ESTABLISHING ENGINEERING S-CURVES TO EVALUATE SUPERVISION ENGINEER ALLOCATIONS FOR HIGHWAY CONSTRUCTION PROJECTS

Shih-Hsu WANG[a], Wei-Chih WANG[b], Pei-Yuan HSU[b], Ci-Hao CHEN[b], Kun-Chi WANG[b]

[a]Department of Civil Engineering, R.O.C. Military Academy, No. 1, Wei-Wu Road,
Fengshan District, Kaohsiung 830, Taiwan
[b]Department of Civil Engineering, National Chiao Tung University, 1001, University Road,
Hsin-Chu 300, Taiwan

Abstract. An engineering consultant firm needs to allocate engineers to supervise a highway construction project in each month during the construction phase. Properly assigning the supervision engineers under a cost-plus-fixed-fee contract has been a key factor affecting the profitability of the firm and the quality assurance of the project. Assigning too many engineers will be a waste, while allocating too few engineers may harm the supervision quality. This work proposes a two-stage model to develop engineering S-curves (called ES-curves) for planning and controlling the engineering supervision schedule. In the planning stage, a predictive ES-curve model is established based on historical ES-curves. In the controlling stage, an ES-curve is built according to the relationships between the engineering progress and construction progress. A cluster analysis and regression analysis are applied to the model development. A case study demonstrates that the produced ES-curves can help management in planning and evaluating when to increase or decrease the number of supervision engineers assigned to a project.

Keywords: S-curve, supervision engineers, resource allocation management, engineering consultant firms, cluster analysis, regression analysis.

Introduction

The Taiwanese government has spent an average of approximately $670 million US dollars annually on numerous highway construction projects during the period 2003 to 2012 (TANEEB 2012). The construction qualities of these projects greatly influence the country's economic development and the comfort level of taxpayers. The project client or owner (called the Taiwan Area National Expressway Engineering Bureau, or TANEEB) has adopted a form of cost-plus-fixed-fee contract to manage the entrusted supervision companies (i.e. engineering consulting firms) to appropriately supervise the quality of construction work performed by of various contractors.

Although TANEEB views the contract as a cost-plus-fixed-fee one, an upper limit of the total supervision manpower (i.e. total man-months) is determined at the outset. Unless the contractual work is altered (e.g. additional construction work caused by change orders increases the amount of supervision work), this upper limit of total man-months remains. Under this limit, an entrusted consulting firm is reimbursed according to the number of engineers (i.e. man-months) allocated.

While assigning too many engineers is inefficient and likely to exceed the upper limit, assigning too few engineers makes it impossible for them to monitor the overall quality of the construction work effectively (Chen 2011). Preferably, the consulting firm should consume the total supervision man-months (in order to charge all of the contractual costs) immediately when the whole construction work is completed. Thus, properly planning the required consulting engineers to supervise a highway construction project throughout the construction phase greatly affects the profitability of the engineering consultant firms.

In practice, prior to the construction, an engineering firm must develop an engineer-allocation schedule to plan when and how many engineering man-months will be spent on the project. This schedule is mainly designed according to the practitioners' experience. During the construction phase, the problem is how to adjust the planned engineering schedule to respond to current construction progress. So far, no practical tools or academic models have been devised to deal with this problem.

This work proposes a two-stage model to develop engineering S-curves (called ES-curves) for planning

Corresponding author: Wei–Chih Wang
E-mail: weichih@mail.nctu.edu.tw

Taylor & Francis
Taylor & Francis Group

and controlling the engineering supervision schedule. In the planning stage, a predictive ES-curve model is established based on historical ES-curves. In the controlling stage, an ES-curve is built according to the relationships between the engineering progress and construction progress. A cluster analysis and regression analysis are applied to the model development.

1. Past research

1.1. Consulting fees

Several studies have investigated the reasonableness of consulting fees (Bubshait *et al.* 1998; Feldmann *et al.* 2008). To obtain a high quality of technical service, a technical service provider (consulting firm) must be compensated with a fair service fee; a charge concept acceptable to most proprietors or project clients (Hoxley 2000; Ling 2004). However, many studies have indicated a significant disparity between the service fee determined by the project clients and the cost deemed reasonable by consulting firms (Flyvbjerg *et al.* 2002; Molenaar 2005; Gransberg *et al.* 2007). Researchers in the past usually recommended addressing this cost difference issue by adjusting the service fee rate (Carr, Beyor 2005; Bubshait *et al.* 1998). However, the rate is affected by a consultant's qualifications, service quality, and service scope. A project client usually will not agree to an increased service fee if charge conditions remain unchanged.

1.2. Engineering manpower allocation

Very few models have been developed specifically for dealing with the resource (engineer) allocation problems of construction supervision projects. For example, Li (2004) developed an integer programming method to assign supervision engineers to multiple building projects with an objective of minimizing supervision costs. However, applying this method in practice is not easy, because many critical assumptions must be made in advance, such as monthly workloads required (in terms of costs), the monthly workload provided by each engineer, and the percentages of time each type of engineer will be assigned to the different projects.

Other studies have focused on allocating engineering manpower to design projects (but not supervision projects). For instance, Ogunlana *et al.* (1998) devised a system dynamic (SD) model to manage the detailed design process of an engineering project. They used SD modelling to experiment with possible scenarios to examine the effectiveness of various policies (or strategies), including progress control, manpower allocation, estimation of workload, and the realization of underestimated work, in meeting different management objectives. These objectives included meeting schedules, minimizing labor days expended, and reducing the cost of design. Running the model requires detailed data, which are usually not well recorded (Ogunlana *et al.* 1998), including planned man-hours/days, planned manpower loading, total documents produced, average design productivity

(documents per man-day), amount of rework, and effort spent checking and reworking.

Moreover, Yang and Chou (2011) identified that finding a systematic way to assign consulting staff to incoming projects to ensure profitability is important to a consulting engineering company. They developed a multi-objective optimization model to assist consulting engineering firms in assigning engineering teams to incoming projects over a planning time horizon. The multiple objectives they considered included maximizing the overall profit, minimizing the variation of workloads, minimizing the maximal overtime, and maximizing the average utilization percentage. In addition, their model focused on the team level, not the engineer level.

1.3. S-curves for construction control

An S-curve is a graphical representation of the cumulative progress of a construction project, from start to finish. The S-curve is normalized according to two basic parameters, namely the relevant quantity total (e.g. project costs) and the duration of the project. The vertical scale shows the cumulative project progress in dollars or as percentage completed. The horizontal scale shows time (Cioffi 2005; Chao, Chien 2009). The shape of the S-curve usually has a smaller slope at the beginning and near the end and a larger slope in the middle, indicating that progress is slow in mobilization and demobilization periods, but faster when the bulk of the work takes place (Chao, Chien 2009).

S-curves have been widely used in project controls of construction projects and are valuable in reporting the status and predicting the future of projects (Miskawi 1989). For example, in application schedule control, a planned S-curve is used as the target against the actual progress at any point in time during the implementation phase of a construction project to evaluate whether the project is on schedule (Cioffi 2005).

Several management tools have been proposed based on the concept of an S-curve. For instance, earned value techniques use S-curve-based parameters to derive a management index to support project costs and schedule control (Fleming, Koppelman 2005; Cioffi 2006; Warburton 2011). Cheng *et al.* (2011) developed a mathematical model to forecast the short-interval (i.e. four-month) actual S-curve of expenditure from a general contractor's point of view. By considering project cash flows in activity networks with fuzzy durations and costs, Maravas and Pantouvakis (2012) developed project optimistic and pessimistic S-surfaces to assist in risk management.

So far, all previous S-curve models have focused on construction progress, but no models have been designed to manage the engineering progress of construction supervision projects (Chen 2011).

2. Analysis of current practice in Taiwan

When the design of a highway project is almost finished, an engineering consulting company is selected by the

project client (TANEEB) to supervise the incoming construction work comprising a mixture of many roads, bridges, and tunnels. If the volume of construction work is high, it is usually divided into several contracts and executed by various contractors. In that case, the engineering firm is responsible for a mega supervision project that supervises the overall construction work conducted by multiple contractors.

The TANEEB applies a cost-plus-fixed-fee method to a supervision contract to ensure that the entrusted consulting company will allocate sufficient onsite engineers to satisfactorily supervise the jobs. Under such a contract, the consulting company must first submit an engineer allocation plan/schedule that indicates the types and number of supervision managers/engineers assigned to the project. To do so, the consulting firm will usually devise a supervision organization of the project.

Figure 1 displays an example of the organization of a mega supervision project. In this example, the project has a project manager, a supervisor, and a vice supervisor in the project section. The organization also has four functional sections (contract, Q/A, surveying, and safety sections) and four supervision sections (to supervise the four construction contractors). Table 1 presents the engineer allocation schedule for this example project. Table 1 also indicates that this example project has a total of 1,277 man-months with a duration of 57 months of project supervision (from May 2007 to February 2012). Notably, several factors influence the development of this engineering allocation schedule, including project size, project complexity, construction interfaces, construction specification requirements, and site conditions. However, this schedule is devised mainly based on practitioner experience.

Next, the TANEEB needs to review and approve the submitted schedule before it becomes an as-planned engineer allocation schedule. This as-planned schedule specifies the sum of the man-months (SM) for all managers and engineers in each month (e.g. the "sub-total man-months" at the bottom of Table 1), the cumulative man-months in each month, and the total man-months (TM) for the whole project. Notably, the approved TM is the aforementioned upper limit.

As construction proceeds, the consulting firm uses this as-planned schedule as a target schedule to assign engineers. They will avoid allocating more engineers than specified in the as-planned sub-total man-months for each month and the total man-months for the whole project, unless absolutely necessary. The TANEEB also uses this schedule to monitor how the consultant allocates man-months to the project.

The consulting firm may need to update this as-planned engineer allocation schedule according to the construction progress. For instance, if construction is delayed, they should hold off on assigning additional engineers until the construction progress has caught up with the schedule. A revised schedule also needs to be reviewed and approved by the TANEEB.

Generally, allocating too many engineers in a situation of slow or delayed construction progress will be a waste of engineers because the man-months may be insufficient to supervise the rest of the project. Conversely, assigning too few engineers may harm the supervision quality because certain supervision work may be ignored. While current practice is experience based, a tool that can help generate or update the engineering schedule should be useful in supporting supervision management when allocating engineers.
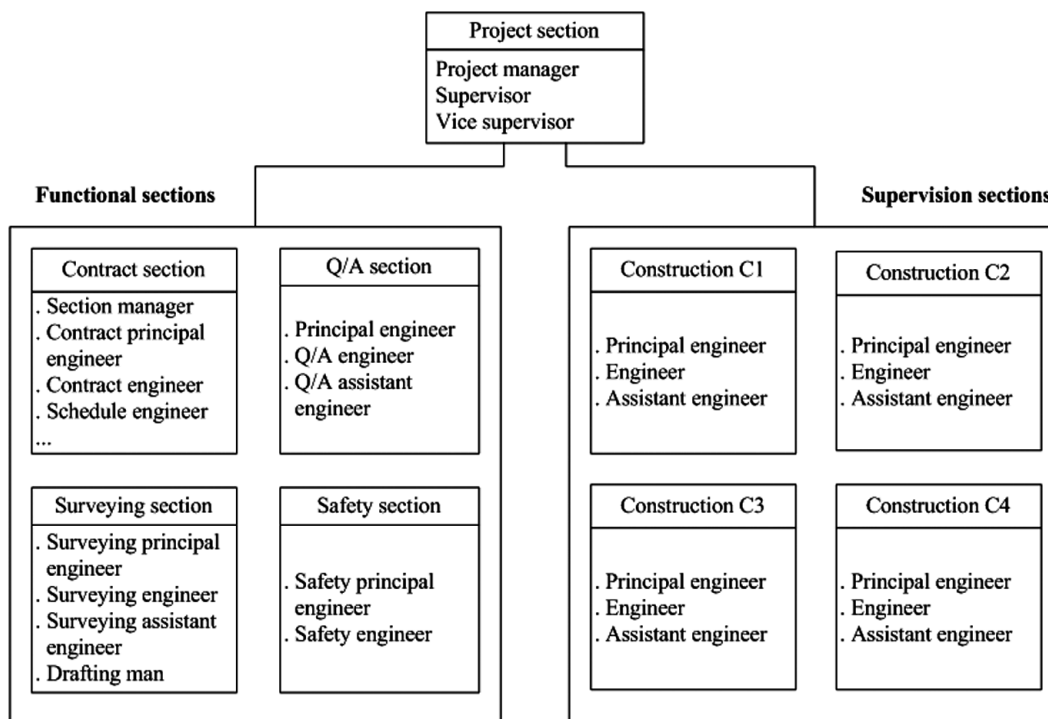

Fig. 1. Example of the organization of a mega supervision project

Table 1. Example of engineer allocation schedule for a mega supervision project

Supervision engineers	Persons	2007 5	6	7	8	9	10	11	12	⋯	2011 9	10	11	12	2012 1	2	Man-months
Project section	2.5																
Project manager	0.5		0.5	1	1	0.5	0.5	0.5	0.5	⋯	0.5	0.5	0.5	0.5	0.5	0.5	29.5
Supervisor	1		0.5	1	1	1	1	1	1	⋯	1	1	1				53.5
Vice supervisor	1			1	1	1	1	1	1	⋯							50.0
Contract section	3.5																
Section manager	0.5		0.5	1	1	0.5	0.5	0.5	0.5	⋯	0.5	0.5	0.5	0.5	0.5	0.5	29.5
Contract/schedule engineer	1			1	1	1	1	1	1	⋯	1	1					52.0
Contract principal engineer	1		0.5	1	1	1	1	1	1	⋯	1	1					52.5
Contract engineer	1		0.5	1	1	1	1	1	1	⋯	1						52.5
Surveying section	9.5																
Surveying principal engineer	1		0.5	1	1	1	1	1	1	⋯	1	1					52.5
Surveying engineer	1		0.5	1	1	1	1	1	1	⋯	1	1					52.5
Surveyor	1									⋯							17.0
Drafting man	1		0.5	1	1	1	1	1	1	⋯	1						52.5
Safety section	2																
Safety engineer	2		0.5	1	1	1	1	1	1	⋯	1	1					71.5
Q/A section	5																
Principal engineer	1		0.5	1	1	1	1	1	1	⋯	1	1					53.5
Q/A engineer	2			1	1	1	1	1	1	⋯	1	1					86.0
Q/A assistant engr.	2			1	1	1	1	2	2	⋯							76.0
Construction C1	4																
Principal engineer	1			1	1	1	1	1	1	⋯	1	1					42.0
Engineer	1						1	1	1	⋯	1	1					37.0
Assistant engineer	2									⋯							63.0
Construction C2	4																
Principal engineer	1					1	1	1	1	⋯	1	1	1	1	1	1	53.0
Engineer	2					1	1			⋯	1	1	1	1	1	1	76.0
Assistant engineer	1					1				⋯	1						45.0
Construction C3	5																
Principal engineer	1							1	1	⋯			1	1			39.0
Engineer	2									⋯							61.0
Assistant engineer	2									⋯							48.0
⋯																	
Sub-total man-months	0	0	5.0	14.1	15.1	14.2	16.2	18.3	18.3	⋯	14.2	13.1	7.1	3.0	3.0	3.0	1277
Cumulative man-months	0	0	5.0	19.1	34.2	48.4	64.6	82.9	101.2	⋯	1247.8	1260.9	1268	1271	1274	1277	1277

3. Proposed model

Based on the as-planned engineer allocation schedule shown in Table 1, this work proposes to establish an engineering S-curve (ES-curve), similar to construction S-curves, to support allocating supervision engineers. In this ES-curve, the horizontal axis displays the duration of supervision as a percentage and the vertical axis shows the cumulative progress of allocated man-months as a percentage.

The proposed model is divided into two stages, namely, a planning and a controlling stage, depending on whether or not the construction schedules are available. In the planning stage, the construction schedules are unavailable and the proposed model applies a regression analysis to derive an ES-curve directly from historical ES-curves. In the controlling stage, the construction project schedules are available. The proposed model uses a cluster analysis to categorize the progress differences (between the engineering progress and construction progress), applies a regression analysis to generate predictive equations for the progress differences, and then develops an ES-curve.

3.1. Development of an ES-curve model in the planning stage

In the planning stage, the approved ES-curves of six historical highway supervision projects are used. To build the ES-curve of a project, the percentages of the supervision duration (the x-axis) and their corresponding cumulative progress percentages (the y-axis) are processed. For instance, Table 2 displays the calculations for the percentages of the supervision duration (the x-axis) and their corresponding cumulative progress percentages (the y-axis) in each month for the same example shown in Table 1. Consider the second month (July 2007) as an illustration. The supervision duration percentage is 3.51% (=2/57), while the corresponding cumulative progress percentage is 1.5% (=19.1/1,277).

As a result, the ES-curve of the project can be plotted in an x-y plan. The ES-curves of the other five historical projects were built in a similar way.

Using the software SPSS, various equation types (such as Sigmoid, Logistic, Weibull, and Gompertz) are tested to fit these six historical data shown in the same x-y plan. The Sigmoid equation type with four parameters is identified as having the R-square value of 0.9946, indicating that this nonlinear equation could explain a high percentage of the variance in the dependent variable (i.e. cumulative engineering progress) (Ratkowsky 1990; Draper, Smith 1998).

Figure 2 displays the derived regression line of the ES-curve in the planning stage. That is, in the planning stage, the value of the cumulative engineering progress, y_x (a percentage), at supervision time x (in percentage) is calculated as follows:

$$y_x = -0.126 + \frac{1.2911}{1 + e^{-\left(\frac{x - 0.5352}{0.2277}\right)}}. \qquad (1)$$
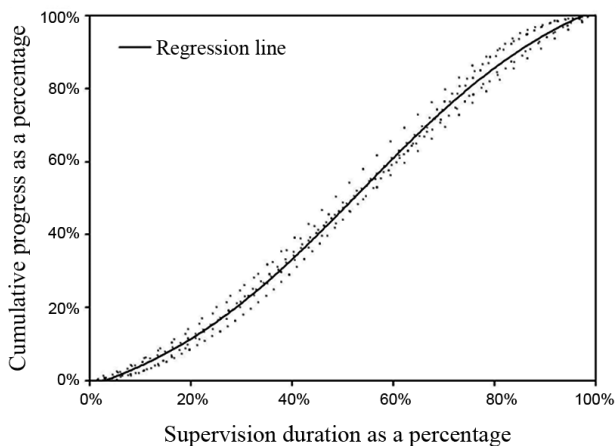

Fig. 2. Derived regression line of ES-curve in the planning stage

Table 2. Data of the supervision duration and cumulative progress for the same example project shown in Table 1

Year		2007							...	2011				2012	
Month	5	6	7	8	9	10	11	12	...	9	10	11	12	1	2
Number of month	0	1	2	3	4	5	6	7	...	52	53	54	55	56	57
Supervision duration as a percentage (x-axis)	0	1.75% (1/57)	3.51% (2/57)	5.26% (3/57)	7.02% (4/57)	8.77% (5/57)	10.53% (6/57)	12.28% (7/57)	...	91.23% (52/57)	92.98% (53/57)	94.74% (54/57)	96.49% (55/57)	98.25% (56/57)	100% (57/57)
Sub-total man months		5.0	14.1	15.1	14.2	16.2	18.3	18.3	...	14.2	13..1	7.1	3.0	3.0	3.0
Cumulative subtotal man-months	0	5.0	19.1	34.2	48.4	64.6	82.9	101.2	...	1,247.8	1,260.9	1,268	1,271	1,274	1,277
Cumulative progress as a percentage (y-axis)	0	0.4%	1.5%	2.7%	3.8%	5.1%	6.5%	7.9%		97.7%	98.7%	99.3%	99.5%	99.8%	100%

3.2. Development of the ES-curve model in the controlling stage

Since supervision engineers are needed according to the progress of the construction, the engineering progress should highly relate to the construction progress during any point in the construction period. In the controlling stage, given that the construction schedules are available, the steps to develop an ES-curve are as follows:

1. Step 1: Establish the engineering and construction S-curves for each historical project.
2. Step 2: Calculate the progress differences between the engineering and construction S-curves for each historical project.
3. Step 3: Use cluster analysis to categorize the progress differences for each historical project.
4. Step 4: Use regression analysis to derive prediction equations for the progress differences for each zone of each project.
5. Step 5: Integrate the regression results across all historical projects.

Next, the ES-curve developed from steps 1~5 can be applied to estimate the engineering progress of a new project.

4. Preparation of historical projects

Four historical mega supervision projects (i.e. project A, B, C, and D) comprising 18 highway construction projects are used to develop the controlling-stage ES-curve model. The left of Table 3 shows the project number for each construction project. For example, mega supervision project A contains five construction projects, namely, A1, A2, …, and A5. Project A has a total of 4,171 man-months.

As indicated earlier, the organization of a mega supervision project includes two parts, namely, a functional section and a field supervision section. Before developing the ES-curve model, the total man-months (MMs) for each mega project must be processed to obtain the man-months allocated to construction project i (denoted as z_i). The value of z_i is calculated as follows:

$$z_i = (m \times p_i) + n_i; \qquad (2)$$

$$p_i = q_i / Tot, \qquad (3)$$

where q_i is the (contractual) construction cost of project i; p_i is the construction cost of project i as a percentage of the total construction cost (denoted as Tot) for all construction projects under the mega project; and m is the total contractual MMs allocated to the function sections across all projects. The value of $m \times p_i$ represents the MMs allocated to the function section of project i, assuming that m is shared according to the scale construction cost, and n_i is the MMs assigned to the supervision section of project i. Then, the MMs allocated to project i (i.e. z_i) is equal to the sum of the MMs allocated to the function section (i.e. $m \times p_i$) and the supervision section (n_i).

Table 3. Man-months allocated to each construction project of four mega supervision projects

Construction project No.	Construction cost ($) (q)	% of construction cost (p = q/Tot)	Man-months		Man-months (z = m×p+n)
			Functional (m)	Supervision (n)	
A1	4,239,333,000	23.63%		385	909.59
A2	3,183,470,000	17.74%		390	783.83
A3	4,499,413,979	25.08%	2,220	437	993.78
A4	3,099,972,342	17.28%		345	728.62
A5	2,920,587,218	16.28%		394	755.18
Total cost (*Tot*)	17,942,776,539	100%		4,171	4,171
B1	2,421,889,000	21.47%		234.5	504.38
B2	2,186,858,358	19.39%		244	487.73
B3	2,982,150,000	26.44%	1,257	249	581.35
B4	1,986,942,000	17.61%		225.5	446.86
B5	1,702,609,000	15.09%		218	407.68
Total cost	11,280,448,358	100%		2,428	2,428
C1	1,266,138,000	15.06%		141.56	277.11
C2	2,107,597,000	25.07%		156.42	382.07
C3	2,268,396,000	26.99%	900.06	207.56	450.49
C4	2,763,462,000	32.88%		173.4	469.34
Total cost	8,405,593,000	100%		1,579	1,579
D1	2,820,133,814	32.81%		187.45	485.6
D2	1,374,071,000	15.98%		177.17	322.38
D3	3,300,945,000	38.40%	908.71	164.46	513.4
D4	1,100,908,680	12.81%		97.21	213.62
Total cost	8,596,058,494	100%		1,535	1,535

Table 3 summarizes the calculated MMs allocated to each construction project. For example, mega project A has 4,171 man-months in total. These MMs are divided into 909.59, 783.83, 993.78, 728.62, and 755.18 MMs to supervise the five construction projects A1, A2, A3, A4, and A5, respectively. The calculation details for these MMs can be found in Chen (2011).

After calculating the total MMs for each construction project i, the MMs of the jth month for the project (denoted as $MM_{i(j)}$) is calculated as follows:

$$MM_{i(j)} = (m_j \times p_i) + n_{i(j)}, \qquad (4)$$

where m_i is the total MMs allocated to the function sections in the jth month, and $n_{i(j)}$ is the MMs assigned to the supervision section of project i in the jth month.

4.1. Step 1: Establish the engineering and construction S-curves for each historical project

After using Eqn (4) to calculate the MMs for each month of project i (that is, $MM_{i(j)}$), the cumulative MMs as a percentage of each month (and the supervision time as a percentage) can then be obtained. Consequently, an ES-curve for each construction project i can be established. In addition, a construction S-curve is obtained in the conventional way. Thus, the ES-curve and construction S-curve can be plotted together for each construction project. For example, Figure 3 shows the engineering and construction S-curves of project A1.

4.2. Step 2: Calculate the progress differences between the engineering and construction S-curves for each historical project

This step collects data of the progress differences between the engineering and construction progress at any point during supervision. The difference between the two progresses in the jth month for project i (denoted as $\Delta y_{i(j)}$) is calculated as follows:

$$\Delta y_{i(j)} = eng_{i(j)} - con_{i(j)}, \qquad (5)$$

where $eng_{i(j)}$ and $con_{i(j)}$ are the cumulative progresses (as percentages) of the engineering and construction progress in the jth month of project i.
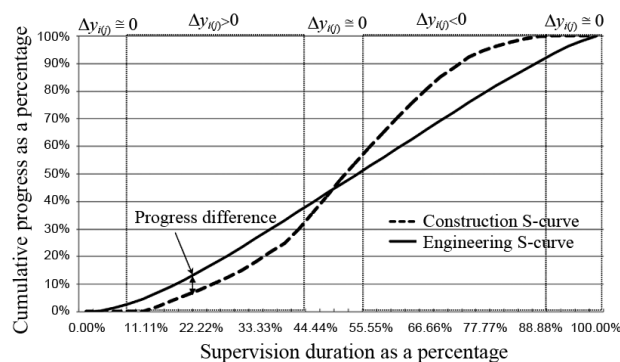
Fig. 3. Example of the engineering and construction S-curves for project A1

Take project A1 as an example (see Fig. 3). The total supervision duration is 36 months. When the supervision duration is in the 8th month (i.e. 22.22% of the 36 months), the values of the cumulative engineering progress ($eng_{i(j)}$) and cumulative construction progress ($con_{i(j)}$) are 14.53% and 7.98%, respectively. Thus, the progress difference ($\Delta y_{i(j)}$) is 6.55% (= 14.53 – 7.98%). A positive value of $\Delta y_{i(j)}$ indicates that the engineering progress is ahead of the construction progress, and vice versa. Figure 3 also shows that is positive in the first half part of the project, but negative in the latter half of the project.

4.3. Step 3: Use a cluster analysis to categorize the progress differences for each historical project

Instead of using a single regression equation to predict the progress differences for the overall project supervision, this study develops five regression equations with respect to the five duration zones of project supervision. For example, in Figure 3, the progress differences can easily be classified into five zones, namely, $\Delta y_{i(j)} > 0$ (in the front half), $\Delta y_{i(j)} \cong 0$ (at the beginning, in the middle, and at the end), and $\Delta y_{i(j)} < 0$ (in the latter half). The steps to conduct the cluster analysis are as follows:

1. Step 3.1: Input the progress difference data, $\Delta y_{i(j)}$, (from the first month to the last month of the project supervision) for each historical project. For example, the construction of project A1 lasted 36 months. The progress difference data are 0.21%, 1.47%, …, −1.82%, and 0.00% from the first month to the 36th month, respectively.

2. Step 3.2: Apply hierarchical cluster and K-means cluster analyses to classify the progress difference data ($\Delta y_{i(j)}$) to define the various zones of each historical project. Hierarchical cluster and K-means are two common cluster analysis methods. Similar to Milligan (1980) and Fisher *et al.* (1992), this study exploits the advantages of both methods to yield the smallest variances of the $\Delta y_{i(j)}$ values within each cluster and lead to statistically significant differences of the means of the $\Delta y_{i(j)}$ values among clusters. To do so, the appropriate number of clusters is first determined using the hierarchical cluster. In each cluster, the mean of the $\Delta y_{i(j)}$ values is generated. The means of all clusters are then used as the initial seeds for the K-means cluster analysis. Additionally the hierarchical cluster and K-means cluster analyses are performed using a statistical data analysis system, SPSS.

3. Step 3.3: Select suitable number of clusters. Researchers such as Sharma (1996) have suggested using the Root-Mean-Square Standard Deviation (RMSSTD), coefficient of determination (R^2) and Semi-Partial R-Square (SPRSQ) to determine whether the resulting number of clusters is appropriate. RMSSTD measures homogeneity within each cluster by summing the standard deviations (STDs) of the clusters. A lower RMSSTD value indicates high homogeneity in each cluster.

The value of R^2 is equal to one minus the ratio of the variance before clustering and after clustering. If the number of clusters and the number of progress difference data are the same, R^2 is equal to one. If all progress difference data are placed in only one cluster, R^2 is equal to 0. A higher R^2 value generally implies improved clustering because more progress difference data are categorized into various clusters to develop more reliable fit regression equations. However, too many clusters are difficult to manage, e.g., requiring additional duration zones for analysis. SPRSQ determines if some intermediate clusters should be combined in the clustering process by measuring the difference in R^2 when the number of clusters changes. A small SPRSQ value refers to a situation in which the combination of clusters does not increase high variances (i.e. the data in each combined cluster resemble each other); this combination is thus acceptable. Conversely, a high SPRSO value (i.e. the data in each combined cluster are highly varied) implies that current clusters should be further disaggregated. Overall, a clustering number with low RMSSTD, high R^2, and low SPRSQ is preferred. In addition, this clustering number should be as small as possible.

For example, Table 4 displays the values of RMSSTD, R^2, and SPRSQ associated with various numbers of clusters (i.e. 2~7 clusters) for project A1. In the table, it can be found that the values of RMSSTD (1.0516; too high), R^2 (0.1059; too low), and SPRSQ (0.9717; too high) evaluated for the two-cluster situation are not suitable. It is preferred that the data ($\Delta y_{i(j)}$) are grouped into more than two clusters (i.e. from three to seven clusters) because the values of RMSSTD, R^2, and SPRSQ are more suitable. Finally, since the smallest number of clusters is preferred, the progress differences are divided into three clusters as follows: $\Delta y_{i(j)} > 0$ (in the first half), $\Delta y_{i(j)} \cong 0$ (at the beginning, in the middle, and at the end), and $\Delta y_{i(j)} < 0$ (in the latter half).

4. Step: 3.4: Generate the clusters. Table 5 displays the progress differences classified per zone. For example, the progress differences located in the first three months are classified as zone 1 ($\Delta y_{i(j)} \cong 0$) for project A1, with values of 0.21%, 1.47%, and 2.96%, respectively.

Table 4. Calculated values of RMSSTD, R^2, and SPRSQ for project A1

Number of clusters	RMSSTD	R^2	SPRSQ
2	1.0516	0.1059	0.9717
3	0.3663	0.8658	0.0108
4	0.3513	0.8766	0.0249
5	0.3138	0.9015	0.0364
6	0.2492	0.9379	0.0112
7	0.2257	0.9491	0.0277

Table 5. Progress differences classified in each zone of project A1

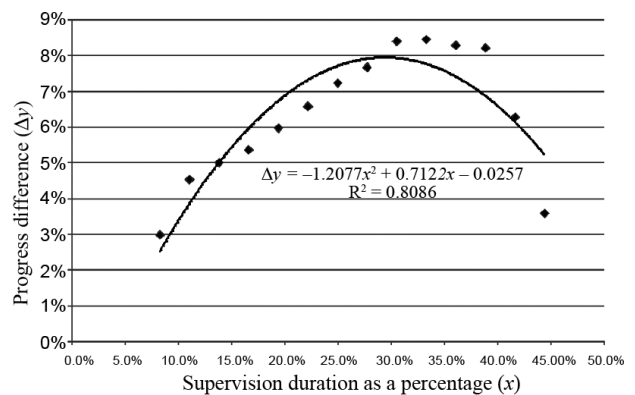| Duration zone	Description	Number of month	Progress difference
1	$\Delta y_{i(j)} \cong 0$	1, 2, 3	0.21%, 1.47%, 2.96%
2	$\Delta y_{i(j)} > 0$	4, 5, 6,7, 8, 9,10,11, 12, 13, 14, 15, 16	4.51%, 4.98%, 5.34%, 5.95%, 6.55%, 7.20%, 7.64%, 8.37%, 8.42%, 8.26%, 8.18%, 6.25%, 3.57%
3	$\Delta y_{i(j)} \cong 0$	17, 18, 19, 20	0.93%, −1.80%, −4.32%, −6.80%
4	$\Delta y_{i(j)} < 0$	21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32	−8.96%, −11.12%, −12.87%, −14.58%, −15.79%, −16.15%, −16.54%, −15.76%, −14.38%, −12.77%, −10.85%, −9.02%%
5	$\Delta y_{i(j)} \cong 0$	33, 34, 35, 36	−6.43%, −3.92%, −1.82, 0.00%

Fig. 4. Regression line for fitting the data of progress differences in zone 2 of project A1

4.4. Step 4: Use regression analysis to derive prediction equations of the progress differences for each zone of each project

After classifying the progress differences into five zones in each project, the progress difference data as a percentage (the *y*-axis) corresponding to the supervision duration as a percentage (the *x*-axis) are plotted in an *x-y* plan. A regression line can be found to best represent these data. For example, Figure 4 shows the progress difference data in zone 2 ($\Delta y_{i(j)} > 0$, from 4th month to 16th month) for project A1. A quadratic regression line with a value of R square = 0.8086 is then identified as fitting these data. The regression lines of the other four zones in project A1 can be found in a similar way, as can the zones of the other 17 projects.

Table 6. Calculated values of R^2 of the regression lines for the five zones of all 18 projects

Construction project No.	R^2				
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
A1	0.9940	0.8086	0.9999	0.9897	0.9999
A2	0.9951	0.9167	0.9921	0.9654	1.0000
A3	0.9969	0.9910	0.9998	0.8336	0.9890
A4	1.0000	0.8304	0.9964	0.9549	0.9993
A5	1.0000	0.1682	0.9886	0.9097	1.0000
B1	0.9999	0.9596	1.0000	0.8290	0.9994
B2	0.9976	0.9181	0.9984	0.7837	0.9998
B3	1.0000	0.9794	0.9988	0.9032	0.9991
B4	0.9989	0.9710	0.9710	0.6911	0.9994
B5	0.9995	0.9236	0.9969	0.8971	0.9997
C1	0.9999	0.9818	0.9994	0.9416	0.9983
C2	1.0000	0.6957	0.9987	0.8894	1.0000
C3	0.9835	0.7078	0.9934	0.9590	0.9998
C4	0.9998	0.8631	0.9342	0.7972	1.0000
D1	0.9976	0.8801	0.9841	0.8696	1.0000
D2	0.9991	0.3494	0.9612	0.8844	0.9994
D3	0.9993	0.7892	0.9994	0.9541	0.9999
D4	1.0000	0.8693	0.9986	0.9142	0.9987
Average R^2	0.9978	0.8113	0.9895	0.8871	0.9990

Table 7. Constants of the regression equations for zone 2 of all 18 projects

Construction project No.	Constants			R^2
	$a_{k(i)}$	$b_{k(i)}$	$c_{k(i)}$	
A1	−1.2077	0.7122	−0.0257	0.8086
A2	−1.1096	0.8120	−0.0130	0.9167
A3	−2.2352	1.1527	−0.0547	0.9910
A4	−2.4885	1.3089	−0.0747	0.8304
A5	−0.7329	0.2948	0.0187	0.1682
B1	−2.5978	1.3967	−0.0699	0.9596
B2	−3.8700	1.9229	−0.1019	0.9181
B3	−2.0909	0.9783	−0.0272	0.9794
B4	−3.1150	1.6421	−0.0960	0.9710
B5	−2.5289	1.3666	−0.0664	0.9236
C1	−1.8816	0.9465	−0.0172	0.9818
C2	−1.2827	0.5236	0.0212	0.6957
C3	−1.5664	0.9993	−0.0495	0.7078
C4	−0.7123	0.4805	−0.0032	0.8631
D1	−2.3461	1.3325	−0.0353	0.8801
D2	−0.5679	0.3238	0.0324	0.3494
D3	−0.9605	0.4932	0.0325	0.7892
D4	−1.5562	0.6345	0.0183	0.8693
Average	−1.8250	0.9623	−0.0284	0.8113

Table 6 lists the calculated values of R^2 of the regression lines for the five zones of all 18 projects. The average values of R^2 of the regression lines are 0.9978, 0.8113, 0.9895, 0.8871, and 0.9990 for zone 1 through zone 5, respectively.

4.5. Step 5: Integrate the regression results across all historical projects

After performing Steps 1~4, the regression lines of the five zones of the 18 historical projects are developed. That is, 18 regression lines are identified for each zone. To derive a generalized regression equation for each zone k ($k = 1 \sim 5$), an average method is applied. That is, a predicted progress difference (as a percentage) for zone k, denoted as Δy_k, is derived as follows:

$$\Delta y_k = a_k x^2 + b_k x + c_k; \qquad (6)$$

$$a_k = (a_{k(1)} + a_{k(2)} + ... + a_{k(i)} + ... + a_{k(18)})/18; \quad (7)$$

$$b_k = (b_{k(1)} + b_{k(2)} + ... + b_{k(i)} + ... + b_{k(18)})/18; \quad (8)$$

$$c_k = (c_{k(1)} + c_{k(2)} + ... + c_{k(i)} + ... + c_{k(18)})/18, \quad (9)$$

where x is the given supervision duration as a percentage, and $a_{k(i)}$, $b_{k(i)}$, and $c_{k(i)}$ are the constants of the derived regression equation for zone k for project i.

Consider zone 2 as an example. Table 7 represents the derived constants of $a_{k(i)}$, $b_{k(i)}$, and $c_{k(i)}$ of the regression equations for zone 2 for 18 projects. Based on

Table 8. Regression equations to calculate the progress differences for each zone

Duration zone	Start time of the zone (%)	Finish time of the zone (%)	Regression equations
1	0.00%	7.46%	$\Delta y_{k=1} = 2.4316x^2 + 0.3062x - 0.0003$
2	7.46%	43.41%	$\Delta y_{k=2} = -1.8250x^2 + 0.9623x - 0.0284$
3	43.41%	57.11%	$\Delta y_{k=3} = 0.4129x^2 - 0.9969x + 0.4134$
4	57.11%	89.73%	$\Delta y_{k=4} = 2.2970x^2 - 3.4797x + 1.2056$
5	89.73%	100.00%	$\Delta y_{k=5} = -2.1319x^2 + 4.7062x - 2.5759$

Eqns (7)–(9), the constant values of a_k, b_k, and c_k ($k = 2$) are calculated as − 1.8250, 0.9623, and − 0.0284, respectively. Hence, the generalized regression equation for zone 2 is $\Delta y_{k=2} = -1.8250x^2 + 0.9623x - 0.0284$.

In addition, the start time and finish time of a particular zone are determined by averaging the start times and finish times of the zone for all 18 historical projects. Table 8 presents the derived regression equations to calculate the progress differences for the five zones.

5. Estimate the engineering progress of a new project

In the controlling stage, the derived regression equations shown in Table 8 can be applied to estimate the engineering S-curve according to the estimated construction S-curve of a new project. That is, the value of a cumulative engineering progress (y_x) at supervision time x (as a percentage) is calculated as follows:

$$y_x = con_x + \Delta y_k, \tag{10}$$

where con_x and Δy_k are the cumulative construction progress and progress difference at time x, respectively.

5.1. Integration of two-stage models

Eqn (1) predicts the ES-curve in the planning stage (zone 1), while Eqn (10) represents the ES-curve in the controlling stage (zones 2~5). Table 9 summarizes the proposed final regression equations for each zone.

5.2. Deviation measures

Deviation between the actual and developed ES-curves is measured using the value of root-mean-square error (RMSE) (Chao, Chien 2010). The formula to calculate this measurement is as follows:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(estimated_j - actual_j)^2}{n}}, \tag{11}$$

where n – number of progress measurements for a project; $estimated_j$ – estimated progress for measurement j; and $actual_j$ – actual progress for measurement j. RMSE directly measures the average deviation in percentage terms.

6. Case study

6.1. Project description

A highway mega supervision project is used to test the proposed model. This mega project (hereafter, the test

project) is responsible for supervising ten construction projects. The total contractual costs of the ten construction projects is approximately 21.4 billion New Taiwan dollars (i.e. about 766.7 million US dollars; 1 US dollar is about 30 NT dollars). Each construction project ranges from 1.5 billion to 3.3 billion NT dollars. Other information related to the supervision project includes: (1) the total supervision fee is approximately 770 million NT dollars; (2) the total supervision duration is 67 months (March 2004 ~ September 2009); and (3) the total contractual man-months of this project is about 4,800 man-months.

6.2. Planning stage

After awarding the test project to an engineering firm, the firm submitted a monthly engineer-allocation schedule according to their experience. Figure 5 displays the actual ES-curve of the engineering schedule for this test project during the planning stage. Currently, no methods can be used to evaluate the appropriateness of this submitted ES-curve. Now, based on the proposed model, a predicted ES-curve can be generated using Eqn (1) and is also plotted in Figure 5. Moreover, according to Eqn (11), the RMSE obtained is 0.023403 (2.3403%) after comparing the actual and predicted ES-curves.
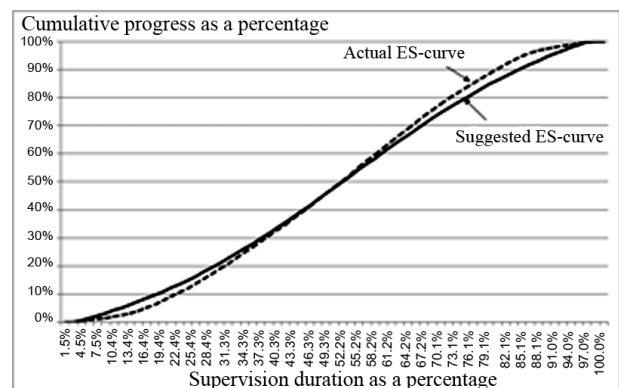

Fig. 5. Actual and suggested ES-curves of the testing project during the planning stage

Table 9. Proposed final regression equation applied in each duration zone

Duration zone	Start time of the zone (%)	Finish time of the zone (%)	Regression equations	Applied stage
1	0.00%	7.46%	$y_x = -0.126 + \dfrac{1.2911}{1 + e^{-(\frac{x-0.5352}{0.2277})}}$	Planning stage
2	7.46%	43.41%	$\Delta y_{k=2} = -1.8250\, x^2 + 0.9623\, x - 0.0284$ $y_x = con_x + \Delta y_k$	Controlling stage
3	43.41%	57.11%	$\Delta y_{k=3} = 0.4129 x^2 - 0.9969 x + 0.4134$ $y_x = con_x + \Delta y_k$	Controlling stage
4	57.11%	89.73%	$\Delta y_{k=4} = 2.2970 x^2 - 3.4797 x + 1.2056$ $y_x = con_x + \Delta y_k$	Controlling stage
5	89.73%	100.00%	$\Delta y_{k=4} = -2.1319 x^2 + 4.7062 x - 2.5759$ $y_x = con_x + \Delta y_k$	Controlling stage

6.3. Controlling stage

As the test project progresses, the ten construction projects were subsequently tendered and the construction work followed. In Figure 6, the dotted S-curve presents the actual integrated construction schedule by weighting the ten individual construction schedules based on their corresponding percentages of construction costs for the total construction costs. For example, an integrated construction progress of this actual construction schedule is the sum of the weighted construction progresses ($= p_i \times con_{i(j)}$) of the ten construction projects. Additionally, Figure 6 also plots the actual ES-curve of this test project.

At this stage, the construction schedule is known, so the second-stage model is applied. Hence, Figure 6 presents the suggested ES-curve based on the derived regression equations (Table 9). According to Eqn (11), the RMSE obtained is 0.035682 (3.5682%) after comparing the actual and suggested ES-curves.

Comparing the actual and suggested ES-curves shown in Figure 6, the major difference is that the latter allocates supervision engineers with respect to the construction progress. For instance, in approximately 43.3% of the duration of the supervision, slow progress of construction work occurred. At that time, the suggested ES-curve correspondingly slowed down the allocation of additional supervision engineers, while the actual ES-curve did not.

6.4. Discussions

After conducting the case study, several lessons are identified from discussions with a manager of the case-study firm and a section officer of the TANEEB. These lessons are summarized as follows:

1. Under a cost-plus-fixed-fee contract, engineering firms tend to consume as many of the total man-months as possible in the early stages. As a result, the engineering man-months are likely to be insufficient in the latter stages of the project when a large amount of construction work still has n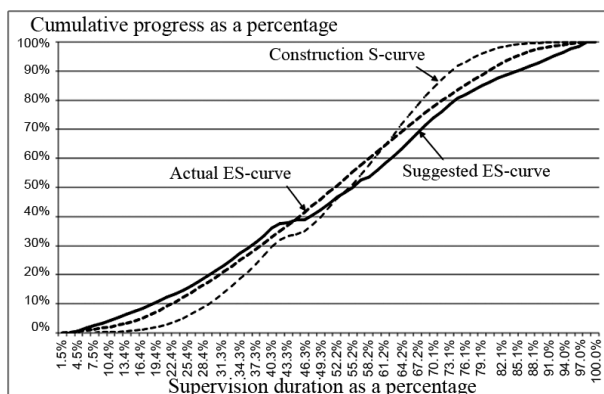ot been done. In such a case, the TANEEB will usually not grant additional cost compensations to the engineering firm, because allocating engineers appropriately is considered to be firm's responsibility. For example, in the case study, the total construction duration was actually extended and no cost compensations were granted to the engineering firm. Supposing the firm could reduce the man-months in the controlling stage, when the construction delay occurred (as suggested by the proposed model), engineering costs should not be overrun. The ES-curves generated by the proposed model should help the engineering firm improve its resource allocation decisions.

2. Currently, the TANEEB reviews the submitted engineering schedules based on experience. The proposed model can also provide an efficient reference to support the reviewing process. That is, the stage-1 ES-curve can help to review the submitted engineering schedule, while the stage-2 curve can check if the initial (as-planned) engineering schedule should be updated to reflect the construction status.

As indicated in Section 3.1, six historical supervision projects were selected by the research team for developing the ES-curve model in the planning stage. These projects were selected mainly for two reasons. First, data for engineer allocation schedules (similar to the data shown in Table 1) of these projects were thoroughly documented for analysis. Second, the approved schedules of these projects were successfully implemented as target schedules to assign supervision engineers during the planning stage. However, according to some practitioners involved in these projects, these approved schedules were occasionally unsatisfactory during the controlling stage. For instance, additional engineers over a specified target amount were not allowed when construction work was accelerating. Therefore, the proposed model is divided into two stages for developing ES-curves.

Conclusions

This study has illustrated the problem of assigning supervision engineers under a cost-plus-fixed-fee contract in a highway construction project. Assigning an appropriate number of supervision engineers is critical to the profitability of an engineering firm and the quality assurance of construction projects. The proposed two-stage model has contributed to developing engineering S-curves to resolve this engineer allocation problem for supervision projects.

There are possible areas for future research. First, the proposed model may be modified so that it applies to building projects. Notably however, the supervision work of a building project is normally contracted based on a percentage fee method. Second, the current planning-stage model is developed based on six mega projects, while the current controlling-stage model is designed according to four mega projects and 18 construction

Fig. 6. Actual and suggested ES-curves of the test project during the control stage

projects. Data from additional projects should be collected and included into the regression database. Third, as mentioned earlier, several factors influence the engineer allocations. We recommend that future research further clarify the effects of these factors and incorporate them into the development of the planning-stage model. Fourth, other optimization theories, such as genetic algorithms, may be substituted for regression analysis in a search for improved solutions.

Acknowledgements

References

Bubshait, A. A.; Al-Said, F. A.; Abolnour, M. M. 1998. Design fee versus design deficiency, *Journal of Architectural Engineering* 4(2): 44–46.
http://dx.doi.org/10.1061/(ASCE)1076–0431(1998)4:2(44)

Carr, P. G.; Beyor, P. S. 2005. Design fees, the state of the profession, and a time for corrective action, *Journal of Management in Engineering* 21(3): 110–117.
http://dx.doi.org/10.1061/(ASCE)0742–597X(2005)21:3(110)

Chao, L. C.; Chien, C. F. 2009. Estimating project S-curves using polynomial function and neural networks, *Journal of Construction Engineering and Management* 135(3): 169–177.
http://dx.doi.org/10.1061/(ASCE)0733–9364(2009)135:3(169)

Chao, L. C.; Chien, C. F. 2010. A model for updating project S-curve by using neural networks and matching progress, *Automation in Construction* 19(1): 84–91.
http://dx.doi.org/10.1016/j.autcon.2009.09.006

Chen, C. H. 2011. *Model for allocating monthly supervision manpower of engineering firms*: MS Thesis. Department of Civil Engineering, National Chiao Tung University, Taiwan.

Cheng, Y. M.; Yu, C. H.; Wang, H. T. 2011. Short-interval dynamic forecasting for actual S-curve in the construction phase, *Journal of Construction Engineering and Management* 137(11): 933–941.
http://dx.doi.org/10.1061/(ASCE)CO.1943–7862.0000358

Cioffi, D. F. 2005. A tool for managing projects: an analytic parameterization of the S-curve, *International Journal of Project Management* 23(3): 215–222.
http://dx.doi.org/10.1016/j.ijproman.2004.08.001

Cioffi, D. F. 2006. Designing project management: a scientific notation and an improved formalism for earned value calculations, *International Journal of Project Management* 24(2): 136–144.
http://dx.doi.org/10.1016/j.ijproman.2005.07.003

Draper, N. R.; Smith, H. 1998. *Applied regression analysis*. 3rd ed. John Wiley and Sons, New York. 736 p.

Fisher, D.; Xu, L.; Zard, N. 1992. Ordering effects in clustering, in *Proceedings of the Ninth International Conference on Machine Learning*, 1992, Morgan Kaufmann, San Mateo, CA, 163–168.

Feldmann, M. L.; Chrusciel, D.; Pohlmann, A.; Shelley, II, M. C.; McCool, K.; Morton, A. D.; Ahoy, C. K. 2008. Architectural and engineering fees from the public institutional perspective, *Journal of Management in Engineering* 24(1): 1–11.
http://dx.doi.org/10.1061/(ASCE)0742–597X(2008)24:1(2)

Fleming, Q. W.; Koppelman, J. M. 2005. *Earned value project management*. 3rd ed. Project Management Institute, Newtown Square, PA. 232 p.

Flyvbjerg, B.; Holm, M.; Buhl, S. 2002. Underestimating costs in public works projects: error or lie? *Journal of the American Planning Association* 68(3): 279–295.
http://dx.doi.org/10.1080/01944360208976273

Gransberg, D. D.; Puerto, C. L.; Humphrey, D. 2007. Relating cost growth from the initial estimate to design fee for transportation projects, *Journal of Construction Engineering and Management* 133(6): 404–408.
http://dx.doi.org/10.1061/(ASCE)0733–9364(2007)133:6(404)

Hoxley, M. 2000. Are competitive fee tendering and construction professional service quality mutually exclusive? *Construction Management and Economics* 18(5): 599–605.
http://dx.doi.org/10.1080/014461900407400

Li, C. Y. 2004. *Model analysis of manpower optimization assignments for construction inspection project*: MS Thesis. Department of Civil Engineering, National Central University, Taiwan.

Ling, Y. Y. 2004. Consultancy fees: dichotomy between A/Es need to maximize profit and employers' need to minimize cost, *Journal of Professional Issues in Engineering Education and Practice* 130(2): 120–123.
http://dx.doi.org/10.1061/(ASCE)1052–3928(2004)130:2(120)

Maravas, A.; Pantouvakis, J. P. 2012. Project cash flow analysis in the presence of uncertainty in activity duration and cost, *International Journal of Project Management* 30(3): 374–384.
http://dx.doi.org/10.1016/j.ijproman.2011.08.005

Milligan, G. W. 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms, *Psychometrika* 45(3): 325–342.
http://dx.doi.org/10.1007/BF02293907

Miskawi, Z. 1989. An S-curve equation for project control, *Construction Management and Economics* 7(2): 115–124.
http://dx.doi.org/10.1080/01446198900000016

Molenaar, K. R. 2005. Programmatic cost risk analysis for highway megaprojects, *Journal of Construction Engineering and Management* 131(3): 343–353.
http://dx.doi.org/10.1061/(ASCE)0733–9364(2005)131:3(343)

Ogunlana, S.; Lim, J.; Saeed, K. 1998. Desman: a dynamic model for managing civil engineering design projects, *Computers and Structures* 67(5): 401–419.
http://dx.doi.org/10.1016/S0045–7949(97)00148–X

Ratkowsky, D. A. 1990. *Handbook of nonlinear regression models*. New York: Marcel Dekker, Inc. 241 p.

Sharma, S. 1996. *Applied multivariate techniques*. New York: John Wiley and Sons. 512 p.

TANEEB. 2012. *Annual budget report*. Taiwan Area National Expressway Engineering Bureau, Ministry of Transportation and Communications, Taiwan.

Warburton, R. D. H. 2011. A time-dependent earned value model for software projects, *International Journal of Project Management* 29(8): 1082–1090.
http://dx.doi.org/10.1016/j.ijproman.2011.02.008

Yang, I. T.; Chou, J. S. 2011. Multi-objective optimization for manpower assignment in consulting engineering firms, *Applied Soft Computing* 11(1): 1183–1190.
http://dx.doi.org/10.1016/j.asoc.2010.02.016

Shih-Hsu WANG. Assistant Professor in the Department of Civil Engineering at ROC Military Academy, Taiwan. He received his PhD degree from the Department of Civil Engineering at National Chiao Tung University. His research interests include national competitiveness and construction consultant service.

Wei-Chih WANG. Professor in the Department of Civil Engineering at National Chiao Tung University, Taiwan. He is a member of the Construction Management Association of Taiwan. His research interests include construction procurement, engineering management, building information modelling, high-tech facility management, and construction simulation.

Pei-Yuan HSU. He received his MS degree from the Department of Civil Engineering at National Chiao Tung University, Taiwan. His research interests include building information modelling and structural engineering.

Ci-Hao CHEN. He received his MS degree from the Department of Civil Engineering at National Chiao Tung University, Taiwan. His research interests include engineering management and resource allocation management.

Kun-Chi WANG. PhD student in the Department of Civil Engineering at National Chiao Tung University, Taiwan. His research interests include building information modelling and construction management.