Estimation of contractor's project overhead rate as research on building cost

Li-Chung Chao
Department of Construction Engineering, National Kaohsiung First University of Science and Technology, Taiwan
(e-mail: chaolc@ccms.nkfust.edu.tw)

Abstract

The contractor's project overhead costs are the on-site related costs for production support in undertaking a project, such as those for supervision, office, utilities and services. Unlike direct costs, they are not directly connected with the performance of any particular element of a project, but are required for running the project as a whole. Despite the recommended practice of estimating the contractor's field overhead for a project as line items, the alternative method of applying a selected rate as a percentage of direct cost still is used widely. As an ongoing research on building cost, Empirical models of project overhead rate have been developed from historical data in the present study. The nature and significance of the contractor's project overhead are first explored, along with factors that have an influence on the overhead rate. The bid data for 173 building projects collected from a large construction firm in Taiwan covering a variety of work is analyzed and a classification is established whereby the projects' overhead rates range between 0.0171 and 0.2912. The data is divided into two parts for model development and model testing according to order of time. Then, two methods of estimation are compared in their modeling and prediction errors: the cluster center method of categorizing projects into 24 groups and the multiple regression method using four variables, i.e. size, duration, type of work, and location. Either method can achieve an average error of about 3% of direct cost in modeling as well as in prediction. An improvement on the subjective rate-applying method, the model may coexist with the itemized estimation method as a checking mechanism and is potentially useful to contractors as well as owners and consultants. Implications for the building industry and recommendations for future research are also discussed.

Keywords: cost estimation, project overhead, empirical model, statistics

1. Introduction

The contractor's project overhead costs are the on-site related costs for production support in undertaking a project. Unlike direct costs, they are not directly connected with the performance of any particular element of a project, but are required for running the project as a whole. Depending on the practice of categorizing costs, they generally include costs of supervision, office, utilities and services, insurance, safety etc. The sum of direct costs and project overhead costs is the contractor's project construction cost representing all expenditures internal to a project and essential for completing it according to specifications. In contrast, the project's share of the contractor's home- office overhead costs and profit (so called the bid markup) is business-oriented and external to the project, for which a higher or lower level may be charged as deemed appropriate. Since the project overhead costs often constitute a greater part of the contract price than the markup, project overhead estimation deserves no less attention than markup determination. Even the owner in preparing the budget of a project should not overlook the importance of a fair estimate of project overhead costs. Traditionally, to obtain a reasonably accurate estimate of the contractor's project overhead costs, the formal practice is to establish line items and calculate how much is needed for each item, based on a plan that meets project conditions and requirements. For example, salaries of management and supervision are estimated according to the planned field office organization and the durations of the positions. Costs of various categories of insurance and bonds are estimated individually as a percentage of the estimated bid price and relevant direct costs. Usually the estimator is assisted by a checklist and evaluates each possible item in turn. Detailed examples can be found in McCaffer and Baldwin [9], Halpin [7], and Diamant [6] etc.

However, precise definition of project overhead charges is time consuming and may not guarantee the correctness of the results. Hence, when all estimates are complete, often the ratio of project overhead to direct cost (referred to herein simply as "project overhead rate") is calculated as representing the level of project overhead and compared with those for past similar projects as a check on abnormality. On the other hand, the alternative method of applying an experience-based rate (say 10%) to cover the contractor's project overhead costs still is used widely. Some contractors and owners/ consultants even fix all the contractor's overheads and profit as a percentage (say 20%) of the estimated direct cost to arrive at the bid price or budget for a project. Naturally, such a simple method is prone to inaccuracy, as the applied rate often is selected without the support of a modeling methodology. To remedy the rate-applying method's subjectivity and to improve its accuracy, this paper presents a research that set out to develop more reliable models founded on historical data, which can be used for estimating or checking for coming projects. Establishing the relationships between project attributes and project overhead rate, two empirical methods, i.e. cluster center and linear regression, are compared about their performance in modeling and prediction, based on a contractor's past bid data. The objective is to find out what would be the best method and how accurate the best method could be according to some error measures for evaluation. Related researches and factors influencing project overhead are reviewed prior to describing the data and the models. Model results are discussed and model limitations addressed at the end.

2. Related researches

Holland and Hobson [8] found from a survey that there is a lack of consistency among construction firms concerning categorization of costs as direct costs and indirect costs, i.e. project and home-office overheads. For example, the construction manager is equally likely categorized as project overhead or home-office overhead, whereas field engineers may be categorized as direct cost as well as project overhead. However, the inconsistency of different firms does not affect the models in the present study, since they are oriented towards one firm, as long as it has its own way of categorization and categorizes costs consistently for different projects. In any case, to avoid confusion, the firm in provision of data for model development has to be specific as to what constitutes project overhead costs to it.

Determination of the bid markup rate to be applied on top of the estimated project construction cost has attracted much research interest over the past decades. For example, Ahmad and Minkarah [1], Chua and Li [4] studied factors that influence the markup, which can be categorized broadly into internal and external factors, or environment, company, and project factors. Using identified factors as inputs, many markup models built upon past bid experiences, case data, or experts' opinion have been proposed, such as those in Chua et al [5], Chao and Liou [2], and Chao [3]. Although they serve a different purpose from the present study, the modeling methodology employed may be referred to, since both deal with very complex estimation problems.

3. Factors influencing project overhead

Like direct costs, the project overhead costs are entirely project-oriented and consumed on site. The project overhead rate presumably varies according to some project conditions and project features as reviewed broadly below. Generic project factors identified by previous researches to have an effect on the bid markup may also influence the level of project overhead, e.g. a small size project located in a remote area with a long duration area is likely to have both a high bid markup and a high project overhead rate. First, common to all construction projects, many overhead costs exhibit economies of scale, e.g. the larger the work, the lower the manpower required for supervision per unit of work due to a more efficient deployment, and so the project size as represented by total direct cost could be a factor. Next, the charges for several items of overhead costs such as office rents and utility fees are mostly in proportion to the time that a project lasts, and hence project duration is likely another factor. Average direct cost per month derived from the above two factors can indicate the intensity of activity and may be considered as an alternative measurement of project size.

The main type of work of a project, i.e. road, building, etc, influences the number of specialty trades involved, the proportion of labor cost, and the character of the site, concentrated or spreading, and thus has an effect on supervision, coordination, and transportation requirements, which impact on the overhead rate. Similarly, the project location, urban or rural, influences setting-up and maintenance costs of offices, shops, and quarters. In particular, the country in which a project is located is an important factor as the contractor's operation is subject to customs, practices, and laws applicable. In developing countries, numerous taxes and fees are levied on a construction project by various authorities, resulting in higher overhead costs in some categories that have to be allowed for in estimates. In a large country, China, for example, there are even regional differences in this regard.

Other possibly relevant project attributes include scope of contract, i.e. construction only or design and build (D/B), and proportion of subcontracted work. D/B and subcontracting affect resources distribution and lead to changes in interfaces and communication links with effects on costs for attendance, coordination, and engineering in support of production. Last but certainly not least, the quality level required of a project is believed to impact significantly on the staff and documentation efforts for quality control and hence overhead costs. The quality level is to some extent implied by work type and project size, e.g. large mass rapid transit contracts tend to require higher quality than common road contracts. However, it would be better to stand as a separate factor since same size facilities such as buildings can have very varied qualities.

4. Description of data

Ideally, actual costs of completing projects should be used for model development. However, estimated costs of a firm in preparing bids were the only data available to this study and so the models built later would try to capture knowledge containing in the firm's estimates, whose consistency and level of noise would affect model performance. The estimate summaries of 200 projects comprising bids submitted between March 2000 and March 2006 were collected from a large general contractor whose business spreads all over Taiwan covering a variety of types of work. For each project, the data shows project name, owner, project address, bid date, project duration, and a summary of various categories of direct costs, project overhead costs, and markups. Although the firm has standardized cost classification and reporting forms for use by all estimators, some problems in the data were found upon close examination. For example, safety costs for some projects are missing as a result of being bill items and conveniently included in direct costs, while a few projects have major materials as owner-supplied and hence a lower direct cost, which would artificially jack up project overhead rates. Since detailed estimates were not available for corrections, the data for 27 projects was discarded and a usable collection of 173 projects was complied at last.

Numbered from one to 173 in chronological order, the sample projects range between 60 millions and 30 billions NT$ in bid amount (note: 1NT$ ≈ 0.03 US$) and between 3 months and 106 months in duration. The following are classified in this study as project overhead costs: salaries for administration and supervision, office and shops, utilities and services, insurance and bonds, transport, safety, surveys and tests, environmental protection, and public relation. The calculated project overhead rates range between 0.0171 and 0.2912, with a mean of 0.0793 and a standard deviation of 0.0428.

All the projects were located domestically and of the construction-only contract. As the firm has long since adopted a policy of subcontracting almost entire work, there is little difference in proportion of subcontracting among the projects. Moreover, the data available gives very limited project information that does not allow a separate indicator for quality level to be set up. Therefore, only the following factors were considered as inputs in modeling: total direct cost, total duration, mean direct cost per month, classification of work, and classification of location. First, the coefficients of correlation among the quantifiable factors and project overhead rate for the projects as shown in Table 1 were examined. The strong and positive correlations among direct cost, duration, and mean direct cost and between duration and overhead rate appear reasonable. The weak correlation of project overhead rate with either direct cost or direct cost per month may be explained by the lessening of effects of economies of scale by large projects' more demanding quality and other factors at the same time, e.g. the overhead rates of the largest ten projects with a concentration of metros and tunnels average 0.0964 versus the overall mean of 0.0793.

A classification scheme is required for the two categorical attributes, type of work and location. With respect to type of work, a project is classified according to limited project description available into one of eight groups (Table 2). With respect to location, a project is placed into one of three classifications based on its address in Taiwan (Table 3). Tables 2 and 3 also show the statistics of each group ranked in order of increasing mean project overhead rate. Standard deviations within the groups are quite large and statistical tests of the difference between the mean overhead rates of any two adjoining groups conclude that the null (equality) hypotheses all cannot be rejected except for work type groups #5 and #6, suggesting that different groupings may be used.

Table 1- Coefficients of correlation among quantifiable factors and project overhead rate for sample projects

	Total direct cost	Total duration	Mean direct cost per month	Project overhead rate
Total direct cost	1			
Total duration	0.4697	1		
Mean direct cost per month	0.9289	0.2838	1	
Project overhead rate	0.1059	0.3613	-0.0537	1

Table 2- Statistics of project overhead rates for sample projects by type of work

Group number (type of work)	Number of projects	Mean project overhead rate	Standard deviation of project overhead rate
1 (Site works/parks)	16	0.0573	0.0277
2 (Buildings)	30	0.0597	0.0249
3 (Highways/airfields)	52	0.0665	0.0213
4 (Water structures)	5	0.0726	0.0167
5 (Bridges)	33	0.0728	0.0292
6 (Ports/marine facilities)	9	0.1072	0.0488
7 (Tunnels)	11	0.1217	0.0633
8 (Metros/high-speed rails)	17	0.1465	0.0510

Table 3- of project overhead rates for sample projects by project location

Group number (location)	Number of projects	Mean project overhead rate	Standard deviation of project overhead rate
1 (Cities/townships)	89	0.0730	0.0419
2 (Metropolitan areas)	62	0.0820	0.0360
3 (Remote areas)	22	0.0976	0.0577

5. Description of models

5.1 Data representation

While project overhead rate is the model output and project duration a model input, whether direct cost or mean direct cost per month represents project size, another input, requires consideration. Trials later showed that the conversion of direct cost to mean direct cost per month did not enhance model performance. Furthermore, in the presence of duration, mean direct cost per month is somewhat redundant for modeling the estimation. Therefore, it is direct cost that is used as an input to represent project size. As direct cost, duration, and project overhead rate are quantitatively defined, they are suitably represented by their measurements.

For the other two inputs, type of work and location, both categorical, two ways of representation, decimal and binary, are used. With decimal representation each input is assigned a number according to the order of mean overhead rates, i.e. group numbers in Tables 2 and 3, meaning that an increase in the value of each input variable corresponds to an increase in the overhead rate. With binary representation a series (number of categories minus one) of 0/1 variables are used for each input: seven for work type and two for location, bringing the total number of input variables to 11.

5.2 Arrangement of data for model development

The readied data is arranged into two sets for developing and testing models: 152 cases from 03/2000 to 03/2005 and 21 cases from 03/2005 to 03/2006. Data for the first five years are used as a large base of estimates from which a model is developed. Data for the last year is external to the model being developed and is used as future cases for testing its prediction performance in estimating overhead rates. The above arrangement agrees with the fact that a contractor makes estimates for coming projects based on experiences from various kinds of prior projects, which grow with time. Because disruption of time sequence will violate the reality and a model's prediction performance cannot be fairly tested with a completely new case, for each model shown later the steps of modeling and prediction will use the two successive data sets strictly in chronological order of the bids.

5.3 Error measures for performance evaluation

Three error measures are used to evaluate modeling and prediction performance: root of mean squared error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE), as defined below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(outputted_i - applied_i)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|outputted_i - applied_i\right| \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|outputted_i - applied_i\right|}{applied_i} \qquad (3)$$

where n = number of cases used in the evaluation; $outputted_i$ = overhead rate produced for case i by the model; $applied_i$ = overhead rate actually applied to case i according to the firm's estimates.

RMSE inherits the efficacy of mean squared error (MSE), which is used by most algorithms including regression to represent overall system error to be minimized in modeling. Furthermore, RMSE refers directly to the deviation between model output and target output like MAE. It is therefore a consistent and convenient performance indicator for the present single output problem. In the following, RMSE is used as the main evaluator to measure model accuracy in monitoring performance and comparing different models, while MAE and MAPE are used as secondary measures.

5.4 Cluster center and regression methods

Two methods, cluster center and regression, were used to model overhead rates for a comparison. Using the cluster center method, the projects are classified according to combinations of work type and location into clusters, with the maximum number of clusters at 8x3=24. The mean overhead rates for each cluster are calculated as the modeled rates. Where there is a missing cluster, the mean for the work-type group is used in its stead. Using the regression method, two multiple regression equations involving all four inputs were built: one with the decimal representation and a total of four independent variables and the other with the binary representation and a total of 11 independent variables. The built equations are then used to produce overhead rates for cases within the modeling set as modeled rates and those for cases within the testing set as predicted rates.

For each model above, the RMSE, MAE, and MAPE of modeling representing closeness of fit and those of prediction representing test accuracy are calculated using (1), (2), and (3), respectively. The results are shown in Table 4.

Table 4- RMSE, MAE, and MAPE of modeling (152 cases) and prediction (21 cases)

Models	RMSE of modeling	RMSE of prediction	MAE of modeling	MAE of prediction	MAPE of modeling	MAPE of prediction
Cluster center method	0.0289	0.0395	0.0207	0.0273	0.2870	0.3673
Regression with decimal rep.	0.0336	0.0373	0.0243	0.0273	0.3291	0.4377
Regression with binary rep.	0.0305	0.0409	0.0218	0.0266	0.3099	0.4202

6. Discussion and conclusions

Compared with the large standard deviation (0.0428) of overhead rates, all of the three models represent a significant improvement (error reduction by more than 1% of direct cost) as a result of explaining factors being introduced, meaning that the input factors are relevant for modeling. Overall, the cluster center method outperforms the regression method and the regression model with binary representation outperforms the regression model with decimal representation. This indicates that linear regression with decimal representation is unsuitable for the problem as it fails to improve performance by picking up the extra factors of project size and duration left out by the cluster center method.

An empirical model's accuracy is inevitably affected by the level of noise in the data used for developing the model, so its performance must be judged considering this influence. For the present study, noises in the overhead rates for the sample projects come from over- or under-estimates of direct cost (inaccuracies in the denominators) and over- or under-estimates of overhead cost (inaccuracies in the numerators), both causing the rates to deviate from what they should be. Since a project's cost estimate can achieve ±3 percent accuracy with the total design available [7] and the direct cost constitutes the bulk of it, the best result of about 3 percent error of direct cost achieved by either the cluster center method or the regression method with binary representation is considered acceptable for the problem, although there is room for improvement. However, the fact that the cluster center method using only two factors achieves comparable or better performance in closeness of fit and test accuracy than the regression method with binary representation requires further consideration.

Although the overhead costs of a project have a lot to do with its legal and business environments and have to be considered within a local context, the presented approach is general and can be applied in any country. As the data used for model development relates to a firm's costs, the models constructed are intended for use by that particular firm, but other organizations can use their own data to the same effect. While subject to limited availability of data with a lot of noise, heuristics from this study suggest that suitable factor selection and data representation are required for producing better results. Continual model updates with the buildup of estimates would be helpful for improving performance as the base of cases expands with time.

Because of the exploratory nature of this research, the presented models are just prototypes that still need to be refined and improved. As their effectiveness is limited by the correctness of the bid data, it is suggested that future researches collect actual costs for use in model development. However, even actual cost data is available, it may not be more dependable than estimate data because of errors in assigning and reporting costs. Checking the data's consistency is important whichever is used. As the present study left out some potentially significant factors affecting model accuracy, such as level of required project quality and type of contract, they can be considered for inclusion as well as more detailed classification schemes for work type and location. Based on the findings of this study, the use of a nonlinear model such as artificial neural networks for dealing with the complex relations existing between the inputs and output of the stated problem is called for in attempts to improve modeling and prediction performance.

References

[1] Ahmad, I., and Minkarah, I. (1987). Optimum mark-up for bidding: a preference-uncertainty trade off approach, Civil Engineering Systems, Vol. 4, 170-174.

[2] Chao, L.-C. and Liou, C.-N. (2007) “Risk-minimizing approach to bid-cutting limit determination.” Construction Management and Economics, 25(8), 835-843.

[3] Chao, L.-C. (2007) “Bidding model for design/build projects.” Proceedings of Construction Management and Economics 25th Anniversary Conference, University of Reading, UK, paper PR-0025.

[4] Chua, D.K. and Li, D. (2000). “Key factors in bid reasoning model.” Journal of Construction Engineering and Management, ASCE, 126(5), 349-357.

[5] Chua, D.K.H., Li, D.Z., and Chan, W.T. (2001). “Case-Based Reasoning Approach in Bid Decision Making.” Journal of Construction Engineering and Management, ASCE, 127(1), 35-45.

[6] Diamant, L. (1988) Construction Estimating for General Contractors, Wiley, New York.

[7] Halpin, D.W. (1985) Financial and Cost Concepts for Construction Management, John Wiley & Sons, New York.

[8] Holland, N.L. and Hobson Jr., D. (1999) “Indirect cost categorization and allocation by construction contractors.” Journal of Architectural Engineering, ASCE, 5(2), 49-56.

[9] McCaffer, R. and Baldwin, A.N. (1984) Estimating and Tendering for Civil Engineering Works, Granada, London.