DEVELOPING A CONSTRUCTION-DURATION MODEL BASED ON A HISTORICAL DATASET FOR BUILDING PROJECT

Ming-Chiao Lin[1], Hui Ping Tserng[2], Shih-Ping Ho[3], Der-Liang Young[4]

[1]*Division of Construction Engineering and Management, Department of Civil Engineering,*
National Taiwan University, No. 1 Roosevelt Rd., Sec. 4, Taipei, Taiwan
[2, 3, 4]*Department of Civil Engineering, National Taiwan University, No. 1 Roosevelt Rd., Sec. 4, Taipei, Taiwan*
E-mails: [1]d93521019@ntu.edu.tw; [2]hptserng@ntu.edu.tw (corresponding author);
[3]spingho@ntu.edu.tw; [4]dlyoung@ntu.edu.tw

Abstract. The delay of vast building projects is still a common problem. This situation is extraordinarily severe to steel reinforced concrete (SRC) building projects that keep going to promote a new structure system in Taiwan's construction industry. The aim of this study is to develop a feasible contract duration model based upon few SRC building cases. A logical approach is employed to select and assure the "good" regression model identified when project characteristics were known and external uncertainties were reasonably estimated. Different necessary diagnostics had been adopted to examine the aptness of the model before inference. The cross-validation is used to validate the appropriateness of the variables selected and magnitudes of the regression coefficients. The mean of the square prediction errors (MSPR) is selected to measure the predictive ability of the model proposed, and the result shows that the predictive ability of the selected regression model could be adequate. Finally, several cases are taken to test the predictive accuracy of the model proposed, and the result shows that the actually necessary construction duration is considerably closed to the duration predicted by the mode. It is concluded that the predictive duration model proposed could be applicable to the SRC construction projects with a reasonable reliability.

Keywords: building project, construction, duration prediction, regression, heteroskedasticity, validation.

1. Introduction

People in Taiwan experienced an unprecedented catastrophe, an earthquake measuring 7.3 on the Richter scale, on September 21, 1999. Damage caused by the Chi-Chi Earthquake included thousands of deaths and severely wounded, with 44,338 houses completely destroyed, 41,336 houses severely damaged, and a total of US $9.2 billion worth of damage. As a result, builders and architects of modern buildings which collapsed were detained and accused by authorities. After the impact of the Earthquake, a conservative structural frame system, members of frame with Steel Reinforced Concrete (SRC) with better security, has gradually started to be adopted in the construction industry. In these SRC building contracts, clients had set aside additional duration for construction, but construction delay still generally occurred. The contract duration needed for SRC building does not included the duration of Steel structure construction and the duration of RC construction. Hence, difficulties and complexities of SRC building in the construction phase always give rise to delay. This leads to building cost far more expensive than tradition RC and Steel structure unless the owners take into account durability and building safety. Therefore, SRC structures are not popular in Taiwan today, with only few cases up to writing this research although promotion of new SRC structures in Taiwan's construction industry is still ongoing. However, there is no national specification on the appropriate construction duration for SRC structure, causing problems regarding construction duration for owners and constructors. This leads to cases of falling behind contract schedule and give rises to disputations.

The construction duration has been observed as one of the main criteria for assessing the performance of building projects in the construction industry (Bromilow 1969; Dissanayaka, Kumaraswamy 1999; Kaka, Price 1991; Love *et al.* 2005; Ng *et al.* 2001; Ogunsemi, Jagboro 2006; Walker 1995). A project will be considered successful if it is completed on time, is within budget, and meets the specified quality standards (Chan, Kumaraswamy 1996, 1997). Understandably, schedule overrun brings about project cost overrun. Although the industry participants are aware of the importance of duration in the construction phase of projects, it was observed that significant part of the contracts had not met the stipulated period. Since 1960's, according to Bromilow's research report, only 12.5% of building contracts were completed within the scheduled completed dates and the overall average actual time is more than 40% longer than the original schedule (Bromilow 1969). After four decades, the inability to complete on time is still a prevailing problem. Al-Khalil

and Al-Ghafly (1999) reported that completed public projects overrun approached 70% of the original schedule through a comparison of the outcomes for projects with their original schedules. Odusami and Olusanya (2000) concluded that projects executed an average delay of 51% of a planned duration for most projects in the metropolis. Blyth *et al.* (2004) stated that 50% of contracts were completed after stipulated durations longer than schedules. Iyer and Jha (2006) observed that over 40% of construction projects were behind original schedule and delay lasted for months. Completed projects lasting much longer than the original schedule results in various disadvantages such as: additional cost, reduced contractor's profit, loss of reputation, and delay of client's operating plan. Many provoking disputes emanated from the various reasons of construction delay (Ng *et al*. 2001). The construction duration overrun is problematic in the construction industry and generates much concern. These statues of delay are still universal in the performance of building projects (Aibinu, Odeyinka 2006).

Several methods exist for estimating the contract duration of building construction projects such as the period expected by client, the special consideration, the time requirement for the project work to be done, and the time taken as recorded through historical project information (Kumaraswamy, Chan 1995; Love *et al.* 2005). In practice, most methods estimating project duration in the industry depends on the subjective skill and cognition of the estimators and planners, rather than on objective assessment.

Khosrowshahi and Kaka (1996) stated that forcing the project into a desired time mold can lead to adverse consequences, giving rise to a chain reaction which affects the performance of the organization in other areas. Underestimation of the project duration raises additional events of penalty, disputes, etc. for the contractors and clients. On the other hand, overestimation of the project duration may lose organizational competitiveness in the industry. Both of these could have undesirable effects on project performance and achievement of project objectives.

Accurate and reasonable contract duration may avoid higher bid cost, and decrease the possibility of disputes between contractors and clients. It is also useful to the contractor in assessing the risks of meeting the client's requirement. The aim of this study is to develop a feasible and suitable project contract duration model based on the historical data set, enabling clients and contractors to estimate better accurate construction duration.

2. Literature review

There were a wide variety of approaches in dealing with the factors affecting construction durations. Several predicting models have been developed from these factors. Bromilow (1969) built a relation between the duration and cost of building contracts known as the Bromilow's time-cost (BTC) model. The BTC model revealed that the time taken to construct a project is only highly correlated with the construction size as measured by the final cost. Ng *et al.* (2001) revisited the BTC model with more new

project data due to improved productivity, checked on its appropriateness for various data subgroups including those of a project type, and compared it with previous models developed at different time periods. Results show that different parameter estimates are needed for different project types.

Researches pointed out that the BTC model potentially falls short by not considering factors other than cost when establishing the construction time for a given project (Ng *et al.* 2001; Walker 1995; Nkado 1992, 1995), further developing the construction time model by categorizing the activities during the construction phase into work groups: substructure; superstructure; cladding/envelope; finishes; M&E services; and their sequential start-start times. The durations of these activity groups can be predicted from 11 variables: gross floor area; area of ground floor; approximate excavated volume; building height; number of stories; end use; cladding type; presence of atrium; building location; intensity of services and site accessibility. It was claimed that the model could provide an objective basis for evaluating the implications of the clients' stipulated completed times at early stages of design.

Khosrowshahi and Kaka (1996) assumed that a project can be defined in terms of a series of variables that characterize the project. They identified the most influential variables and combined a number of these variables such as scope; floor; start-months; horizontal-across; build-ability; frame; project-operation; units; abnormality and log cost to develop and propose a model to estimate project duration. Based on these variables, Chan and Kumaraswamy (1999a, b) carried out several investigations in the public housing building construction process to identity a set of significant variables influencing construction duration of projects. The durations of the primary work packages, i.e., piling; pile caps/raft; superstructure; E&M services; finishes and their respective sequential start-start log times; were modeled in terms of the identified sets of critical factors by multiple linear regression exercises, concluding that the model developed was applicable to the public housing industry. Love *et al.* (2005) analyzed certain significant variables, i.e., project duration; project type; procurement method; tender type; gross floor area (GFA) and the number of stories, proposing an alternative model to replace the BTC model. They concluded that the GFA and the number of stories in a building were key determinants of time performance in forecasting construction duration in the project. Blyth *et al.* (2004) developed a predicting duration model based on project characteristics by combining the twenty one most influential project variables. They concluded that the predicting model can achieve the 7% maximum in predicting over duration. However, Kaka and Price (1991) revealed that private building varied significantly and fitted poorly in this model, suggesting that further classification of projects may be required to enhance the accuracy of the relevant variables relationship. Ogensemi and Jagboro (2006) opinioned that the BTC model is a non-linear type form, and introduced a breakpoint between two linear models, forming a piecewise linear model to improve the accuracy of the BTC model. In a

questionnaire employed to determine factors affecting the performance of construction project, climate condition at site had been identified as the most important factor by owners; consultants and contractors because it affected the productivity and time performance of project (Enshassi *et al.* 2009). Zavadskas *et al.* (2010) assumed that risk could make cost and time overrun in the construction projects. They divided project risks into three groups: external; project and internal risk, where weather is considered as external risk. According to the aforementioned existing literature, there are a number of variables which may act individually or in combination to influence project duration. Therefore, it is worthy to further study how suitability of SRC building projects on construction duration. The extent of the effect of those variables is dependent on the nature of the project and external uncertainties. Since construction projects are distinguished diversiform categories such as building, civil and others, the homogeneity of data should be divergent from one another. The study would focus on private SRC building projects to build up a reasonable construction duration model.

3. Methodology of the model development

Multiple linear regression analysis is a widely used statistical procedure for determining the relationship among relevant variables (Blyth *et al.* 2004; Chan, Kumaraswamy 1997, 1999a, b; Love *et al.* 2005). It is difficult to acquire only one best significant combination from the vast and potential independent variables in the objective approach. The identification of potential proper variables requires some intuition and practical experimental experience. Due to the variation in the uses of regression models, no particular subset of explanatory regression variables is the best. A descriptive use of a regression model typically will emphasize precise estimation of the regression coefficients, whereas a predictive use will focus on the prediction errors (Neter *et al.* 1996; Siegel 2003). In general a regression model includes the selection of explanatory diagnostics for examining the appropriateness of a fitted regression model, the remedial measures when the model conditions are not met, and validation of the regression model.

In this study two types of variables are incorporated into the construction prediction model. One is categorized into project characteristics which had some influence on the construction duration performance extracted from the discussions of previous researchers. The other is uncertain external factors which cannot be foreseen by the clients or contractors during construction. Several factors investigated, together with practical common recognition in construction industry, have been used in this predicting duration models, with project characteristics including construction cost, duration, and size of project etc., whereas the uncertain variables are defined as the external condition and internal condition.

Variables in models need to be properly defined. The contract duration is defined as the date from the agreed construction commencement date to the planned completion date. The actual construction duration for a project is considered to start when the date of contract commencement takes effect and end on the date of practical completion on site. The main reason for considering the agreed construction commencement date is that project activities are not usually continuous from project inception to contract commencement due to preparing or awaiting some unexpected events. However, the operation of the contract has become effective. On the other hand, contract initial cost is taken as the tender price or construction budget, whereas the actual cost is the final cost in the construction phase. The contract final cost could be different from the contract initial cost when variation during construction phase is taking into consideration. Except for the aforesaid possible variables, the identification of other potentially proper variables requires an intuition and practical approach.

When projects are executed at different periods, it would be necessary to adjust for fluctuating factors to avoid the disparities and consider the discounted values of project cost in relation to a particular year; the 2006 contract price indices are employed as the common-base set of data.

A total amount of fifty-six building projects with a total value exceeding US $212,977,449 dollars were collected for analysis. The procurement of all these projects uses the traditional design-bid build concept. This database represents new building in Taiwan completed in the period between 2000 and 2006. The contract cost for the project ranges from $842,356 dollars to $11,969,230 dollars, having a mean (M) of $3,803,169 dollars and standard deviation (SD) of $ 2,477,694 dollars. The buildings ranged from 2,251 m^2 to 51,904 m^2 gross floor area (GFA), 2 to 14 stories high, and took 197 to 1080 days to construct. Results showed that only a few contracts were completed on or before the date originally expected. The time performance for the steel reinforced concrete (SRC) building projects was far worse than expected.

The premises of building projects carried out in this research are:

1) All of the contractors are competent and efficacious in setting up the construction process as well as work within the same norm.

2) Mass materials such as steel, windows etc. are provided under the condition that there are no delays occur by the supplier chosen by the client.

3) All buildings have the same structure system (SRC) and most building materials are similar in likelihood.

In multiple linear regression analysis, one important objective is the identification of the best combination of variables in acquiring better significant combinations from the vast and potential independent variables within the objective. There are some techniques in SPSS (10.0) software package which are used as the alternative to the best solution. In the study, the forced entry technique is adapted to conduct the multiple regression analysis. It will be found that different combinations might produce different results during the process of developing the statistical models. A significant level of 10% is set to test.

There is a significant predicting relationship between the project construction duration and a relatively small number of measurable parameters of a project and its environment. These steps identified the suitable combination of the constituent variables of the model as follows:

1) The determination of the dependent variable to dig into the possible predictor (independent) variables from the data set.

2) The observation of the histogram of the data to test the normality in the distribution; if not, the variables are transformed into a suitable form such as the logarithm scale.

3) The observation of the scatter plot to examine the linearity in the distribution of the data; if not, one or two variables need to be transformed into a suitable form such as a logarithm form, resulting in a linear relationship for the linear regression technique to be used accordingly.

4) The regression of predicting an equation model to identify the variables which pass the statistical testing and can reasonably explain the variation within the data.

5) Omitting or adding the model variables.

6) Repeating steps 4 and 5 until the following criteria are satisfied:

− all important and possible variables are included in the model;

− F and t statistical values pass the hypothesis testing value;

− R^2 or adjusted $\overline{R}^2$ has reached a saturation level.

The construction model frame mentioned previously can be exhibited as Fig. 1.

4. Building the regression model

In this study, the aim of the research is to build a feasible and applicable contract stipulated duration model using the prediction regression algorithm for clients and contractors. We present a strategy for the model-building process which involves several phases. The first phase is to identify a subset that includes three explanatory variables quested from the characteristics of the building construction contract to build a fitted regression model. The first phase regression model is fitted with the following results: $\overline{R}^2 = 0.908$, F = 172.057. This should be contently acceptable for the fitted results. In the next steps, the regression model formed in first phase is applied to fit the actual duration; the result is considered "not good" as the stipulated duration results obtained were: $\overline{R}^2 = 0.658$, F = 33.321. When the explanatory variable (contract cost) is substituted with actual cost, the result is only "slightly better" than the former with $\overline{R}^2 = 0.698$, F = 43.416. After a hard trial effort, other characteristics have no effect in heightening the value of $\overline{R}^2$ and F. It is imputed that some unknown factors deeply influence the efficacy of the fitted regression model. In accordance with literature, it is judged that the construction-duration performance could be influenced by a pool of uncertain external or internal factors in the construction phase. Subsequently, two factors, number of change orders and rainy days, are formulated and incorporated one after another into the third phase regression model. The number of change order was taken as the average of change orders of all cases during the construction phase while rainy days were estimated as the average of rainy days which cease construction in the same duration (e.g. the same commence month and completing month that the case was taken) of by-gone three years. This process leads to the selection and identification of the ultimate regression model, fitted with the following "good" results: $\overline{R}^2 = 0.920$, F = 115.315. The procedures of operation are illustrated in Table 1.

The final regression model based on the 56 cases is:

$$\hat{Y}_i = 1.238 + 0.218X_1 - 0.00092X_2 + 0.02177X_3 + 0.663X_4 + 0.01169X_5$$
$$(0.1107)\ (0.0579)\ \ (0.00064)\ \ \ (0.0039)\ \ \ \ (0.0590)\ \ \ \ (0.0034)$$
$$n = 56,\ \overline{R}^2 = 0.920,$$

$$(1)$$

where: $\hat{Y}_i$ is log (predictive construction duration); X_1 is log (contract initial cost); X_2 is GFA / expected contract duration; X_3 is number of stories; X_4 is modified contract duration (estimated rainy days + expected contract duration) / expected contract duration; X_5 is change order; (…) is standard error of the estimated coefficient; $\overline{R}^2$ is adjusted R^2; n is number of case.
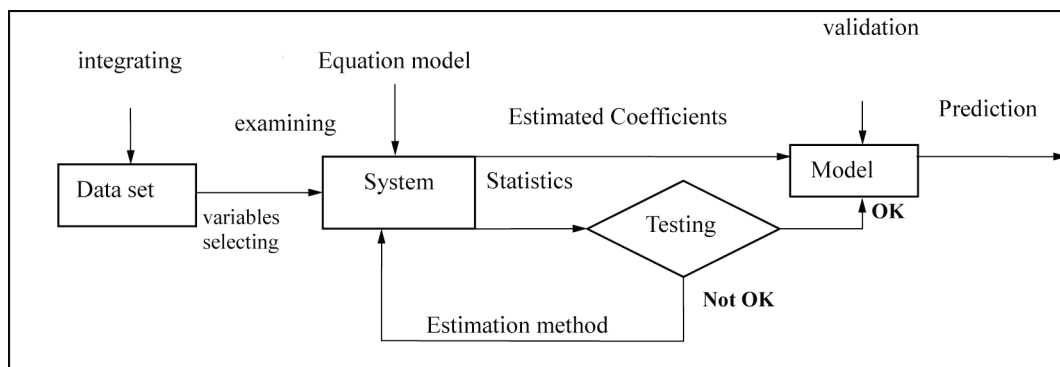

Fig. 1. Construction Model Frame

Table 1. The Cooperation for Potential Explanatory Variables of 4 Different Phases

PHASE	1st	2nd	3rd	4th
constant	1.651 (23.240)	1.730 (11.507)	1.119 (13.767)	1.238 (11.591)
X_1 log(contract initial cost)	0.388 (8.033)	0.499 (4.390)	1.053 (10.352)	0.218 (3.762)
X_2 (GFA/ expected contract duration)	−0.00343 (−6.506)	−0.002990 (−2.675)	−0.00203 (−3.317)	−0.000918 (−1.431)
X_3 Stories	0.02895 (8.616)	0.01870 (2.630)	0.02699 (6.868)	0.02177 (5.622)
X_4 (modified contract duration (estimated rainy days + expected contract duration) / expected contract duration)			0.709 (11.205)	0.663 (11.234)
X_5 Change orders				1.169E−02 (3.455)
R	0.953	0.811	0.949	0.959
$\overline{R}^2$	0.908	0.658	0.901	0.920
F	172.057	33.321	116.236	115.315

∗ The actual construction duration is substituted by the expected contract duration during the 1st phase.

∗ The (…) is *t* value of the parameter.

Table 2. Model Selection Criteria Plot

Regression Model	R^2	$\overline{R}^2$	MSE	AIC	
$X_1 X_2 X_3 X_5$	0.719	0.697	9.130E−02	−263.327	
$X_1 X_2 X_3 X_6$	0.658	0.632	0.10063	−252.425	
$X_1 X_2 X_4 X_5$	0.870	0.860	6.213E−02	−306.434	
$X_1 X_2 X_4 X_6$	0.836	0.823	6.969E−02	−293.572	
$X_1 X_3 X_4 X_5$	0.917	0.910	4.962E−02	−331.620	
$X_1 X_3 X_4 X_6$	0.889	0.881	5.730E−02	−315.494	
$X_1 X_3 X_5 X_6$	0.718	0.696	9.138E−02	−263.222	
$X_1 X_4 X_5 X_6$	0.859	0.847	6.475E−02	−301.811	
$X_2 X_3 X_4 X_5$	0.898	0.890	5.508E−02	−319.914	
$X_2 X_3 X_4 X_6$	0.846	0.834	6.751E−02	−297.139	
$X_2 X_3 X_5 X_6$	0.683	0.658	9.699E−02	−256.551	
$X_1 X_2 X_3 X_4 X_5$	0.920	0.912	4.912E−02	−331.868	**selected**
$X_1 X_2 X_3 X_4 X_6$	0.901	0.891	5.465E−02	−319.903	
$X_1 X_2 X_3 X_5 X_6$	0.719	0.691	9.212E−02	−261.425	
$X_2 X_3 X_4 X_5 X_6$	0.898	0.888	5.560E−02	−317.986	
$X_1 X_2 X_3 X_4 X_5 X_6$	0.921	0.911	4.933E−02	−330.522	

5. Criteria for model selection

For variables reduction, it is necessary to identify a small group of "good" regression models according to some specified criteria. Those criteria could provide time-saving algorithms for identifying the "best" subset. During the comparison of the regression models, more than one criterion is considered in evaluating the possible subsets of independent (X) variables. In this study, those different criteria for comparing the regression models that would be used with the regression selection procedure are R^2, adjusted $\overline{R}^2$, AIC (Akaike Information Criterion) and mean square error (MSE).

The potential **X** variables included six explanatory variables, leading to 2^6 possible subset regression models that could be formed in the selection procedure. Table 2 illustrated an abridgement of all different possible subsets by the all-possible-regression approach (Neter *et al.* 1996). In the "best" subsets algorithms, four criteria points to the same "best" subset, subset (X_1, X_2, X_3, X_4, X_5), which may be regarded as the tentative regression model. The selection of the final regression model depends greatly upon diagnostic results. Residual plots and diagnostic checks are performed mainly to identify influential outlying observation, multicollinearity, heteroskedasticity, etc., and to examine the appropriateness of the fitted regression model.

6. Diagnostics

When a regression model is built, it is important to examine the aptness of the data model before inferences are

made. In this study, some graphic methods for studying the appropriateness of the model build, such as linearity of the regression function or normality of the error terms and the like, as well as several formal statistical tests will be discussed.

Scatter Plot Matrix and Correlation Matrix. A scatter plot matrix facilitates the study of the relationships among the variables by the scatter plots within a row or a column. Scatter plots of the response variable and against each predictor variable can help determine the nature and strength of the bivariate relationships between each of the predictor variables and the response variables. The scatter plot can also find gaps and outliers in the data points.

Table 3 and Fig. 2 show that some of the predictor variables are correlated with each other. The degree of linear association among the predictor variables is moderate or relatively low.

Table 3. Scatter Plot Matrix and Correlation Matrix (Correlations)

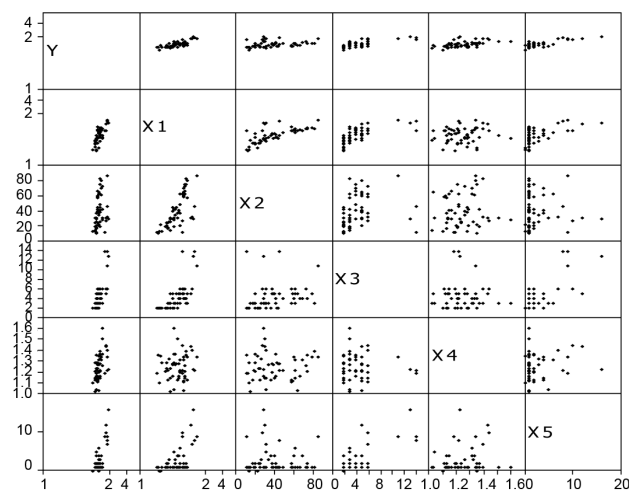|	Y	X_1	X_2	X_3	X_4	X_5
Pearson Correlation						
Y	1.000	0.688	0.265	0.719	0.546	0.750
X_1	0.688	1.000	0.746	0.624	0.126	0.488
X_2	0.265	0.746	1.000	0.207	0.060	–0.014
X_3	0.719	0.624	0.207	1.000	–0.027	0.683
X_4	0.546	0.126	0.060	–0.027	1.000	0.185
X_5	0.750	0.488	–0.014	0.683	0.185	1.000

Fig. 2. Scatter plot matrix among variables

Residual. The following plots of residuals are taken from several informal diagnostic plots of residuals to provide information on any types of departures from the linear regression model.

1) The residual plot against the fitted value in Fig. 3 shows no evidence of serious departures from the model.
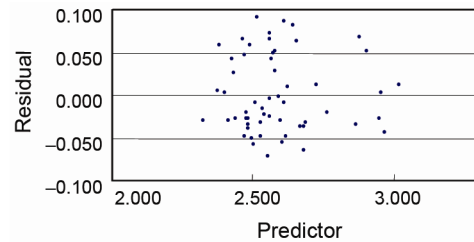

Fig. 3. Residual Plot

2) The normal probability plot of residuals in Fig. 4 illustrates a slight departure from linearity. However, the problem of normality was not considered a serious impact on inferences to be made from the regression.
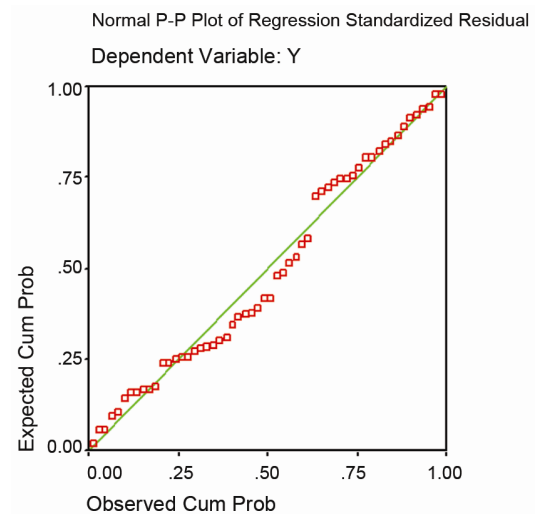

Fig. 4. Normal P-P Regression Standardized Residual Plot

Test for Heteroskedasticity. The homoskedasticity assumption for multiple linear regression states that the variances of the unobservable error conditional on the explanatory variables is constant. When homoskedasticity fails, the standard errors are no longer valid for constructing confidence intervals and t statistics. Similarly, F statistics are no longer F distributed and Lagrange multiplier (LM) statistic no longer has an asymptotic chi-square distribution (Wooldridge 2002). That is, the statistics used to test the hypothesis for assumptions are not valid in the presence of heteroskedasticity. Many tests for heteroskedasticity have been suggested over the years. In this study, the Breusch-Pagan test is applied to test for heteroskedasticity (Wooldridge 2002). We assumed that the ideal assumption of homoskedasticity holds, and we required the data to tell us otherwise. The steps for testing for heteroskedasticity are abbreviated. Regressing the squared OLS (ordinary least squares) residuals on the independent variables, produced R-squared = 0.082; thus, LM = 4.592, and the p-value = 0.287. The p-value (the smallest significance level for test) exceeds the desired significance level. Therefore, we fail to reject the null hypothesis of homoskedasticity in the model proposed at 10% level. We may conclude that heteroskedasticity is not a problem to the study model.

Identifying Outlying Observations. The outlying or extreme cases may involve large residuals and have dramatic effects on the fitted least square regression function. There is a need to identify the outlying cases carefully and decide whether they were to be retained or eliminated. A case may be regarded as outlying with respect to its Y value, its X value, or both. Not all outlying cases have a strong influence on the fitted function. The following steps were performed to determine if the regression model under consideration is heavily influenced by one or a few cases in the data set.

Identifying Outlying Y Observations. Frequently, the detection of outlying Y observation is based on an examination of the residuals. In the study, we utilize the mean of studentized deleted residual for the diagnosis of outlying Y observation (Neter *et al.* 1996). A formal test performed by means of the Bonferroni test procedure to determine whether the case with the largest absolute studentized deleted residual is an outlier. If the regression model is appropriate, no case is an outlier due to a change in the model; each studentized deleted residual will follow the t distribution. The studentized deleted residual in Fig. 5a shows that cases 21 and 27 have the largest absolute studentized deleted residual. Case 27 which has the largest absolute studentized deleted residual is an outlier resulting from a change in the model. Using the Bonferroni simultaneous test procedure with a family significance level of $\alpha = 0.10$: $t(1-\alpha/2n;n-p-1) = t(0.9992;50) = 3.506$, where n is number of cases, p is number of parameters.

Since $t(27) = (2.474;3.506)$, we may conclude that case 27 is not an outlier. The other case 21 is also found to be not outlier.

Identifying Outlying X Observations. The leverage value is a useful indicator in a multivariable setting deciding whether or not a case is an outlier with respect to the X value. The leverage values greater than $2p/n(= 0.214)$ are considered as outlying cases with regard to their X values (Neter *et al.* 1996). The leverage value in Fig. 5b shows that cases 53 and 55 have the largest value, between 0.325 and 0.394. Although these values exceed 0.214, they do not exceed 0.5, and hence could indicate a moderate leverage. We shall need to ascertain how influential those cases are in the fitting of the regression function. The following three measures of influence are widely used in practice, each based on the omission on a single case to measure its influence.

1) *Influence on Single Fitted Value*:

The guideline for identifying influential cases indicates that a case is influential if the absolute value of DFFITS exceeds 1 for small to medium data sets (Neter *et al.* 1996). Fig. 5c shows that those DFFITS values of case 27, 49, 51, 53 and 56 were much lower than 1. We may conclude that those cases were not influential to require remedial action.

2) *Influence on All Fitted Value – Cook's distance*:

The Cook's distance measure considers the influence of any one case on all fitted values. The Cook's D value in Fig. 5d depicts that case 49 has the largest distance value, $D = 0.1596$, and is lower than the critical value of 0.904. It may be concluded that case 49 does not influence the regression fit.

3) *Influence on the Regression Coefficients – DFBETAS*:

The guideline for identifying influential cases indicates whether the absolute value of DFBETAS exceeds 1 for small to medium data sets or not (Neter *et al.* 1996). All values of DFBETAS were much lower than 1. We may claim that there were no influential factors which require remedial action.

All three influence measures (DFFITS, Cook's distance, and DFBETAS) have been ascertained to have no influential cases on the fitting of the regression function.

Multicollinearity Diagnostics – VIF. The VIF is a formal method of detecting the presence of multicollinearity which is widely used. The largest VIF value among all X variables is often used as an indicator of the severity of multicollinearity. A maximum VIF value in excess of 10 is frequently taken as an indication that multicollinearity may unduly influence the least squares estimates. From Table 4, it is observed that all values of parameters are lower than 10, point ingout that there is no serious multicollinearity in the model.

Table 4. VIF Value of model parameters

|	X_1	X_2	X_3	X_4	X_5
VIF	6.422	3.930	2.906	2.536	1.112

7. Model validation

Validation of the regression model involves the appropriateness of the variables selected, the magnitudes of the regression coefficients, the predictive ability of the model, and the like. Cross-validation is used to validate a regression model by splitting the data into two sets. The number of cases for a model-building set should be at least 6–10 times the number of variables in the pool of predictor variables (Neter *et al.* 1996), the other is for a model-validation set. In present study, the entire data collected is not enough to make an equal split for the five independent variables selected to develop the regression model. Therefore it was determined that the validation data set is smaller than the model-building data set. The collected thirty-cases set was used for estimating the regression model, and the twenty-six cases set was developed to validate the stability of the model. The result can be compared for consistency with the regression coefficients between the two models illustrated in Table 5.

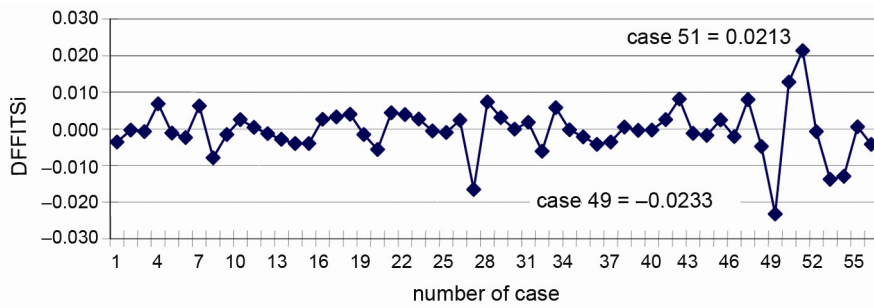**Table 5.** Comparing for the Regression Coefficients of Two Models

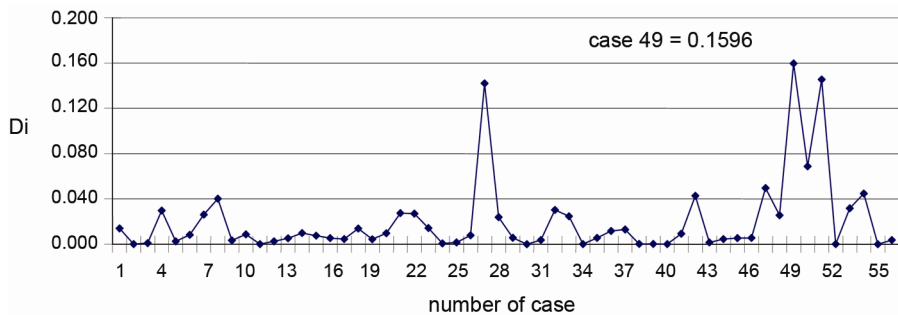| cases	R^2	$\overline{R}^2$	β_0	β_1	B_2	β_3	β_4	β_5
30	0.945	0.934	1.186	0.262	−0.00133	0.02029	0.650	0.01098
26	0.869	0.837	1.260	0.186	−0.00511	0.02217	0.684	0.01297
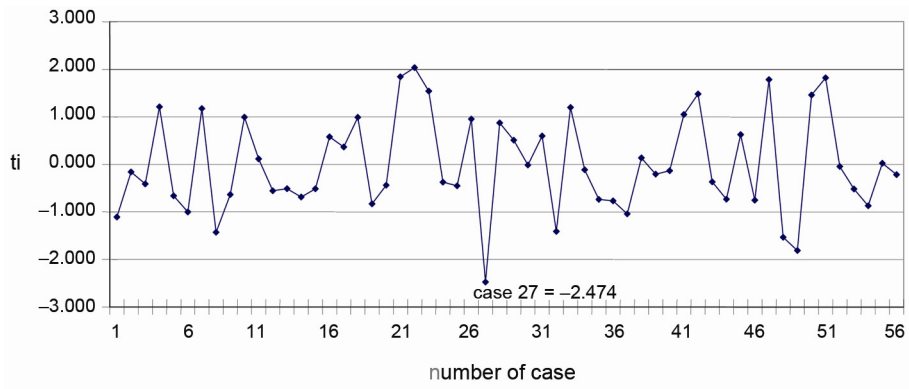
a) Studentized Deleted Residual
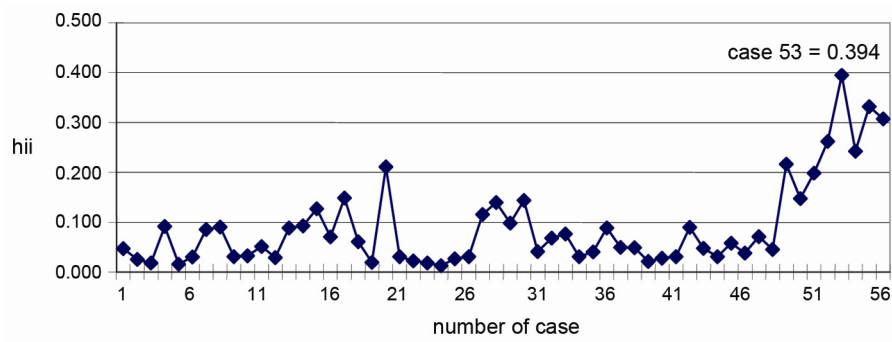
b) Centered Leverage Value

c) DFFITS

d) Cook's Distance

Fig. 5. Outlying Observation Case Plot

8. The Predictive ability of the regression model

Furthermore Neter *et al.* (1996) declared that a mean of measuring the predictive ability of the regression model selected is to use this model to predict each data set, followed by calculating the mean of the squared prediction errors, as denoted by MSPR, which stands for mean squared prediction error:

$$\text{MSPR} = \frac{\Sigma_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}, \qquad (2)$$

where Y_i is the value of the response variable in the ith validation case; $\hat{Y}_i$ is the predicted value for the ith validation case based on the model-building data set; n is the number of cases in the validation data set.

If the mean squared prediction error MSPR is fairly close to MSE based on the regression fit to the model-building data set, it shows that the error mean square MSE for the selected regression model is not seriously biased and gives an appropriate indication of the predictive ability of the model (Neter *et al.* 1996). It was found that the MSPR = 0.002572 is quite close to MSE = 0.002610, highlighting that the predictive ability of the selected regression model could be adequate in the future.

9. Discussion of the results

Some limitation on the applicability of this proposed model arises from the size of samples and the range of building functions that it encompasses. The sample size of 56 sets a limitation, but the adjusted R-squared of 0.920 indicates that the sample size was sufficient to produce a significant model for prediction construction duration. The coefficient of GFA per Contract duration is negative, which indicates that construction duration tends to decrease with GFA per Contract duration for the project sample. The outcome also had been acquired by Love *et al.* (2005). On the other hand, coefficients of other explanatory variables are positive, which indicates that project construction duration tends to increase.

During construction stage, change order of projects might greatly affect the duration of projects; the more change orders, the more influences. At the model building stage of this research, expected change order adopt the average number of change order in construction stage of all 56 cases. For SRC buildings in Taiwan, they are still at early stage of development. Due to characteristics of SRC buildings in construction process, it is far more complicated than traditional RC structure, and leads to more change orders. The average number of change orders is up to 2.7 times and is shown on the equations of model building from this research. As for the estimation of duration by the model proposed in this study, it might be operated and judged through the contents of construction design and tender. In the situation of no change order, the number of change order might be set as zero.

The reason why rainy day is considered in the prediction model of this research is due to the lack of attention in literature. However, the condition of climate do affects the productivity performance of construction projects. Rainy days do influence the implementation of schedule. There are two rainy seasons in Taiwan which are plum raining and typhoon seasons. When construction projects encounter these two seasons, it would be difficult to push construction on schedule. Therefore, this factor is considered as a variable and single out.

The modified contract duration shown in equation 1 is formed by estimated rainy days plus construction days during expected construction period of contract. The estimated rainy day is based on average rainy days of the past three years in the same beginning and finished date. For example, the expected construction duration of one SRC building is ten months from February 1 of 2009 to November 31 of 2009, the estimated rainy day is adopted as average rainy days from 2006 to 2008 in the same beginning and finished date.

Fig. 6 illustrated the accuracy of the model with low deviation and residual. The result showed that the model possesses effective ability to predict construction duration of SRC building project.
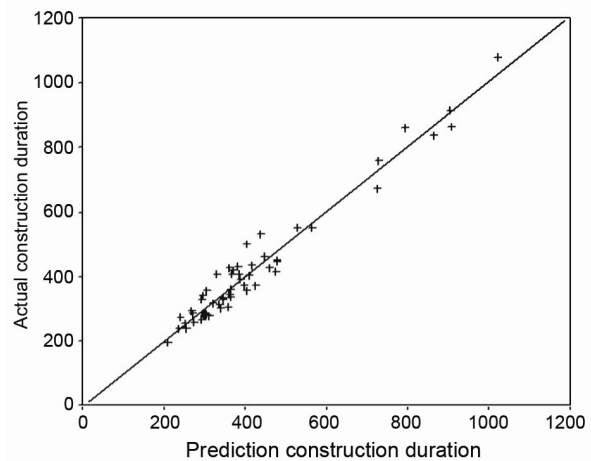

Fig. 6. Predicting Construction Duration vs. Actual Construction Duration

In order to further confirm the predictive ability of the model proposed in practical application, other different but similar 11 SRC construction projects not in the data set used for modeling, as they finished after the sampling, were taken to test. The information of their basic data and the outcomes of prediction are shown as follows: contract initial cost of US $843,750–US $20,090,750, expected contract duration days of 300–900, a gross floor areas of 1,579–60,380 m², stories of 3–15, change orders of 1–9, and actual construction days of 292–1060. The error percentage of the forecasting construction duration came –7.37% to 3.30%. The results showed effectiveness of the model in this study to forecast construction duration for SRC structure. Furthermore, it also elucidated that a case had unreasonable contract duration as underestimated by the client, whereas the actually necessary construction duration was considerably close to the duration predicted by the model proposed. The model could be an objective and reliable tool to client and contractor for estimating the actual necessary duration and further evaluating contract duration of SRC building.

10. Conclusions

In this study, a set of 56 SRC construction cases were used to develop a construction-duration prediction model for SRC buildings. This research identifies the significant factors that could be derived from building project characteristics and uncertain factors. A logical approach is employed to select the "good" regression model when the contract cost, the gross floor area and stories are known, while the numbers of change orders and rainy days are rationally estimated. Necessary diagnostics are adopted to examine the aptness of the model before inference. The cross-validation is used to test the appropriateness of the variables selected and magnitudes of the regression coefficients. The MSPR is also selected to measure the predictive ability of the model proposed, with result showing that the adequately predictive ability of the model. Furthermore, additional 11 newly finished cases are taken to test the predictive accuracy of the model individually, and the result shows that the actually necessary construction duration is considerably closed to the predictive duration. According to our forgoing derived process, it is sufficiently easy for clients to determine a suitable and applicable contract duration of SRC building. It can provide contractors an objective basis for assessing the completion duration to decide what policy to implement for the SRC construction project. This model also can facilitate a rapid appraisal of design change and weather factors on the timely performance of building projects. In other words, it could allow clients and contractors to pre-determine some arrangement for alleviating the influence of external and internal uncertainty. Overall, it is concluded that the process of development is a pragmatic approach, and the prediction model is an indispensable, fast, cost-efficient, and relatively easy forecasting tool to be utilized in practical construction management.

Acknowledgments

References

Aibinu, A. A.; Odeyinka, H. A. 2006. Construction Delays and Their Causative Factor in Nigeria, *Journal of Construction Engineering and Management* ASCE 132(7): 667–677. doi:10.1061/(ASCE)0733-9364(2006)132:7(667)

Al-Khalil, M. I.; Al-Ghafly, M. A. 1999. Important Causes of Delay in Public Utility Projects in Saudi Arabia, *Construction Management and Economics* 17(5): 647–655. doi:10.1080/014461999371259

Blyth, K.; Lewis, J.; Kaka, A. 2004. Predicting Project and Activity Duration for Buildings in the UK, *Journal of Construction Research* 5(2): 329–347. doi:10.1142/S1609945104000097

Bromilow, F. J. 1969. Contract Time Performance – Expectations and Reality, *Building Forum* 1(3): 70–80.

Chan, D. W. M.; Kumaraswamy, M. M. 1996. An Evaluation of Construction Time Performance in the Building Industry, *Building and Environment* 31(6): 569–578. doi:10.1016/0360-1323(96)00031-5

Chan, D. W. M.; Kumaraswamy, M. M. 1997. A Comparative Study of Causes of Time Overruns in Hong Kong Construction Projects, *International Journal of Project Management* 15(1): 55–63. doi:10.1016/S0263-7863(96)00039-7

Chan, D. W. M.; Kumaraswamy, M. M. 1999a. Forecasting Construction Duration for Public Housing Projects: a Hong Kong Perspective, *Building and Environment* 34(5): 633–646. doi:10.1016/S0360-1323(98)00040-7

Chan, D. W. M.; Kumaraswamy, M. M. 1999b. Modelling and Predicting Construction Durations in Hong Kong Public Housing, *Construction Management and Economics* (17): 351–362. doi:10.1080/014461999371556

Dissanayaka, S. M.; Kumaraswamy, M. M. 1999. Comparing Contributors to Time and Cost Performance in Building Projects, *Building and Environment* 34(1): 31–42. doi:10.1016/S0360-1323(97)00068-1

Enshassi, A.; Mohamed, S.; Abushaban, S. 2009. Factors Affecting the Performance of Construction Projects in the Gaza Strip, *Journal of Civil Engineering and Management* 15(3): 269–280. doi:10.3846/139237302009.15.269-280

Iyer, K. C.; Jha, K. N. 2006. Critical Factors Affecting Schedule Performance: Evidence from Indian Construction Projects, *Journal of Construction Engineering and Management* ASCE 132(8): 871–881. doi /10.1061/(ASCE)0733-9364(2006)132:8(871)

Kaka, A.; Price, A. D. F. 1991. Relationship Between Value and Duration of Construction Projects, *Construction Management and Economics* 9(4): 383–400. doi:10.1080/01446199100000030

Khosrowshahi, F.; Kaka, A. P. 1996. Estimation of Project Total Cost and Duration for Housing Project in UK, *Building and Environment* 31(4): 373–383. doi:10.1016/0360-1323(96)00003-0

Kumaraswamy, M. M.; Chan, D. W. M. 1995. Determinants of Construction Duration, *Construction Management and Economics* 13(3): 209–217. doi:10.1080/01446199500000025

Love, P. E. D.; Tse, R. Y. C.; Edwards, D. J. 2005. Time-Cost Relationships in Australian Building Construction Projects, *Journal of Construction Engineering and Management* ASCE 131(2): 187–194. doi /10.1061/(ASCE)0733-9364(2005)131:2(187)

Neter, J.; Kutner, M. H.; Nachtsheim, C. J.; Wasserman, W. 1996. *Applied Linear Regression Models*. 3rd ed. Irwin, Burr Ridge, Illinois. 435 p.

Ng, S. T.; Mak, M. M. Y.; Skitmore, R. M.; Lam, K. C.; Varnam, M. 2001. The Predictive Ability of Bromilow's Timecost Model, *Construction Management and Economics* 19(2): 165–173. doi:10.1080/01446190150505090

Nkado, R. N. 1992. Construction Time Information System for the Building Industry, *Construction Management and Economics* 10(6): 489–509. doi:10.1080/01446199200000048

Nkado, R. N. 1995. Construction Time-Influencing Factors: the Contractor's Perspective, *Construction Management and Economics* 13(1): 81–89. doi:10.1080/01446199500000009

Odusami, K. T.; Olusanya, O. O. 2000. Client's Contribution to Delays on Building Projects, *The Quantity Surveyor* 30 (January/March): 30–34.

Ogunsemi, D. R.; Jagboro, G. O. 2006. Time-Cost Model for Building Projects in Nigeria; *Construction Management and Economics* 24(3): 253–258. doi /10.1080/01446190500521041

Siegel, A. F. 2003. *Practical Business Statistics*. 5th ed. McGraw – Hill. 513 p.

Walker, D. H. T. 1995. An Investigation into Construction Time Performance, *Construction Management and Economics* 13(3): 263–274. doi:10.1080/01446199500000030

Wooldridge, J. 2002. *Introductory Econometrics: A Modern Approach*. 2nd ed. South-Western of Thomson, Mason, Ohio. 896 p.

Zavadskas, E. K.; Turskis, Z.; Tamošaitienė, J. 2010. Risk Assessment of Construction Projects, *Journal of Civil Engineering and Management* 16(1): 33–46. doi:10.3846/1392-3730.2010.16.33-46

STATYBŲ PROJEKTO REALIZAVIMO TRUKMĖS MODELIO, PAGRĮSTO ISTORINIAIS DUOMENIMIS, KŪRIMAS

M.-C. Lin, H. P. Tserng, S.-P. Ho, D.-L. Young

S a n t r a u k a

Vėlavimas realizuojant stambius statybų projektus tebėra dažna problema. Ši situacija ypač rimta dirbant su gelžbetoninių statybų projektais, kuriais populiarinama naujų konstrukcijų sistema Taivanio statybų sektoriuje. Šiuo tyrimu siekiama sukurti tinkamą sutarties trukmės modelį, pagrįstą keliais gelžbetoninės statybos atvejais. Atrenkant ir užtikrinant, kad būtų sudarytas „geras" regresijos modelis, kai projekto savybės žinomos, o išoriniai neapibrėžtumai pakankamai įvertinti, taikytas loginis metodas. Prieš darant išvadas modelio tinkamumas buvo išnagrinėtas naudojant skirtingas būtinos diagnostikos priemones. Naudojant kryžminį patikrinimą pagrindžiamas pasirinktų kintamųjų tinkamumas ir regresijos koeficientų vertės. Siekiant įvertinti siūlomo modelio tinkamumą prognozuoti, apskaičiuotos vidutinės kvadratinės prognozavimo paklaidos (angl. *Mean of the Square Prediction Errors, MSPR*). Rezultatas rodo, kad pasirinktasis regresijos modelis prognozuoja gana gerai. Pagaliau pasirinkus kelis atvejus išbandomas siūlomo modelio prognozių tikslumas. Rezultatas rodo, kad faktinė būtina statybų trukmė gana artima modelio prognozuojamai trukmei. Daroma išvada, kad siūlomą trukmės prognozių modelį galima taikyti gelžbetoninės statybos projektuose, o jo rezultatai bus gana patikimi.

Reikšminiai žodžiai: statybų projektas, statybos, trukmės prognozės, regresija, heteroskedastiškumas, pagrindimas.

Ming-Chiao LIN. Ph.D. candidate of the Department of Civil Engineering of National Taiwan University. He also is the director of the Department of Construction of Taiwan Tzu Chi Foundation. His research interests include techniques of project planning and control, construction planning and scheduling, grey prediction theory and value management.

Hui Ping TSERNG. Full professor at the Department of Civil Engineering of National Taiwan University. He also is corresponding member of Russian Academy of Engineering. He has a PhD in Construction Engineering and Management and he is official reviewer or editorial of board member of several international journals. His research interests include advanced techniques for knowledge management, construction project management, management information system, GPS/Wireless Sensor Network, and automation in construction.

Shih-Ping HO. Associate Professor of Construction Management at National Taiwan University. He taught at Stanford University in 2010 as endowed Shimizu Visiting Associate Professor. He is on the Editorial Board of Engineering Project Organization Journal. His research focuses on game theory modeling and analysis, the internationalization of A/E/C firms, the governance of Public-Private Partnerships (*PPPs*), strategic management, and knowledge sharing.

Der-Liang YOUNG. Life-time distinguished professor at the Department of Civil Engineering of National Taiwan University. He is the 54th academic award winner of the MOE of the Republic of China. He has a PhD in Hydraulics/Hydrology. He is currently the editors-in-chief of the Journal of Structural Longevity. His research interests include computational fluid mechanics, computational hydraulics and hydrology, meshless numerical methods and finite element analysis.