

Development of Classification Model for the Level of Bid Price Volatility of Public Construction Project focused on Analysis of Pre-Bid Clarification Document

Y.E. Jang^a, J.S. Yi^a, J.W. Son^a, H.B. Kang^a, and J. Lee^a

^aDepartment of Architectural and Urban Systems Engineering, Ewha Womans University, South Korea
E-mail: jsyi@ewha.ac.kr

Abstract –

The purpose of this paper is to classify the level of formation of the bid price by using the type of uncertainty inherent in the bid document as a variable. To this end, the research examined the factors of the project related to the bid price presented in the previous study. Next, the pre-bid clarification document, which can be used to check the uncertainty of the bid documents, is used as a surrogate variable. Through these input variables, this research implemented two kinds of models using four algorithms: one predicts the level of bid price with uncertainty of bid document and the other predicts the level of bid price without uncertainty of bid documents. As a result, the model that predicts the level of the bid price reflecting the uncertainty of the bid document shows about 24 percent better performance than the model that predicts the bid price without reflecting the uncertainty of the bid document.

Keywords –

Risk Management, Bid Price Risk, Bid Price Volatility, Uncertainty of Bid Document, Pre-Bid Clarification, Bid Price Average, Bid Price Range, Machine Learning (ML), Classification Model, Public Construction Project

1 Introduction

Contracts for construction projects are made through competitive bidding. Bidding is a method of concluding a contract with a bidder who offers the most favourable content, that is, generally the lowest price, by letting multiple bidders submit applications with their price. Therefore, the most important issue for the bidder is how much the bid price should be presented [12]. This is because the bid price affects the likelihood that the bidder will be able to obtain a satisfactory profit from the project, as well as the likelihood that the bidder will be able to win the project.

Since the construction contract is concluded by contracting to implement the construction object through mutually agreed price based on the information given prior to the contract, which means bid document, the price required for the implementation of the construction object is also based on the bid document. Since the bid document can be operative as a contract document after the contract is concluded, bidders will thoroughly review the bid document from the bidding stage. In other words, the bid document serves as a substantive basis document for the bidders to calculate the bid price.

The bid document is a fundamental and essential communication tool between the client and the bidder. If the content of the bid document is uncertain, the intention of the construction object may become ambiguous and cause a mistake during construction phase [8], which may lead to construction rework, disputes and claims.

Understanding the uncertainty of the bid document can therefore enhance mutual communication among stakeholders and ultimately encourage the clients (i.e., owners) to improve the quality of the bid document [9]. In the United States, a system is in place to address such uncertainties in the bid document through a system that encourages bidders to inquire to their owners during the pre-bid clarification on the uncertainties inherent in the bid document and responds to them.

Bid price may be expressed in many ways, but generally it can be expressed as follows:

$$B_i = C_i(1 + M_i) \quad (1)$$

C_i means the total project cost of the project that bidder i expects, and M_i means the markups that bidder i internally reflect by analysing the project's uncertainty, or risk [1]. In other words, the result M_i reflecting the uncertainty in the bidding stage is included in the bid price B_i , which cause increase of total bid price. Conversely, if the uncertainty decreases, the bid price is determined at a lower level. In particular, a project that eliminates the uncertainty of bid document through the

pre-bid clarification procedure has been studied to form a bid price at a level more stable than the other projects [14].

Risk management is carried out throughout the lifecycle of the project, but initial preemptive risk management is crucial in that risk management at the bidding stages minimizes the damage that will occur at the later stages of the project [8]. Failure of risk analysis at the bidding stage is one of the root causes of project failure, which can cause significant damage to bidders and construction industry as well as project failure [15].

However, due to the many uncertain factors to be considered in predicting risk, it is difficult for bidders to offer bid prices that simultaneously satisfy satisfactory profits and the likelihood of winning a project [10]. In particular, there is a growing need to eliminate risks at the bidding stage of larger construction projects that involve relatively more uncertainty [18]. This uncertainty ultimately increases the price of the project [16], but due to the difficulty in figuring out the level of risk within a limited amount of time during the bidding process of the construction project, bidders have difficulties in determining the bid price as well as difficulty in deciding whether to participate in the bid [7].

For this reason, in practice, rather than predicting a reasonable level of risk in deciding whether to participate in a bidding for a construction project, it relies on subjective factors such as experience, speculation, and intuition. In other words, there is a need for a decision support tool that can be utilized by practitioners who have difficulties in accurate price predicting due to many factors affecting the bid price.

Therefore, this study considers uncertainty that can be observed in pre-bid clarification document as surrogate variable (i.e. proxy variable) in order to measure uncertainty in bid document and compares the level of performance of the bid price prediction between the model including that proxy variable and the model without it.

2 Theoretical Backgrounds

2.1 Risk Management of Construction Project

According to the PMBOK (Project Management Body Of Knowledge), risk management research can be said to consist of 1) Risk Identification, 2) Risk Assessment, and 3) Risk Plan and Control. Risk assessment can be defined as an estimate of the potential impact of an uncertain factor on a project based on an understanding of this uncertainty [4]. However, due to the difficulty of considering too many variables to be considered, and too much time to be

taken into account, risk assessment studies in practice as well as in risk assessment remain qualitative [2]. Although the most important prerequisite for quantitative analysis is to obtain real data [22], existing risk management studies have mainly focused on higher-level risk factors, and these studies have the problem that it is difficult to identify the results through actual data.

2.2 Factors Affecting Bid Price in Preliminary Studies

Construction projects can be divided into building, transportation, and various types of plant depending on the characteristics of the construction object. Among them, although construction and transportation are relatively different in size (e.g., construction (20.2%), transportation (25.6%)), the difference is not significant. In previous risk management research, rather than dividing building and transportation, there was a considerable case in which the risks of construction projects were covered in a comprehensive manner.

However, since the data analyzed by this study is the transportation project data, the risk factor list is extracted from the total of 5 studies including the study using the transportation project data in the preceding research. Seven variables were selected by matching whether the variables corresponded to the same or inclusive relation.

In this study, 5 prior risk management studies were used to extract risk factors for model input variable selection. Risk factors extracted from each study were presented in Table 1.

Table 1. Risk factors from literature review

No.	Risk Factors
1 (Chan and Au 2009)	Poor employer's reputation to honor payment on time; Amount of liquidated damages being higher than expected; Poor financial capability of the employer; Very tight contract period; Non-ideal project cash flow; Large portion of works subcontracted to nominated subcontractors; Low intensity of work; High degree of difficulty; Onerous contract conditions and rigid specifications; Possibility to have public objections
2 (El-Mashaleh 2012)	Project type; Project size; Quality of bid documents; Terms of payment; The contract includes an "adjustment for changes in cost" sub-clause; Cash flow requirements of the project; Availability of labor required for the project; Availability

	of materials required for the project
3 (Agnieszka and Edyta 2013)	Availability of equipment required for the project; Promote the reputation of the firm; Improve the experience of firm's personnel; Increase the possibility of building a long-term relationship with the client; Identity of the client; Financial capability of the client; Reputation of the client regarding his commitment of making timely payments; Influence of the client in making recommendations in the construction market; Reputation of the consultant regarding his independence in making "fair determinations" between the contracting parties; Amount of work currently at hand; Current financial standing of the firm; Availability of other projects in the market
4 (Dominic and Simon 2014)	Type of work; Past experience with similar projects; Contract documents; Owner's reputation; Value of the project; Need of work; Current involvement in other projects; Size of the project; Profits from similar past projects; Time of project duration; Criteria of bid selection
5 (Delaney 2018)	Location of the project; Time for the preparation of the bid; Possible subcontractors; Necessity for specialized equipment; Degree of complexity of works

The variables extracted from previous studies are as follows (Table 2).

Table 2. Factors affecting bid price in preliminary studies

No.	Factors
1	Number of bidders
2	Working days
3	Engineer's estimate
4	Project location
5	Bid preparation days
6	Project type

In Chapter 5, these variables are included in the training dataset in Model 1 and 2 so that both can predict the volatility level of the bid price.

2.3 Pre-Bid Clarification Document

Bidders are provided with bid document by which they obtain the desired construction object at the bidding stage. Bidders will carefully examine the bid document from the bid announcement date to the start date of the bid, and then determine the bid price. In this way, the bid document describes the criteria and procedures for the design, construction method, material, and quality inspection to complete the construction object. Since the bid documents themselves are contractually valid, bid documents may be used as a basis for making important judgments when claims and disputes arise in the future. If the risk factors included in the bid document cannot be reviewed in advance, it may become a factor of future project costs. Therefore, it is very important to analyse the uncertainty of bid documents in terms of risk management [14].

If the bidders believe that there is a problem due to uncertain contents of the bid documents during the pre-bid clarification process, they can ask the owner to review the contents. In other words, pre-bid clarification is a process in which a bidder inquires to the owner if the content of the bid document provided by the owner (e.g., construction contract, design document such as plans, special conditions of construction contract) is unclear [14]. The owner who receives the query from the bidder generally has an obligation to respond within a fixed period. If the query determines that there is a problem with the bid document, he or she will issue an amendment document, which is called addendum, instead of simply responding. Thus, the elimination of uncertainty in bid documents through pre-bid clarification can reduce the occurrence of abnormally low bidding prices and winning prices [19], the New York State Office of General Service, [17]) proposed the policy making process of inquiry and response process of Pre-bid clarification document in the bidding system of construction project as follows:

PRE-BID INQUIRY & RESPONSE POLICY

Background

The pre-bid inquiry and response process is important and beneficial for both the bidder and OGS. It helps to ensure more accurate contractor estimates and bids and fewer ambiguities. Providing consistent responses to prospective bidders also helps to avoid or minimize contract change orders, claims and disputes.

Therefore, considering that the risks presented in the pre-bid clarification documents address uncertainties arising from all kinds of bid documents, the pre-bid clarification document is used as a proxy for uncertainty of bid document in the bidding phase of the project. In the pre-bid clarification document, the following is a table summarizing the five common types of contents that affect the bid price. In Chapter 5, the variables that

can be extracted according to these types (Table 2) are constructed and reflected in Model 2 and compared with Model 1.

Table 3. Types affecting bid price in pre-bid clarification document

No.	Type and example
1	<p>BI 1. Discrepancy (Mutual contradiction)</p> <p>Example (Unsolved) Inquiry #34: <i>Comparing Layout drawing L-1 (pg 17) with pavement elevation plan C-1 (pg 37), the dimensions of the Alhambra Ave EB on-ramp (right) are not the same. Please clarify or re-issue correct drawings.</i> Response #2: <i>Please bid per current documents.</i></p>
2	<p>BI 2. Error</p> <p>Example (Solved) Inquiry #10: <i>Please check the State Quantity regarding Bid item 35 Rubber HMA – GAP?</i> Response #2: <i>Quantities have been verified. Bid per current contract documents.</i></p>
3	<p>BI 3. Omittance</p> <p>Example (Solved) Inquiry #4: <i>The Special Specifications require the installation of some plants and the irrigation. The irrigation plans shows two truck standpipe. It is safe to assume that there no water at this location either water must be truck or water must be developed.</i> Response #3: <i>Correct, the contractor is responsible for bringing water to the site. Water cannot be developed at the site, such as installing a well to capture underground water.</i></p>
4	<p>BI 4. Insufficient information</p> <p>Example (Unsolved) Inquiry #58: <i>?</i> Response #2: <i>Please bid per current contract documents.</i></p>
5	<p>BI 5. Alternative information</p> <p>Example (Unsolved) Inquiry #113: <i>We are requesting clarification on the requirements for Uniaxial Geogrid specified under Section 88–1.02R Uniaxial Geogrid. Specification requires a punched and drawn polypropylene uniaxial geogrid with minimum ultimate tensile strength of 7600 lb/ft. Review of available uniaxial geogrid products indicates no manufacturer supplies polypropylene uniaxial geogrid. We are requesting Caltrans to allow uniaxial geogrid with minimum ultimate tensile strength of 7600 lb/ft meeting the requirement of Section 88–1.02D(1) and geogrid shall have a regular and defined open area with an open area of 50 to 90 percent. Geogrid shall be manufactured from high tenacity polyester (PET) or high density polyethylene (HDPE).</i> Response #2: <i>Any proposal not specifically identified in the bid documents will only be considered by the Engineer after contract award. Please bid per the current contract documents.</i></p>

3 Machine Learning

Machine learning can be roughly divided into two

types, one is prediction method that derives regression equation as statistical analysis, and the other is classification method which determines the category of data. The bid price volatility level classification model that predicts the level of actual bid price after learning data by using machine learning needs a proper classification algorithm. In this study, four classification algorithms were used for the model: Decision Tree (Tree), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Neural Net (NN).

The classification algorithm using Tree algorithm is a methodology commonly used in machine learning. It aims at generating a model that predicts the level of an output variable based on several input variables. Tree generally has merit that it can be understood only by a brief description. In addition, compared with other techniques, there is no need to process the data, so the data pre-processing procedure is simple, and the processing speed is fast, so that it can be applied stably to a large dataset. However, the Tree model is very sensitive to the initial settings, which can lead to different results each time [13]. In addition to this, there is a disadvantage that it is difficult to solve a complicated classification problem such as XOR operation with the Tree algorithm.

SVM is one of a kind of machine learning and is a supervised learning model for pattern recognition and data analysis. It is mainly used for classification. Given a set of data belonging to either of the two levels, the SVM algorithm creates a non-stochastic binary linear classification model that determines to which level new data belongs based on a given set of data [20]. The SVM algorithm finds the boundary with the largest width. SVM can be used in nonlinear classification as well as linear classification. In order to perform nonlinear classification, it is necessary to map the given data to the high dimensional feature space. In order to this efficiently, a kernel trick is used. The learning effect of SVM is known to be very good among classification algorithms, and it is useful for classification of proteins in the field of medicine.

KNN is one of the representative classification algorithm used in machine learning. It categorizes data into a principle that weights neighbors' contributions so that the closer the neighbors contribute more to the average than the farther neighbors. For example, the most common weighting scheme is to give each neighbor a weight of $1/d$ when d is the distance to the neighbor. This principle allows KNN to classify data effectively, but it can be very sensitive to the local structure of the data. This is called 'Majority Voting', and this phenomenon occurs when the level distribution is biased. In other words, more frequent levels of data tend to dominate the prediction of new data because more frequent levels of data tend to be the majority of

the K Nearest Neighbors [11].

The NN algorithm, also known as the Artificial Neural Network, is a classification algorithm that can be useful for learning highly complex data. The main advantage of the NN algorithm is that it can learn from the observed data and produce the desired approximate function. In order to utilize this NN algorithm, it is necessary to set the connection pattern between the neurons of the other layers, the learning process of updating the weight of the connection, and the activation function of converting the weight input of the neuron into the activation output. The NN algorithm belongs to Deep Learning among the machine learning. It can solve complex problems such as XOR problem, and it is generally known that it has high performance in classification.

4 Data

In order to classify the bid price volatility level through the analysis of the bid data, this study selected the public construction project in transportation ordered by Caltrans, California, USA as analysis data.

There are various types of private standard contracts used worldwide, such as the International Federation of Consulting Engineers (FIDIC), Joint Contracts Tribunal (JCT), New Engineering Contract (NEC), and American Institute of Architects (AIA). In Caltrans' case, Federal-Aid Construction Contracts (FHWA-1273) is used as the standard contract terms when they carry out a general project. Unlike private projects, the contract conditions of these public projects are used without any modification of special conditions in most cases. This means that the construction project from Caltrans uses a bid document that does not deviate significantly from the standard. These standardized contracts and the bid documents that contain them will not only reduce the influence of numerous factors affecting the bid price, but also the nature of the bid documents, as described in the scope of this study, it is believed to be effective in examining the impact of uncertainty in the bid documents on the risk measurement indicators.

The California State Department of Transportation is a very large ordering organization in the United States, and it has about 450 civil engineering works each year based on its strong economic power. According to the State Transportation By The Numbers: A Compendium of State Summaries, a statistical survey conducted by the US Department of Transportation, the California State of California is a large administrative area that accounts for about 4.3 percent of the US territory. 281,617 km of roads and 8,521 km of trains, (US Department of Transportation 2015). The US State Department of Transportation is investing \$ 1.7 billion annually in infrastructure to reorganize aging facilities

and build new facilities, accounting for about 10 percent of the US Department of Transportation's overall budget, the largest figure in the nation's 50 states. Therefore, the California Department of Transportation has thousands of standardized construction project data that have been carried out to facilitate data collection and analysis. However, due to the long experience of ordering and project management knowledge, the uncertainty of bid documents of the California State Department of Transportation is estimated to be relatively small compared to other ordering organizations. If the uncertainty of these bid documents decreases, I think it may be a little difficult to see the relationship and its effectiveness.

In addition, the California State Department of Transportation's bidding process includes a Pre-bid clarification process that eliminates the uncertainty of the bid documents through the bid question and answer process. In addition to the Pre-bid clarification document generated during this process, and that the data are in accordance with the purpose and characteristics of this study.

5 Datamining: Modelling

There are two models to be implemented in this study:

- Model 1: Implemented with input variables NOT including uncertainty of bid document related
- Model 2: Implemented with input variables including uncertainty of bid document related

Therefore, this Chapter deals with specifying the input, output, and output variable classes required to implement these two models.

5.1 Input Variables

The input variables to be used in the model consisted of the variables related to the bid price derived from the preliminary studies and the variables related to the uncertainty of the bid document (Table 4).

Table 4. Model input variables

Model	Input variable
1	<ul style="list-style-type: none"> ▪ Number of bidders ▪ Working days ▪ Engineer's estimate ▪ Project location ▪ Bid preparation days ▪ Project type
2	<ul style="list-style-type: none"> ▪ Number of bidders ▪ Working days

- Engineer's estimate
- Project location
- Bid preparation days
- Project type
- Number of unsolved BI. 1
- Number of unsolved BI. 2
- Number of unsolved BI. 3
- Number of unsolved BI. 4
- Number of unsolved BI. 5
- Number of unsolved BI Total
- Solved ratio
- Number of addenda

5.2 Output Variable and Class Designation

The output variable of the model should be a risk measure of the bid price. In this study, the following two risk measures were used: Bid Average Risk and Bid Range Risk.

5.2.1 Bid Average Risk

Bid Average Risk is the ratio of the average price which can be considered the agreed bidding price to the project's base price (engineer's estimate), which measures the risk and can be expressed as follows:

$$\text{Bid Average Risk} = \frac{\text{Average Bid Price}}{\text{Engineer's Estimate}} \quad (2)$$

For example, if the bid price of project A and B are both \$10 billion and each average bid price of project A and B is \$10 billion and \$13 billion, the Bid Average Risk of Project A and B are 1.1 and 1.3 respectively. And for project B, bidders are expecting more risk.

5.2.2 Bid Range Risk

Bid Range Risk refers to the difference between the maximum bid price and minimum bid price as a result of the bid compared to the engineer's estimate, and it can be expressed as follows:

$$\text{Bid Range Risk} = \frac{\text{Max Bid Price} - \text{Min Bid Price}}{\text{Engineer's Estimate}} \quad (3)$$

Since the difference between the maximum and minimum bid price is also affected by the project size, the risk cannot be determined simply by measuring difference between the maximum and minimum bid prices. For example, in a project with \$10 billion and a project with \$1 billion of engineer's estimate, the difference between max and min bid price is all the same at \$2 billion, but the difference between the uncertainties of the two projects cannot be considered same.

Therefore, it is appropriate to compare the difference between max and min bid price against the project's base price, in this study, which is regarded as engineer's

estimate.

5.2.3 Output Variable and Class Designation

According to the above discussion, the output variables of Models 1 and 2 are shown in Table 5.

Table 5. Model output variables

Model	Output variable
1	<ul style="list-style-type: none"> ▪ Bid Average Risk ▪ Bid Range Risk
2	<ul style="list-style-type: none"> ▪ Bid Average Risk ▪ Bid Range Risk

The classification model predicts the class divided by a certain criterion rather than the value of these output variables themselves. For this purpose, the level of each output variable is specified and their classes are designated based on a certain criterion (Table 6).

Table 6. Class designation

Output variable	Class	Range	Ratio
Bid Average Risk	++	≥ 1.1	25%
	+	≥ 1.0 and < 1.1	27%
	-	≥ 0.9 and < 1.0	26%
	--	< 0.9	22%
Bid Range Risk	++++	≥ 0.27	25%
	+++	≥ 0.2 and < 0.27	23%
	++	≥ 0.13 and < 0.2	24%
	+	< 0.31	28%

6 Results

This Chapter discusses the implementation of the models in accordance with Chapter 5. The following table shows the accuracy of the algorithms and output variables of Models 1 and 2 (Table 7), where the accuracy is a predicted ratio of the total data, and the expression is as follows:

$$\text{Accuracy (\%)} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (4)$$

Table 7. Accuracy of classification models by algorithms

Algorithm	Accuracy (%)			
	Model 1		Model 2	
	Bid Average Risk	Bid Range Risk	Bid Average Risk	Bid Range Risk

Tree	31.5	34.3	61.7	60.6
SVM	34.6	40.5	59.5	60.5
KNN	32.8	36.1	58.4	56.1
NN	37.5	42.5	63.9	65.8

First, according to the Table 7, it was showed that the prediction accuracy of the model slightly differs according to the algorithm. The NN showed the highest accuracy in all cases, followed by SVM, KNN, and Tree, although with a few exceptions.

Second, the bigger difference is between Model 1 and Model 2. In Model 1, mean accuracy is somewhat low at 36.22%. In Model 2, which includes variables related to uncertainty of bid documents, mean accuracy is 60.81% as shown in Table 7.

Third, there was a difference in the accuracy depending on the output variables. Except for only two cases in Model 1 and Model 2, both models showed better performance in predicting Bid Range Risk than Bid Average Risk.

7 Conclusion

Although it is important to manage risk at the initial stage of the construction project, that is, at the bid stage, there are many cases where it fails to calculate the appropriate bid price because it is not done properly. In particular, the bid document is one of the important risk factors in the bid phase because it contains many uncertainties in the document due to the characteristics of the one - time construction project. However, since it is difficult to analyze all of the uncertain risk factors affecting the bid price in a short time, in business, it relies on the experience of practitioners or experts in predicting bid prices, and many studies have focused on qualitative solutions to risk management. Quantitative solutions are also based on statistical techniques based on virtual data or expert questionnaires rather than based on actual bid data, which has limitations in verification. In other words, there is a lack of empirical analysis of the effect between the uncertainty of the bid document and the actual bid price.

In order to meet both the need for decision support tools that can be used to calculate the final bid price in business and the academics' need for risk management research based on actual bid data, this research has developed a model to classify the bid price volatility level for the project bid data. The bid price volatility to be considered in this paper means the average of the bid price and the range of the bid price over the engineer's estimate. In order to reflect the uncertainty of bid documents, which were previously regarded as uncontrollable risk, the research sought to improve the accuracy of the classification model by using pre-bid

clarification document as a proxy variable.

First, variables that can be obtained from the bid data were extracted and variables with high contribution were chosen by analyzing 13 previous researches related to risk factors. After that, 14 variables were identified as input variables in the model through the numeralization or categorization of the variable data.

Pre-bid clarification document analysis showed that the project with high uncertainty of bid documents formed a relatively higher average of bid prices than those without, and the range of bid prices was relatively broad. This confirms that the uncertainty of the bid document affects the bid price.

In order to examine whether the information related to the uncertainty of the bid documents gives better results in terms of accuracy in classifying the level of volatility of bid prices, a classification model that includes 8 variables related to uncertainty of bid documents and a classification model that does not include those variables were developed and accuracy of both models were compared. As a result of the analysis, the model that reflects the uncertainty of the bid documents showed accuracy of 63.9% for the average bid price and 65.8% for the range in bid price volatility level classification, whereas the model that does not reflect the uncertainty of the bid documents showed accuracy of 37.5% for the average bid price and 42.5% for the range in bid price volatility level classification. Consequently, it is confirmed that the result is better when the information related to the uncertainty of the bid document is reflected.

This study ultimately has the following contributions. In the bidder's view, the bid price should be calculated to increase the likelihood of winning the project and to guarantee the profit of the project within the limited bid preparation time. The bidder will have the opportunity to fix or strategically change their bid price by comparing their bid price with the bid price average and range of the current project predicted through the cases of the past projects. In other words, the results of this study can be used as a more reasonable decision support tool that bidders can use when determining the bid price based on actual bid data. In addition, from the view of the owner, it can be expected that unnecessary design change and increase of project cost will be reduced by contracting with the bid price which fully reflects the result of the risk analysis including uncertainty of the bid price. Moreover, the bid document is a document that the owner provides to the bidder. The result of this study that the uncertainty of the bid document affects the bid price suggests the positive effect of providing bid document with high quality.

Although the actual bid price is calculated on the basis of the actual bid documents containing the contents for the purpose of construction, there has been

a limit to the difficulty in analyzing the actual bid data as the risk management research conducted based on the higher level of qualitative risk factors. This study has significance in that variables were obtained from the actual bid data, and implies that the level of risk management research has deepened to the detail level based on the actual bid documents by reflecting uncertainty of bid document through pre-bid clarification document as a proxy variable, consequently providing the foundation of research that can make more effective risk management risk studies.

8 Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1A09083708).

Also, this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. NRF-2016R1A2B4015977).

9 References

- [1] Abotaleb, I. S., & El-adaway, I. H. (2016). Construction bidding markup estimation using a decision theory approach. *Journal of Construction Engineering and Management*, 143(1), 04016079.
- [2] Ahmad, I., & Minkarah, I. (1988). Questionnaire survey on bidding in construction. *Journal of Management in Engineering*, 4(3), 229-243.
- [3] Chan, E. H., & Au, M. C. (2009). Factors influencing building contractors' pricing for time-related risks in tenders. *Journal of Construction Engineering and Management*, 135(3), 135-145.
- [4] Chua, D. K. H., & Li, D. (2000). Key factors in bid reasoning model. *Journal of Construction Engineering and Management*, 126(5), 349-357.
- [5] Chua, D. K. H., Li, D. Z., & Chan, W. T. (2001). Case-based reasoning approach in bid decision making. *Journal of construction engineering and management*, 127(1), 35-45.
- [6] Delaney, J. W. (2018). *The Effect of Competition on Bid Quality and Final Results on State DOT Projects* (Doctoral dissertation, State University of New York at Buffalo).
- [7] Dikmen, I., Budayan, C., Talat Birgonul, M., & Hayat, E. (2018). Effects of Risk Attitude and Controllability Assumption on Risk Ratings: Observational Study on International Construction Project Risk Assessment. *Journal of Management in Engineering*, 34(6), 04018037.
- [8] Doloi, H. (2012). Cost overruns and failure in project management: Understanding the roles of key stakeholders in construction projects. *Journal of construction engineering and management*, 139(3), 267-279.
- [9] Duzkale, A. K., & Lucko, G. (2016). Exposing uncertainty in bid preparation of steel construction cost estimating: II. Comparative analysis and quantitative CIVIL classification. *Journal of Construction Engineering and Management*, 142(10), 04016050.
- [10] El-Mashaleh, M. S. (2012). Empirical framework for making the bid/no-bid decision. *Journal of Management in Engineering*, 29(3), 200-205.
- [11] Kashyap, D., & Misra, A. K. (2013, July). Software development cost estimation using similarity difference between software attributes. *In Proceedings of the 2013 International Conference on Information Systems and Design of Communication* (pp. 1-6). ACM.
- [12] KPMG International. (2015). *Global Construction Survey 2015*. Climbing the curve.
- [13] Laurent, H., & Rivest, R. L. (1976). Constructing optimal binary decision trees is NP-complete. *Information processing letters*, 5(1), 15-17.
- [14] Lee, J., & Yi, J. S. (2017). Predicting Project's Uncertainty Risk in the Bidding Process by Integrating Unstructured Text Data and Structured Numerical Data Using Text Mining. *Applied Sciences*, 7(11), 1141.
- [15] Maemura, Y., Kim, E., & Ozawa, K. (2018). Root Causes of Recurring Contractual Conflicts in International Construction Projects: Five Case Studies from Vietnam. *Journal of Construction Engineering and Management*, 144(8), 05018008.
- [16] Miller, R., & Lessard, D. R. (2008). *Evolving strategy: risk management and the shaping of mega-projects*. Cheltenham: Edward Elgar.
- [17] OGS. (2018, December 3). *Pre-bid inquiry & response policy*. Retrieved from <https://online.ogs.ny.gov/dnc/contractorconsultant/esb/prebidinquiryresponsepolicy.asp>
- [18] Priemus, H., Flyvbjerg, B., & Van Wee, B. (2008). *Decision-making on mega-projects*. Edward Elgar Publishing.
- [19] Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of Construction Engineering and Management*, 129(2), 198-204.
- [20] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- [21] Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining,

numerical data and ensemble classifiers.
Automation in Construction, 43, 23-29.

- [22] Winch, G. M. (2010). *Managing construction projects*. John Wiley & Son.