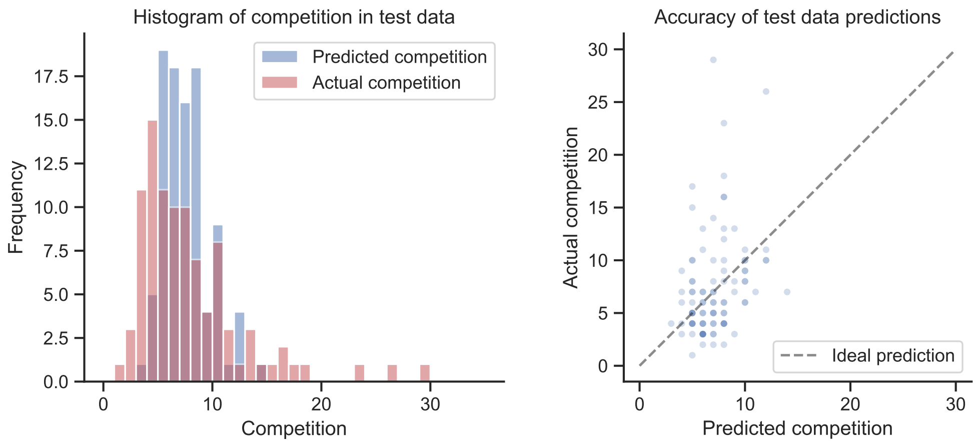

Figure F.19: Visualization of Negative Binomial predictions on the test data that contains only projects with a European tender scope

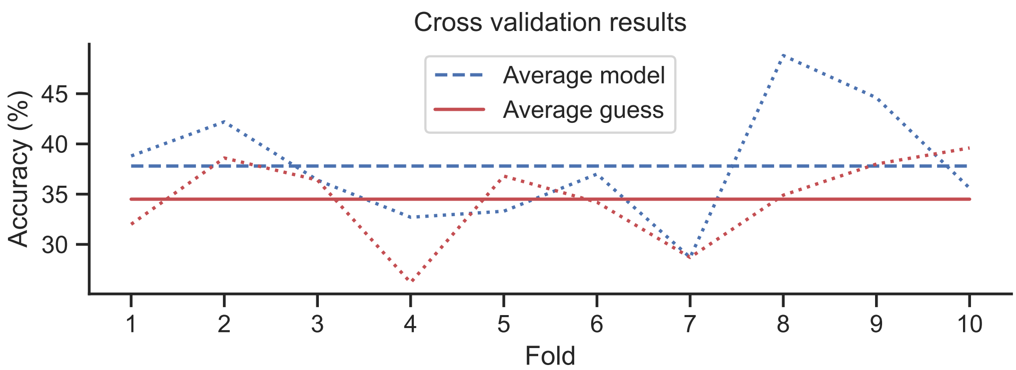

Figure F.20: Negative Binomial Cross validation results on projects with a European tender scope

Table F.20: Negative Binomial Cross validation results on projects with a European tender scope

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	38.8	42.2	36.4	32.7	33.3	37.0	28.7	48.8	44.6	35.6	37.8
Guess accuracy (%)	32.0	38.6	36.4	26.2	36.8	34.2	28.7	34.9	38.0	39.6	34.5

Only general type of work projects

Table F.21: Initial regression results on projects labeled with general type of work

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	32.9	26.0	29.8	32.6	27.1	30.9
Best-guess accuracy (%)	29.7	27.1	24.5	29.7	27.1	24.5
Pearson χ^2	1311.3	NaN	NaN	670.5	NaN	NaN
Allowed Pearson χ^2	744.9	NaN	NaN	744.9	NaN	NaN
Deviance	1290.1	NaN	NaN	668.0	NaN	NaN

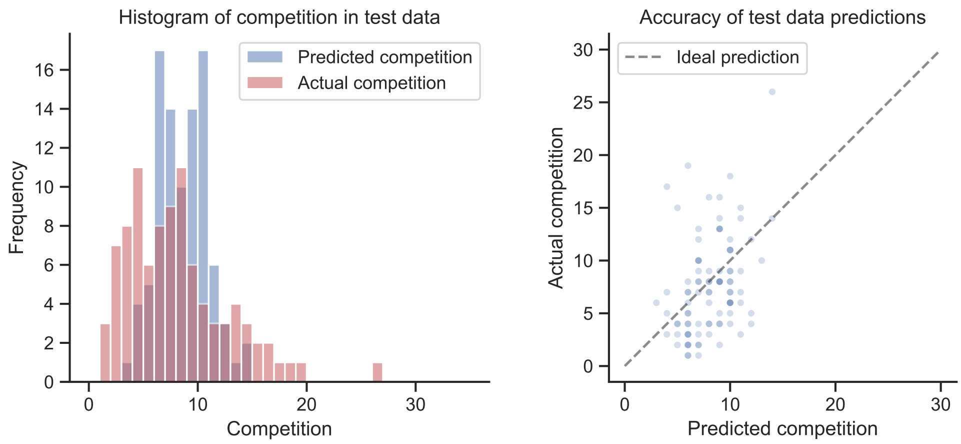

Figure F.21: Visualization of Negative Binomial predictions on the test data that contains only projects labeled with general type of work

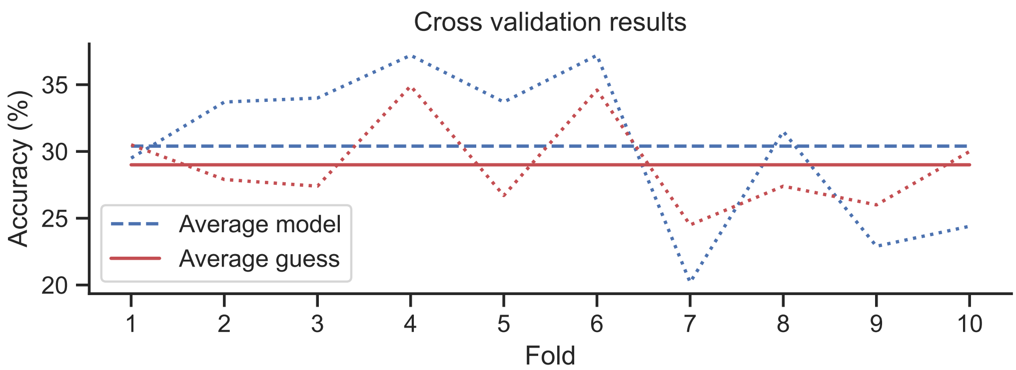

Figure F.22: Negative Binomial Cross validation results on projects labeled with general type of work

Table F.22: Negative Binomial Cross validation results on projects labeled with general type of work

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	29.5	33.7	34.0	37.2	33.7	37.2	20.2	31.5	22.9	24.4	30.4
Guess accuracy (%)	30.5	27.9	27.4	34.9	26.7	34.6	24.5	27.4	26.0	30.0	29.0

Only ground work projects

Table F.23: Initial regression results on projects labeled ground work

Index						
Model	Poisson	Poisson	Poisson	Negative Binomial	Negative Binomial	Negative Binomial
Input data	Training	Validation	Testing	Training	Validation	Testing
Model accuracy (%)	34.2	19.4	32.6	33.8	19.4	30.2
Best-guess accuracy (%)	30.6	19.4	30.2	30.6	19.4	30.2
Pearson χ^2	479.6	NaN	NaN	262.1	NaN	NaN
Allowed Pearson χ^2	283.6	NaN	NaN	283.6	NaN	NaN
Deviance	465.9	NaN	NaN	256.6	NaN	NaN

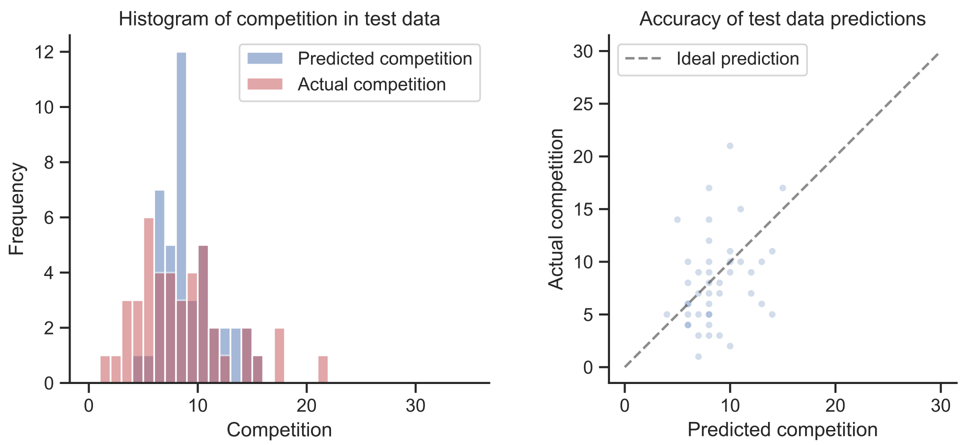

Figure F.23: Visualization of Negative Binomial predictions on the test data that contains only projects labeled ground work

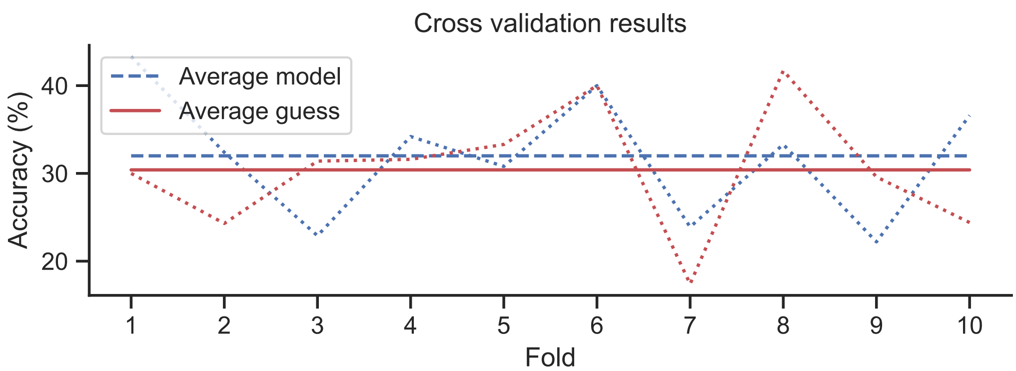

Figure F.24: Negative Binomial Cross validation results on projects labeled ground work

Table F.24: Negative Binomial Cross validation results on projects labeled ground work

	1	2	3	4	5	6	7	8	9	10	average
Model accuracy (%)	43.3	32.4	22.9	34.2	30.8	40.0	23.9	33.3	22.2	36.6	32.0
Guess accuracy (%)	30.0	24.3	31.4	31.6	33.3	40.0	17.4	41.7	29.6	24.4	30.4