

# A Strategy Bayesian Model to Predict Profit of Construction Projects

Sunghyuk Park and Sangyong Kim

MSc. Student, School of Construction Management and Engineering, University of Reading, Reading, United Kingdom

PhD. Candidate, School of Construction Management and Engineering, University of Reading, Reading, United Kingdom

<http://dx.doi.org/10.5659/AIKAR.2011.13.3.49>

**Abstract** Competitive bidding in construction is concerned with contractors making strategic decisions in respect of determination of bid price if contractors opt to bid. This study presents a strategy model for deciding optimum tender price with reflecting appropriate profit in competitive bidding using Bayesian regression analysis (BRA). The purpose of the developed model is to help contractors to secure suitable profitability by predicting the actual profit based on key variables. They may affect construction cost at bidding phase, ultimately which help contractors to secure high quality output. The model was tested empirically by application to a bidding dataset collected from a large South Korea contractor. BRA allows contractors to estimate more accurate actual profit by reflecting not only objective information but also subjective experiences and judgments. Consequently, the model can contribute to improvement of decision-making process for setting an optimum tender price.

*Keywords: Optimum Tender Price, Competitive Bidding, Bayesian Regression*

## 1. INTRODUCTION

Profit maximization has been the single most important business objective (Nellis and Parker, 1992). In the competitive bidding, however, the lowest price is the key criterion in selecting contractors (CIOB, 2009), which does not allow the contractors to have the tender price to reflect sufficient profit and risk contingencies. Furthermore, the competition has often made the occurrence that the contractors win the bids at an abnormal tender price (Gorgan, 1992). It is serious problems such as poor quality, delay or claims against clients to compensate the loss of profit (Lo et al, 2007). Egan (1998) contends that the construction industry must move away from the lowest price to the best value. Despite such criticism on the competitive bidding, it is still the favoured route that clients meet their contractors. Thus, the contractors have to calculate the optimum tender price not only to win the bid but also to achieve appropriate profitability. It is possible by improving the accuracy

of cost estimating through assessment of various factors affecting construction cost in bidding phase.

Until now, many studies have developed to more efficiently and accurately forecast construction cost. Cost models introduced and have developed to achieve the goal. However, most of the models were developed by academic researchers, and there is no any reference on what contractors actually use them in real business (Laryea and Hughes, 2009). The main reasons are the lack of experts to be able to deal with the model and the lack of data for analysis and distrust on the accuracy of the models by senior management (Christodoulou, 1998). Also, the models has been quantitative so that the bid decision process can be made clear and objective, rather than subjective (Kim and Reinschmidt, 2010). However, in real situations, subjective assessment is the most commonly used by contractors (Shash, 1995; Mochtar and Arditi, 2001). Given the factors, it is necessary to develop practical and reliable strategy model to help contractor's decision making in bidding phase. It should be able to reflect the subjective assessment for dealing with both quantitative and qualitative factors affecting construction cost.

Bayesian statistics is an appropriate technique to deal with both the factors so that it is possible to reflect prior probability. The concept of Bayesian approach to statistics is to use the prior information (subjective experience or judgment) and the information contained in the data (objective factors) (Bolstad, 2007). Given the feature of the Bayesian statistics, it can be effectively used to develop a strategy model. It should be handled by subjective experiences and knowledge.

This study developed a strategy model to help contractors to decide optimum tender price considering appropriate profit

Corresponding Author: Sangyong Kim, PhD. Candidate

School of Construction Management and Engineering,  
University of Reading P.O. Box 219, Whiteknights, Reading,  
Reading, United Kingdom

Tel :+44 118 378 7182 e-mail: rd026992@reading.ac.uk

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

using Bayesian regression analysis (BRA). Thus, the model helps the contractors to adjust their estimated tender price in the adjudication stage of cost estimating, ultimately which contributes to achieving the business objectives (appropriate profitability, quality, completion on time and budget) of both the clients and contractors.

**2. RESEARCH BACKGROUND**

The purpose of frequentist statistics has been obtained by disregarding any prior knowledge about the process being measured. Yet disregarding the prior knowledge is significantly inefficient so that the useful information measured by plentiful experiences and various researches exists in many areas. In this aspect the benefit of BRA is to use both sources of information: prior information about the process and information about the process contained in the data, which means that if the reliable prior knowledge is secured. This method can be relatively more relevant especially when the amount of sample lacks (Bolstad, 2007).

BRA is the similar to traditional regression analysis in the basic notion that it analyzes the relationship,  $y=a+\beta x$ , between one dependent variable and one or more independent variables. Bayesian regression is much easier if the model is reparameterized to be “ $y=a_x+\beta(x-\bar{x})$ ” (Bolstad, 2007). The joint likelihood of the sample factors into a part dependent on the slope  $\beta$  and a part dependant on  $a_x$  (Bolstad, 2007). Independent priors are used for the slope  $\beta$  and intercept  $a_x$ . They can be either normal priors or “flat” priors, and the joint prior is the product of the two priors. The joint posterior is proportional to the joint prior times the joint likelihood. Since both the joint prior and joint likelihood factor into a part dependent on the slope  $\beta$  and a part dependant on  $a_x$ , the joint posterior is the product of the two individual posteriors. Each of them is normal where the constants can be found from the simple updating rules.

The parameters of posterior distribution calculated by BRA are changed by joint likelihood function and prior distribution (see Figure 1). However, if the prior knowledge is incorrect, and the difference of likelihood and priors is significant, it will be irrelevant to judge the reliance of posterior.

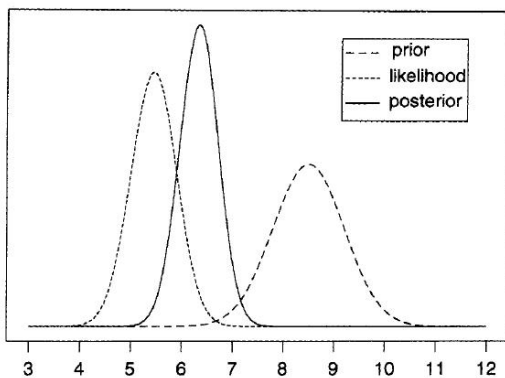


Figure 1. Prior, likelihood and posterior (Bolstad, 2007)

Bolstad (2007) introduced an indicator random variable. This study gives a small prior probability of indicating original prior that

is misspecified. The general mixture prior is  $P(I = 0)g_0(\theta) + P(I = 1)g_1(\theta)$ , where  $g_0$  and  $g_1$  are the original prior and a more widely spread prior, respectively. The joint posterior of distribution of  $I$  and  $\theta$  is found given the data. The marginal posterior distribution of  $\theta$ , given the data, is found by marginalizing the indicator variable out. It will be the mixture distribution:

$$g_{mixture}(\theta|y_1, \dots, y_n) = P(I=1|y_1, \dots, y_n)g_0(\theta|y_1, \dots, y_n) + P(I=0|y_1, \dots, y_n)g_1(\theta|y_1, \dots, y_n) \tag{1}$$

This posterior, Eq. (1), is very robust against a misspecified prior. If the original prior is correct, the mixture posterior will be very similar to the original posterior. However, if the original prior is very far from the likelihood, the posterior probability,  $P(I=0|y_1, \dots, y_n)$ , will be very small, and the mixture posterior will be close to the likelihood like Figure 2. This has resolved the conflict between the original prior and the likelihood by giving much more weight to the likelihood (Bolstad, 2007).

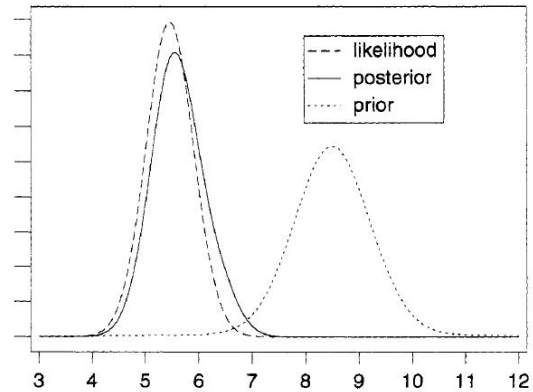


Figure 2. Mixture posterior and the two components (Bolstad, 2007)

**3. METHODOLOGY**

The two-stage methodology was adopted to achieve the aim of this study. First, data collection and analysis were performed for finding the most appropriate factors that affect the profitability of a project. Firstly, the collected data were analyzed by multiple regression analysis (MRA) to secure coefficients of the data variables versus a dependent variable. And then, the results were evaluated by expert survey and the reliance of sample size and characteristics in order to abstract the more important factors affecting the profitability.

Second, simulation was performed for developing models to predict the profit of a project with the data abstracted from previous stage. In the first step, traditional regression model was developed as a benchmark against which a Bayesian regression model could be measured. In the next, a survey targeting six professionals in construction industry was conducted to find informative priors for Bayesian regression model. Lastly, the Bayesian regression model was developed by using the informative priors, which was compared to the result of the traditional regression model to confirm the reliance of the Bayesian model.

### 4. DATA COLLECTION AND ANALYSIS

#### (1) Data collection

The survey questionnaire was designed to enable respondents to add any further factors that they considered necessary for inclusion to the list of related factors. The review of related researches supplies a list of factors that is further re-examined in expert interviews. They are given the task of adding to the prepared list or crossing off factors from the list that is irrelevant from their perspective. The confidence of data was come by step by step selecting procedure as shown in Figure 3.

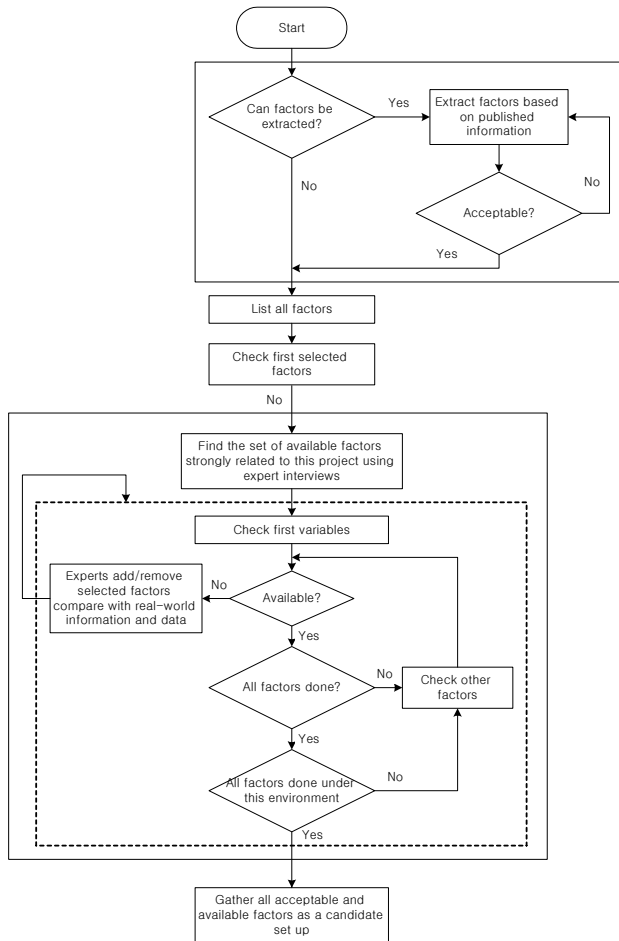


Figure 3. Selecting procedure of factors

#### (2) Analyzing coefficients of factors affecting the profit

MRA was used for selecting the most reliable factors affecting the profits of a project. The analysis is one of the useful methods not only to look at the relationship between a dependent variable and more than two independent variables but also to examine whether there are underlying factors that could explain combinations of variables that vary together by considering correlation matrix of pairwise correlations (Hinton et al., 2004). The independent variables are completed year, location, number of basement, number of story, type of building, gross floor area, duration, building cover ratio, floor area ratio, numbers of parking, type of structure, retaining wall, foundation and finished materials (floor, ceiling, internal and external wall). The goal of the MRA is to create

a model that contains all of the independent variables that will the best and most efficiently predict the dependent variable. To create the best results, five methods in the MRA are used, such as 'enter', 'stepwise', 'remove', 'backward' and 'forward'. This study analyzed every method to look at which method is the most efficient and found that 'backward' method is the most appropriate analysis procedure. The R2 and adjusted R2 value in the Table 1 indicate the amount of variance in the dependent variable. In most cases R2 is usually adopted, but taking the adjusted R2 value can be useful when a number of variables exist. Although the R2 value of 'backward' method was slightly lower than that of 'Enter' and 'Remove' methods at 0.769 and 0.794 respectively, the adjusted R2 of 'backward' is significantly higher than others at 0.678 and 0.560 each. ANOVA value indicates how important the regression models are. If F value is high and P value is less than 0.05, the model can explain the significant amount of variation in dependent variable. The ANOVA value of 'backward' model explained the most significant amount of variation in the dependent variable, with the results estimated at  $F(9,23)=8.491$ ;  $P<0.01$ . Furthermore, 'backward' model estimated more number of independent variables having positive coefficient than other models. Therefore, this study adopted 'backward' method for MRA.

Table 1. Comparison five analysis methods

Item	Enter	Stepwise	Remove	Backward	Forward
R <sup>2</sup>	.794	.446	.794	.769	.446
Ad. R <sup>2</sup>	.560	.338	.560	.678	.388
ANOVA	F(17.15) =3.398; P<0.05	F(3.29) =7.771; P<0.01	F(17.15) =3.398; P<0.05	F(9.23) =8.491; P<0.01	F(3.29) =7.771; P<0.01
Number variable	3	3	3	7	3

Table 2 gives nine models produced by MRA using a 'backward' method. Given R2 value in the table, independent variables in every model accounted for over 76.9% of variance in the dependent variables, and the deviation between highest and lowest value was 4.9%. Adjusted R2 value also showed the similar figures. Compared to the two values, model nine with the highest adjusted R2 value and relatively high R2 value was judged to the most reliable model. The model nine can consider more factors affecting a dependent variable.

Table 2. Comparison of nine models created by 'backward' MRA

Model	R <sup>2</sup>	Ad. R <sup>2</sup>	F	Sig.	Number variables
1	.794	.560	3.398	.011	17
2	.794	.588	3.850	.005	16
3	.793	.610	4.335	.002	15
4	.788	.623	4.784	.001	14
5	.785	.638	5.348	.001	13
6	.782	.651	5.970	.000	12
7	.778	.661	6.676	.000	11
8	.774	.671	7.526	.000	10
9	.769	.678	8.491	.000	9

According to Table 3, independent variables simultaneously abstracted from model nine were year, location, building coverage ratio, floor area ratio, type of structure, floor finished materials, internal wall finished, number of basement and duration, were contained in model nine. P-values were also bigger than 0.05. The figures of B column give the coefficients of the independent variables versus dependent variable. However, it is impossible to compare the coefficient of each independent variable with the figures of B column since the scales of the figures are different. Beta value is the standardized coefficient values for solving the problem. Of these independent variables, the most significant variable affecting dependent variable was floor area ratio that has the largest absolute Beta value, and next was internal wall finished.

Table 3. Coefficient of model nine

Model	B	Beta	t	Sig.
(Constant)	1265.2	-	2.885	.008
Year	-.629	-.312	-2.875	.009
Location	1.565	.576	3.882	.001
N. of basement	.509	.237	1.287	.211
Contract duration	.043	.192	1.690	.105
9 Building coverage ratio	.001	.388	2.759	.011
Floor area ratio	-.020	-.799	-5.201	.000
Type of structure	3.230	.393	2.386	.026
Floor finished	-1.756	-.439	-2.707	.013
Internal wall finished	-2.603	-.608	-4.152	.000

### (3) Evaluating the main factors

The nine main factors affecting the profit of a project were calculated from MRA. However, it is necessary to evaluate the factors with experiences and information of professionals and sample size of data to confirm real values. From the reason, of the main independent variables calculated from MRA, some were deleted or modified, and some were added.

The independent variable, completed year, had significantly high p-value (0.009) in the model nine, but the fundamental cause that it affected profit margin is fluctuation price rather than the completed year itself. Moreover, Jaggar (2002) said that constructors can expect to add more profit margin in times of a boom as there will be a surplus of contracts and not enough constructors around to carry them out and vice versa. This means that tender price is intentionally adjusted by construction market condition. Given the substantive causes that the completed year influences the profitability, it is more reasonable to use the fluctuation price versus construction period and the economic condition as independent variables rather than completed year of the project. However, only fluctuation price was used as a independent variable instead the completed year so that it is difficult to judge whether the bidders of projects collected in this study applied the condition or not. To use the fluctuation price, construction price indices made by KICT (Korea Institute of Construction Technology) was utilize. This is construction statistics produced for estimating fluctuation price on the inputs - materials, labour and equipments - related to construction direct cost, based on the price of specific time.

Two factors, finished materials of floor and internal wall, were excluded from the independent variables for developing models

due to the lack of reliance in terms of the size and type of the samples. Table 4 showed the type and number of the sample in floor and internal wall. However, the number of vinyl and water paint in each independent variable accounted for most of the samples. This means that it is difficult to interpret that the independent variables sufficiently explained the relationship with the dependent variable.

Table 4. The type and number of samples in floor and internal wall

Floor		Internal Wall	
Type	Number	Type	Number
Vinyl	29	Water Paint	31
OA floor	1	Gypsum	2
Timber	1	-	-
Tile	1	-	-
Laminate	1	-	-
Total Sum	33	Total Sum	33

Type of building, gross floor area and contract sum were added to the independent variables as they are the basic factors in determining the tender price of a project. On the other hand, retaining wall and number of story were also added due to the high risk that the factors can result in the increase of construction cost at construction phase. Especially Elhag et al. (2005) ranked 67 factors which affect the cost of construction projects according to their influence and significance regarding the tender cost. The number of stories and level of uncertainty of soil conditions were ranked 27 and 14 respectively. Consequently, the main factors affecting the project profit are like Table 5.

Table 5. Independent variables index

Variable name	unit	N.B.
Duration	month	
Fluctuation	%	
Location	text	Urban, Semi-Urban, Rural
N. of basement	number	
N. of story	number	
Type of building	text	Residence, Office
Gross floor area	m <sup>2</sup>	
Building coverage ratio	%	
Floor area ratio	%	
Retaining wall	text	Slurry wall, H-Pile+E/A, CIP
Type of structure	text	RC, SPC
Contract sum	won	

## 5. SIMULATION

The purpose of this study is to create the models to estimate contract sum including appropriate profit using BRA. The main difference between traditional MRA and BRA is whether informative prior is used or not. The tender price estimating should be accurate since it directly affects the success of a

project. Moreover, as the construction projects have significant uncertainties, In this situation reflecting the individual experiences can be useful in predicting more accurate tender price. In this aspect, this study applied BRA and traditional MRA to confirm the reliability of the Bayesian model including reflection of the subjective views, and used a survey to secure proper informative priors.

**(1) Data set**

The data of previous study (see Table 5) was used as independent variables, and profit was also used to a dependent variable for developing the models. The data collected from 33 completed projects in South Korea were partitioned into two parts, training set and holdout set in order to avoid bias. After a certain prediction model is completed, it is significantly important to confirm the reliance of the model. To check the model, the method to substitute testing data is generally used. Hair et al.(1992) describe such a procedure for use in dicriminant analysis, Hecht-Nielsen (1990) likewise addresses the use of the method to test the model of neural networks, and Seydel (2003) also uses partitioning the data into two parts. From these reasons, the data in this work is partitioned into training set and holdout set. 29 projects are the training data to create cost models, and the rest 4 projects are the holdout data to test reliance of the models. Furthermore, this study selected the projects with the different arrange of dependent and independent variables as the holdout set in order to test the model (see Figure 6).

**(2) Traditional regression-based model**

This regression-based model was calculated by SPSS 18. The several values of the regression model were shown in Table 6. R value represents correlation coefficient between all entered independent variables and a dependent variable, and R2 value shows the amount of variance in the dependent variable that can be explained by the independent variables (Hinton et al, 2004).

Table 6. Summary of regression-based model

Model	R	R <sup>2</sup>	SD	F	Sig.
1	.802 <sup>a</sup>	.642	2.45	2.769	.029 <sup>a</sup>

Table 7. Coefficient of independent variables

Variable name	B	SD	β	t	Sig.
Intercept	1.432	6.661		.215	.833
Duration	.014	.069	.064	.201	.843
Fluctuation	.204	.216	.304	.945	.358
Location	.973	1.400	.201	.696	.497
N. of basement	-1.503	.993	-.710	-1.513	.150
N. of story	.326	.200	.358	1.627	.123
Type of building	1.815	2.170	.247	.836	.415
Gross floor area	5.45E-05	.000	-.539	-2.085	.053
Building coverage ratio	.070	.103	.151	.681	.506
Floor area ratio	-.009	.005	-.397	-1.849	.083
Retaining wall	-2.287	1.440	-.286	-1.589	.132
Type of structure	5.385	2.751	.678	1.957	.068
Contract sum	-1.983E-06	.000	-.417	-1.623	.124

a. Dependent variable : profit rate

In the result of coefficient, Table 7, duration, location, type of building, building cover ratio give relatively low effect in influencing the profit. However, most of these factors are the fundamental elements in calculating a project cost. Furthermore, location and duration is qualitative factor that should be judged by subjective experiences and knowledge so that the two factors contain complicate risks. For example, according to location, the project cost can increase due to delivery cost, infrastructure condition, claims and so on. From these reasons, these factors were included for calculating this regression model.

**(3) The survey for abstracting informative priors**

Reliable prior probability is significantly important in improving the value of BRA model. Thus, the main purpose of this survey is to secure high-quality informative priors. The targeting of survey is six experts whose everyday job is working in construction sector. Given that the respondents of this survey should judge the impact that each independent variables affect construction cost and profit. Especially, most of independent variables have a possibility to be able to influence the construction cost and profit by a variety of uncertainties while projects are progressed. In this aspect, not only estimators but also such non-estimators as construction manager and site manger were selected to the respondents of the survey. The survey used e-mail to communicate with respondents being in South Korea.

Respondents were required to give score on how important 11 factors listed in the second part of questionnaire are in influencing final project cost. The sheet like Figure 4 was used to score importance of each factor, and the factors were measured by scale of 0 to 10.

Factor	Impact Factor (0=Low, 10=High)										
	0	1	2	3	4	5	6	7	8	9	10
Duration											0
Fluctuation	0										
Location						0					
Number of Story											
Number of Basement						0					
Type of building											0
Type of structure										0	
GFA						0					
BCA										0	
FAR				0							
Retaining Wall											0
Contract Sum										0	

Figure 4. Sheet for scoring importance of factors

The result of the questionnaire is shown in Table 8. The impact score of each factor was rescaled. The factors that respondents recognized the most important in affecting final project cost were ‘contract sum’, ‘type of building’, ‘type of structure’ and ‘number of basement’, whereas ‘building coverage ratio’ was the lowest impact factor.

Singularly, not only did expert 1 and 5 not score the impact of duration but the value of σ is also significantly high, which resulted from characteristics of duration. As mentioned in the first part, although duration is one of the most important factors affecting the profit, it is not meaningful to consider. Thus, it is likely that the duration factor should be considered comparing contract duration and actual duration along with main reasons to make the extension of time.

Table 8. Results of importance of factors

Item	1	2	3	4	5	6	Mean	$\sigma$
Duration	-	0.6	0.8	0.5	-	1	0.1750	0.7092
Fluctuation	0.4	0.2	0.8	0	0.2	0.6	0.1000	0.2360
Location	0	0.6	0	0.8	1	0.2	0.2333	0.3427
N. of basement	0.8	0.2	0	0.2	0.6	1	0.4667	0.1547
N. of story	0.2	0.6	0	0.2	0.6	1	0.3667	0.1987
Type of building	0.4	1	0.8	0.8	0.8	1	0.8000	0.0480
Gross floor area	0.2	1	0.8	0.4	0.8	0.2	0.5000	0.2040
Building coverage ratio	0.8	0.6	0	0.2	0.4	0.8	0.2667	0.2827
Floor area ratio	0.6	0.4	0.6	0.6	0.4	0.8	0.1083	0.4064
Retaining wall	0.6	0	0.4	0	0.2	0.4	0.1333	0.1227
Type of structure	1	0.8	1	0.4	0.6	0.6	0.7333	0.0587
Contract sum	1	0.8	0.6	0.8	0.8	1	0.8333	0.0227

The values of mean and  $\sigma$  calculated by results of a survey were used as informative priors for BRA, and the informative priors were mean and variance in Table 8.

**(4) Bayesian regression-based model**

It is necessary to tell WinBUGS what likelihood distribution to use and what prior distribution to use in order to create Bayesian regression models. WinBUGS requires particular program code like Figure 5 that it can recognize. Likelihood distribution using 12 independents variables and prior distribution of the variables calculated from survey were described in Figure 5.

```

model
{# Standardise x's and coefficients
for (j in 1 : p) { b[j] <- beta[j] / sd(x[ , j])
for (i in 1 : N) { z[i, j] <- (x[i, j] - mean(x[, j])) / sd(x[, j]) }
b0 <- beta0 - b[1] * mean(x[, 1]) - b[2] * mean(x[, 2]) - b[3] *
mean(x[, 3]) - b[4] * mean(x[, 4]) - b[5] * mean(x[, 5]) - b[6] *
mean(x[, 6]) - b[7] * mean(x[, 7]) - b[8] * mean(x[, 8]) - b[9] *
mean(x[, 9]) - b[10] * mean(x[, 10]) - b[11] * mean(x[, 11]) - b[12] *
mean(x[, 12])
# Model
d <- 13; # degrees of freedom for t
for (i in 1 : N) { Y[i] ~ dnorm(mu[i], tau)
# Y[i] ~ ddexp(mu[i], tau)
# Y[i] ~ dt(mu[i], tau, d)
mu[i] <- beta0 + beta[1] * z[i, 1] + beta[2] * z[i, 2] + beta[3] * z[i, 3] +
beta[4] * z[i, 4] + beta[5] * z[i, 5] + beta[6] * z[i, 6] + beta[7] * z[i, 7] +
beta[8] * z[i, 8] + beta[9] * z[i, 9] + beta[10] * z[i, 10] + beta[11] * z[i,
11] + beta[12] * z[i, 12]
stres[i] <- (Y[i] - mu[i]) / sigma
outlier[i] <- step(stres[i] - 2.5) + step(-(stres[i] + 2.5))
# Priors
beta0 ~ dnorm(0, 0.00001)
beta[1] ~ dnorm(0.8421, 0.7091)
beta[2] ~ dnorm(0.4857, 0.236)
beta[3] ~ dnorm(0.5853, 0.3426)
beta[4] ~ dnorm(0.3932, 0.1546)
beta[5] ~ dnorm(0.4457, 0.1986)
beta[6] ~ dnorm(0.2190, 0.0480)
beta[7] ~ dnorm(0.4516, 0.2040)
beta[8] ~ dnorm(0.5316, 0.2826)
beta[9] ~ dnorm(0.6375, 0.4064)
beta[10] ~ dnorm(0.3502, 0.1226)
beta[11] ~ dnorm(0.2422, 0.0586)
beta[12] ~ dnorm(0.1505, 0.0226)
}
    
```

Figure 5. Program code

In next stage total 11,000 simulations were extracted using Gibbs Algorithm to perform Bayesian regression analysis, and then by discarding the first 1,000 simulations to get initial transients the entire 10,000 samples were extracted from posterior distribution. Also, posterior mean, standard deviation, and 2.5th, 50th, 97.5th percentile of the simulations were calculated.

The nod statistics indicates various result values of posterior distribution such as mean, standard deviation and so on (see Figure 6). Node shows the name of the unknown quantity. b0 means intercept of regression model, and b[1], b[2], ..., b[11] indicates each independent variable. Mean shows coefficient of each variables.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b[1]	0.05139	0.04437	4.622E-4	-0.03739	0.05167	0.1362	1001	10000
b[2]	0.04404	0.1402	0.001355	-0.2292	0.04325	0.3183	1001	10000
b[3]	1.068	0.9485	0.01021	-0.8034	1.065	2.92	1001	10000
b[4]	-0.9255	0.5822	0.006488	-2.084	-0.9309	0.2319	1001	10000
b[5]	0.1926	0.1553	0.001704	-0.1197	0.195	0.4968	1001	10000
b[6]	0.3331	1.802	0.01899	-3.233	0.3524	3.879	1001	10000
b[7]	-3.896E-5	2.156E-5	2.16E-7	-6.003E-5	-3.916E-5	4.436E-6	1001	10000
b[8]	-0.04963	0.1072	0.001024	-0.259	-0.04816	0.1662	1001	10000
b[9]	-0.003968	0.004934	5.038E-5	-0.01362	-0.00402	0.00591	1001	10000
b[10]	-6.018	2.166	0.02064	-10.28	-6.022	-1.687	1001	10000
b[11]	3.468	1.935	0.01937	-0.4072	3.49	7.249	1001	10000
b[12]	-1.687E-6	8.078E-7	9.042E-9	-3.302E-6	-1.675E-6	-7.857E-8	1001	10000
b0	9.17	5.994	0.06216	-2.598	9.201	21.32	1001	10000
sigma	2.653	0.4701	0.007285	1.909	2.591	3.75	1001	10000

Figure 6. Node statistics

The MC error means the computational accuracy of the mean and can be made as small as desired by increasing the number of simulations. The On the other hand, posterior standard deviation represents genuine uncertainty and cannot be reduced others than by obtaining additional real data (Spiegelhalter et al., 2003).

Figure 7 shows the values of simulations using markov chain monte carlo (MCMC) and whether they become well-mixed. If the value in this graph is not fluctuated but constantly maintained during interaction, it will be difficult to judge that this simulation have sufficient reliance. In this sense, the values of these simulations became well-mixed since the part to be constantly maintained was not found given the graph of posterior trace. Posterior distributions were also graphed in Figure 8. Every graph of every node indicates that the posterior distributions became well-mixed.

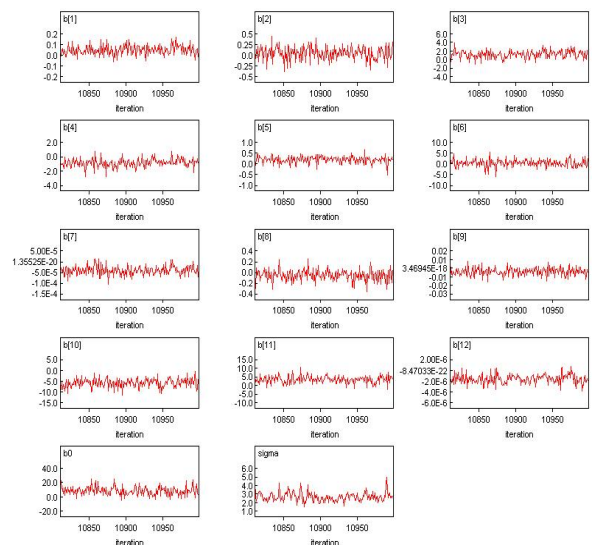


Figure 7. Graph of posterior trace

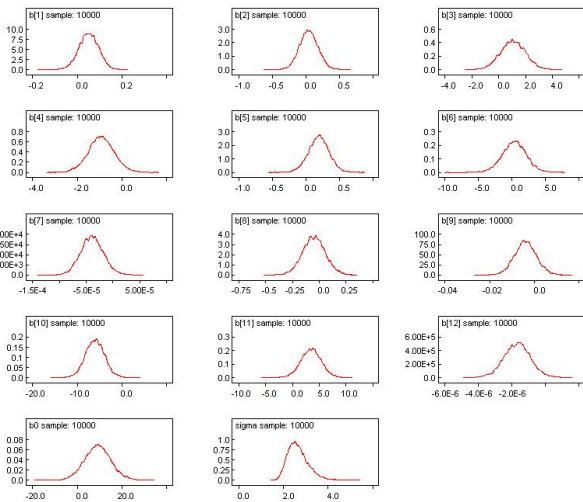


Figure 8. Graph of posterior density

**(5) Analysis and comparison of the model outputs**

Table 9 shows the coefficients of independent variables of traditional MRA and BRA models. Most of independent variables in the two models have the same positive or negative impact to dependent variable. For example, if the number of basement increased, the profit would decrease, which means that this factor has a negative impact to profit margin. However, the coefficient of each variable calculated by BRA was smaller than that by traditional MRA accept for Duration, Location and Retaining wall variables. Especially although the duration and location were significantly low values in correlation with a dependent variable in the traditional MRA model, the values in the BRA model got higher, which means that informative prior, six professionals' views, largely influenced posterior value of the two factors. In the standard error, every variable except the retaining wall variable in BRA was smaller than counterpart, which means that the variability of coefficient reduced.

Table 9. Coefficient of the MRA and BRA models

Variable Name	SPSS		Bayesian	
	B	St. Error	B	St. Error
Intercept	1.4316	6.6608	9.1700	5.9940
Duration	0.0139	0.0694	0.0514	0.0444
Fluctuation	0.2038	0.2155	0.0440	0.1402
Location	0.9735	1.3996	1.0680	0.9465
N. of basement	-1.5028	0.9930	-0.9255	0.5822
N. of story	0.3261	0.2004	0.1926	0.1553
Type of building	1.8147	2.1703	0.3331	1.8020
Gross floor area	-5.46E-05	2.62E-05	-3.89E-05	2.16E-05
Building coverage ratio	0.0703	0.1033	-0.0486	0.1072
Floor area ratio	-0.0094	0.0051	-0.0040	0.0049
Retaining wall	-2.2874	1.4398	-6.0180	2.1660
Type of structure	5.3853	2.7512	3.4680	1.9350
Contract sum	-1.98E-06	1.22E-06	-1.58E-06	8.07E-07

The two regression models were tested by holdout set that were not used to calculate the regression models. Table 10 shows profit predictions of the two models. The profit predictions of BRA model were much more accurate than the model calculated by traditional MRA. The differences of BRA model versus real profit margins was arranged from 1.82 to 0.01, whereas that of traditional MRA model showed from 0.75 to 4.71. Consequently, it is confirmed that BRA model reflecting informative priors measured by six professionals allows improving the accuracy of profit predictions of projects.

Relatively, prediction of test case 2 was calculated to significantly negative results although real profit margin was 1.94%. There are two main reasons why the prediction occurred. The first reason is a lot of risk factors increasing construction cost not to be able to predict in tender phase. Another reason is due to the feature of linear regression. Because the linear MRA models indicate the values on a certain line made by the impacts that independent variables affects a dependent variable. It is difficult that the linear MRA models predict every data accurately. Nevertheless, the results indicated that BRA has sufficient probability to improve the quality of decision making process for optimum tender price.

Table 10. Comparison of predicted profit margin

Test case	Real Profit	Traditional MRA		BRA	
		Prediction	Difference	Prediction	Difference
1	-0.5	4.21	4.71	-0.51	0.01
2	1.94	-0.37	2.31	0.12	1.82
3	6.37	5.39	0.98	6.21	0.16
4	8.37	7.62	0.75	7.95	0.42

**6. SUMMARY AND CONCLUSION**

Bayesian statistics is an appropriate technique to deal with both the factors so that it is possible to reflect prior probability. The concept of Bayesian approach to statistics is to use the prior information and the information contained in the data. It should be handled by subjective experiences and knowledge. This study tried to create the strategy model to predict the actual profit of the construction projects for deciding optimum tendering price by affecting the factors. As a result, this study developed the model to predict the real profit margin of construction projects more accurately using BRA whose reliability was verified by testing and comparing the results. Given the feature of the Bayesian statistics, it can be effectively used to develop a strategy model. Consequently, it is confirmed that BRA model reflecting informative priors measured by six professionals allows improving the accuracy of profit predictions of projects.

The two objectives of this study were accomplished in terms of developing the model to make predictive profit margin at tendering phase using BRA verifying the reliance. Furthermore, contractors will be able to use this analysis to judge how variable factors influence the profit of projects before determining tender prices. Especially if the constructors have sufficient data of previous completed projects and experiences, they can secure the significant reliability of the analysis as well as estimate optimum tender price reflecting appropriate profit.

However, only 33 projects were dealt with to develop the model, and the number of variables was restricted to 12 due to the reason of difficulty in securing sufficient data. It was a significant barrier in improving the accuracy of the model. Moreover, the level of the influence factors has on profits of construction projects. Bayesian software, WinBUGS, also has some limitations in dealing with many independent variables although it has significant benefits. It is impossible to calculate a large number of significant factors. This study used the way to rescale the data to avoid values which might lead to overflow. And, it is to need a lot of time and effort for dealing with the software compared to the traditional MRA.

This study mainly dealt with technical factors as independent variables such as retaining wall, type of structure, the number of basement and story. As a result of the development of construction technology, however, the risk of increase of construction cost by the technical factors has significantly reduced. Instead, external factors such as claim, infrastructure and local legislations are increasingly influencing construction cost. Especially such factors can result in the extension of project time, which can deteriorate the project profitability by increasing final construction cost. In this context, considering more external factors, completed project time, market and bidding condition is required for developing strategy model to decide optimistic tender price reflecting appropriate profit.

## REFERENCES

- Bolstad, W.M. (2007) "Introduction to Bayesian statistics." 2nd ed. Hoboken: John Wiley & Sons, Inc.
- Christodoulou, S. E. (1998) Optimum bid markup calculation in competitive bidding environments using fuzzy artificial neural networks. Doctoral Dissertation, Columbia University, New York, NY.
- CIOB. (2009) "Code of estimating practice." 7th ed. Chichester: CIOB/Wiley-Blackwell.
- Egan, J. (1998) Rethinking construction-report of the construction task force: [www.dti.gov.uk/construction/rethink/report/index.htm](http://www.dti.gov.uk/construction/rethink/report/index.htm)
- Elhag, T.M.S., Boussabaine, A.H. & Ballal T.M.A. (2005) "Critical determinants of construction tendering costs: Quantity surveyors." *International Journal of Project Management*, 23(7): 538-545.
- Grogan, T. (1992) "Low bids raise hidden costs." *Engineering News-Rec.*, 228(13): 30-31.
- Hair, J.F.Jr., Anderson, R., Tatham, R., & W.C. "Multivariate data analysis." 5th ed. Upper Saddle River, NJ: Prentice Hall.
- Henriod, E.E., & Lanteran, J.M. (1988) "Trends in contracting practice for civil works." Task Force on Innovative Practice, World Bank, Washington, D.C.
- Hinton P.R., Brownlow C., McMurray I. & Cozens B (2004) "SPSS Explained." New York: Routledge Inc.
- Jaggar, D., Ross, A., Smith, J. & Love P. (2002) "Building design cost management." Oxford: Blackwell Science.
- Kim H.J. & Reinschmidt, K.F. (2011) "Effects of contractors' risk attitude on competition in construction." *Journal of Construction Engineering and Management*, 137(4): 275-283.
- Laryea, S. & Hughes, W. (2011) "Risk and price in the bidding process of contractors." *Journal of Construction Engineering and Management*, 137(4): 248-258.
- Lo, W. Lin, C.L. & Yan, M.R. (2007) "Contractor's Opportunistic Bidding Behavior and Equilibrium Price Level in the Construction Market." *Journal of Construction Engineering and Management*, 133(6): 409-416.
- Mochtar, K. & Arditi, D. (2001) "Pricing strategy in the US construction industry." *Construction Management and Economics*, 19(4): 405-415.
- Nellis, J.G. & Parker, D. (1992) "The essence of business economics." New York ; London : Prentice Hall.
- Seydel, J. (2003) "Evaluating and comparing bidding optimization effectiveness." *Journal of Construction Engineering and Management*, 129(3): 285-292.
- Shash, A.A. (1995) "Competitive bidding system." *Cost Engineering*, 37(2): 19-20.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003) "Winbugs user manual statistics Version 1.4." Cambridge, UK; MRC Biostatistics Unit.

(Date of Submission : 2011.7.25)