Improving the Estimation of Project Overheads in Construction Companies in Hong Kong

by

Caroline Tak Wa Chan

Doctoral Thesis

Submitted in partial fulfillment of the requirements

for the award of

Doctor of Philosophy of Loughborough University

4 June 2007

Abstract

Project overheads cover the site cost of administrating a project as a whole, rather than a particular work section. Estimation of these items is one of the routine tasks of all parties including the contractors and project owners. Nevertheless, our understanding of this subject mainly lies on the theoretical level due to the limited empirical study in the past. More importantly, estimation of project overheads demands a lot of expertise but still exhibits a high risk of inaccuracy. Therefore, the aims of this research study are to explore the estimation and expenses of project overheads in practice and to devise an efficient model for project overheads estimation.

In the data collection process, surveys and interviews were conducted with large contractors in Hong Kong. A comprehensive review on the expenses pattern, estimating method and accuracy of project overheads was thus developed upon the empirical base. To improve the estimating accuracy and efficiency, an estimating model utilizing artificial intelligence was designed. The model was an artificial neural network (ANN) model adopting group method of data handling (GMDH) algorithm. Input variables were determined by an opinion survey followed by exploratory factor analysis to extract the principal factors for modeling. The model was trained with sixty-three project cases collected from the contractors in Hong Kong and validated by another eight cases which were not being used in the training process before. Satisfactory training and testing results were obtained, together with an identification of five significant variables affecting project overheads. The model had undergone further validations including eight rounds of cross-validation; comparison with linear multiple regression; and comparison with other ANN architectures like multi-layered feed forward network and general regression network. All the results evidenced that the proposed GMDH model was a reliable tool, producing accurate predictions on project overheads. To affirm the applicability of the model, a focus group of seven senior quantity surveyors was conducted. Members of the group conceded that the model was efficient and accurate; and worth further development into a tailored model for individual company.

Acknowledgement

I would like to express my sincere thanks to my supervisor, Dr Christine Pasquire, for her patience and guidance throughout my study. I am also grateful to Ms Helen Newbold of the Department for her kind assistance. Finally, I must thank my family for their continuous support and encouragement over the past few years.

Table of Contents

List of Tables

List of Tables (cont'd)

List of Tables (cont'd)

List of Tables (cont'd)

List of Figures

List of Acronyms and Abbreviations

ANN	Artificial neural network
BOND	Bonds and warranties (factor)
COMP	Project complexity (factor)
DES_VO	Design and variations (factor)
ECON	Economic condition during construction stage (factor)
FCPSE	Full complexity prediction squared error
FPE	Final prediction error
GCV	Generalized cross validation
GMDH	Group method of data handling
GRNN	General regression neural network
HKSMM	Hong Kong Standard Method of Measurement
LOCAT	Location (factor)
MDL	Minimal description length
MLFF	Multi-layered feed forward
OUT_COST	Output – Project overhead cost
PNN	Probabilistic neural network
PQS	Professional quantity surveyor
PROCU	Procurement nature (factor)
PSE	Prediction squared error
RELAT	Contractor's past relationship with project parties (factor)
SC_FIN	Financial strength of sub-contractors (factor)
SC_PER	Sub-contractors performance (factor)
SHAPE	Site and building shape (factor)
SIZE	Project size (factor)
STA_PLT	Extent of staff and plants assigned to project (factor)
SUBSID	Contractor is client's subsidiary (factor)
TEAM	Strength of site management team (factor)

CHAPTER 1

Introduction

1. Introduction

Cost estimating is a technical process of predicting costs of construction (CIOB, 1997). The process is unavoidable to contractors and consultants as it is where the consultants provide the budget for project owner's information and approval; and contractors prepare tenders for bidding and future cost control. Although there are ample text books detailing the theoretical methods of construction cost estimation, not much literature related to the practice of cost estimation is found (Akintoye and Fitzgerald, 2000).

1.1 Rationale of Study

According to the CIOB[1] Code of Estimating Practice (1997), project overheads mean, "the site cost of administering a project and providing general plant, site staff, facilities and site-based services and other items not included in all-in-rates. They are also commonly known as: preliminaries, general cost items or general expenses." They represent the cost-significant items that are required by the works to be carried in the project, and concerned with the whole of the works rather than just a particular work section. As described by the technical information of the National Building Specifications issued by the RIBA Enterprises, these associated costs may either be 'one-off' fixed costs, such as the cost of bringing to site and erecting site accommodation (and subsequent removal) or time-related costs, such as the water supply, lighting and maintenance cost for the site accommodation. Although project overheads normally account for a relatively small percentage of the contract sum (Assaf *et al.* 1999, Hegazy and Moselhi 1995), it is important to the success in bidding a tender and in the running of a project; as explained in the following sections.

1.1.1 Inherent Difficulty to Estimate Project Overheads

In typical bills of quantities, there are mainly four parts requiring contractors to price: Project overheads, Measured works, Prime Cost Sums, and Provisional Sums. For Provisional Sums, no estimation effort is required from the tenderers. For the Prime Cost Sums, tendering contractors are only required to enter the attendance and profit markup against the Sums. Hence, the actual parts of tender that demand greatest estimation effort are the Project overheads and the Measured Works.

While in Hong Kong most of the projects are divided into work packages and sublet to sub-contractors, the tendering main contractors are only required to mark up on the best

[1] Chartered Institute of Building, U.K.

quotations from the subcontractors and suppliers. More careful contractors may adjudicate the sub-contractors' tenders thoroughly with reference to the companies' cost database and current market information. Occasionally, unit rates may also be required to build up the cost for the work item if the work cannot be sublet. In that case, the costs of labour, plant and materials, together with profit markup will be estimated (detailed literature on it will be further elaborated in the Literature Review Chapter). For the Project overheads, however, estimation will be more complicated. This section covers not only site management fee, but also other contractual obligations and requirements like insurance, site cleaning, site facilities, etc. Some of the items are related to the project duration (e.g. management fee, watching), whilst others are reliant on the amount of works to be executed (e.g. insurance, levy). However, there are some items dependent on costs which are not so easy to define or anticipate. For example, when considering the site office and sheds in Hong Kong project, such provision can vary from simple containers to very well-furnished, purpose-made temporary structures. In some occasions, the temporary site offices may be moved into the newly completed building envelope or moved to office apartments of nearby office towers when the site is extremely congested. Another example is protection. Protection to materials can range from a simple shelter to a properly secured, weatherproof and watertight temporary store. If consider protection to completed works, it can imply various degrees of protective measures to the finished works. Therefore, this item depends not only on the amount of works, but also the corporation of various sub-contractors and overall housekeeping of site. During the tender stage, it is difficult for the estimator to realistically forecast the future arrangement based on the tender document. Besides the inherited uncertainty, the tendering period available for the contractor is usually very short when considering the amount of effort required to appraise the whole tender document in detail before preparing an estimate. Therefore, giving an accurate estimate of project overheads in the tender stage is a difficult task. Nevertheless, most of the project owners in Hong Kong (especially for the large projects) demand a breakdown of the project overheads for the purpose of future payment or assessment of prolongation cost.

1.1.2 Importance of Project Overheads Estimation to Contractors

Although estimators may intuitively think that project overhead estimation is not an important task, quite a lot of theorists advocated that accurate project overheads estimation is essential in order to provide a competitive bid. Tah *et al.* (1994) claimed that estimation of direct cost would be very similar among different bids, and the main variations among the competitors' bids were basically the mark-ups and indirect costs (project overheads). They further suggested that in order to improve the tender prices, efforts should be directed to improve the methods of estimating the project overheads. Solomon (1993) also claimed that the main area in a tender where the contractor could seek to gain competitive edge was the adjustment in project overheads. Assaf *et al* (1999) had similar suggestion that reducing and controlling project overheads could bring continual competitiveness to the company. Taylor (1994) put forward the scenario that if contractors sublet all or most of the works to the subcontractors, equal or offsetting bids would be produced by the contractors from their subcontractors' quotations. Hence, the critical components that marked the difference among contractors' bids would be their allowance for overhead and profits. This applies to cases like Hong Kong and a lot of other places where sub-contracting is a common practice in the construction industry. Under that scenario, the risk associated with procurement and installation of the trade works can be easily transferred to the subcontractors through subletting. However, the risks associated with project overheads provision like setting up of site office, site cleaning, site management, etc. are rather difficult to transfer. Any associated risk due to poor estimation or management has to be borne by the main contractor.

Furthermore, as mentioned earlier, estimation of project overheads is a difficult task when compared with other estimating processes. If contractors can develop a reliable system to estimate the project overheads with satisfactory accuracy and efficiency, cost of estimation can be very much reduced. Such cost savings can lower the head office overheads of the company in the long run.

1.1.3 Importance of Project Overheads Estimation to Consultants and Project Owners

In the last section, how an accurate estimation of project overheads brings significant benefits to the contractors is discussed. In fact, the benefit is duly shared by PQSs and project owners as well. Figure 1-1 illustrates the flow of "cost estimates" from contractor to PQS and project owner. If contractor can provide an accurate estimate of project overheads in the bid, the figure can serve as a better piece of cost information for the PQS to prepare a more realistic budget. Project owner can be benefited from the more accurate budget which helps better cost control of the project during the contract period. Moreover, if the estimation of project overheads is being done more accurately, the financial stress of the contractor (due to under-estimation) can be reduced. This can help to lower the claim consciousness of the contractor during the contract period when the contract is awarded to him.

Furthermore, as mentioned in the last section, the overall improvement in contractor's estimation of project overhead cost can reduce the estimating expenses of the contractor. Such reduction in the head office overheads will become savings in the operating cost of the contractor which will be reflected in the contractor's bid.

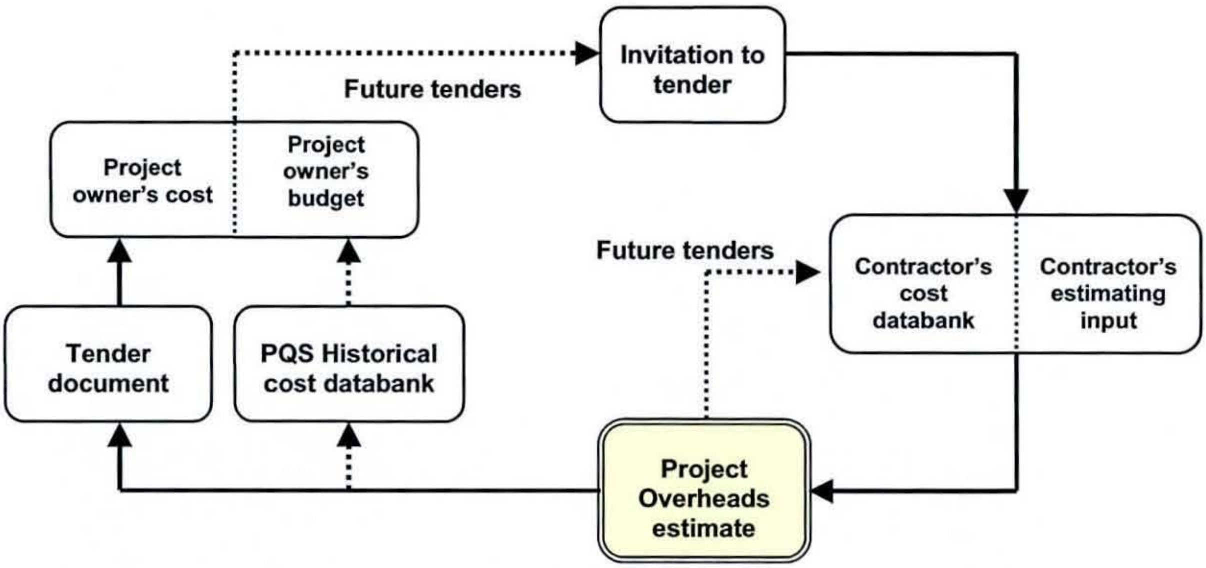

Figure 1-1 Flow of cost data among contractor, PQS and employer in terms of project overheads estimation

In short, estimation of project overheads is an important task in the whole estimating process to all parties, especially the contractors. Nevertheless, the uncertainty inherited with project overhead items makes them difficult to be estimated with satisfactory accuracy.

1.2 Aims and Objectives

No matter how difficult it is to prepare an accurate estimate for project overheads, contractors still need to do so in order to get new projects and maximize their profit. Although the Code of Estimating Practice (The Chartered Institute of Building, 1997) seems to be comprehensive and detailed, there are still a lot of hidden uncertainties (as mentioned before) which make the estimation task for project overheads extremely difficult.

In fact, the CIOB stated very clearly that the guidelines (Schedule) in the Code should not be considered as absolute. Actual pricing of project overheads will vary with different policies and preferences of individual company. Hence, the Code only provides a direction of thinking or consideration to the users. Detailed build-up of the corresponding amount for each project overhead item still requires a lot of estimating effort and judgment from the estimator-in-charge.

Therefore, there is a need for a reliable and accurate estimating tool for project overheads; so as to relieve the difficulty to estimate these items and to achieve cost effectiveness across different parties. Therefore, the aim of this research is to improve the accuracy of project overheads estimation by the use of artificial neural network models. To achieve this, the objectives of this research are set as follows:

1. To explore the nature of construction project overheads and the associated estimation methods in use by practitioners in Hong Kong.
2. To identify critical project overhead items that possess higher risk of genuine under- or overestimation by tenderers in Hong Kong.
3. To identify the overall pattern of project overhead expenses amongst projects in Hong Kong.
4. To develop an estimating model to predict project overheads in an accurate and efficient manner using artificial intelligence.
5. To identify significant factors affecting project overhead expenses in Hong Kong.

In most cases, difficulty at work is arisen from the lack of knowledge about the subject matter. Estimation of project overheads is also facing a similar challenge. The existing literature (especially empirical studies) about project overheads is very limited, causing difficulty to devise an efficient estimating tool. Therefore, the first and second objectives are to widen the understanding of project overheads. The third and the fourth objectives are to develop a reliable and efficient tool to facilitate the estimation process of project overheads. With the identification of the significant factors affecting the project overheads expense, management of the contractor can allow more attention to those items (that are controllable by them) in order to reduce the project overhead cost in the long run. This is the reason why the fifth objective is established.

1.3 Hypotheses

To meet the objectives of this research, three hypotheses are set including:

1. There is substantial inaccuracy in the estimation of project overheads by the large contractors in Hong Kong.
2. Project overhead cost can be estimated more accurately by artificial intelligent modelling than conventional methods such as multiple regression.
3. Project overhead cost can be predicted by simple parameters satisfactorily.

To achieve the first three objectives, the research focused on the nature and perceptions of project overheads, as well as the related estimation methods used by the practitioners. The likelihood of inaccurate estimation was thus established. According to some past studies related to project overheads estimation, there is considerable variation amongst the bidders with a substantial number of them relying on professional judgment during the estimating process. It was considered likely that similar results would be obtained if a similar study was carried out in Hong Kong. Therefore, the first hypothesis was designed as: "There is substantial inaccuracy in the estimation of project overheads by the large contractors in Hong Kong."

When developing an ANN model to estimate project overheads for the fourth objective, a conventional cost estimating method such as multiple regression was used to validate the performance of the ANN model. As most ANN researchers suggested that ANN models outperformed multiple regression, it was anticipated that the validation results of ANN model would be satisfactory. Therefore, the second hypothesis was designed as: "Project overhead cost can be estimated more accurately by artificial intelligent modelling than conventional methods such as multiple regression."

The last objective of this study was to identify the significant factors affecting project overhead expenses in Hong Kong. To achieve this, the GMDH[2] algorithm was used as it can identify the significant inputs to be selected in the modelling process. Such findings

[2] GMDH refers to Group Method of Data Handling. It was developed by Ivahnenko as a vehicle to identify nonlinear relations between input and output variables. GMDH algorithms are applicable in a wide variety of application e.g. data mining, forecasting, pattern recognition and neural network modelling. Details of the GMDH algorithm were explained in Literature Review Chapter.

support the third hypothesis: "Project overhead costs can be satisfactorily predicted by simple parameters."

1.4 Review of Methodology

As a brief review of the methodology used in this study, the methods used to achieve the various research objectives (and to verify the hypotheses) are described. The whole research process was divided into six phases which could be grouped into two main parts: exploratory study of project overheads and the development of the estimation model.

1.4.1 Exploratory Study of Project Overheads

1.4.1.1 Phase 1 - Literature Review

Similar to typical researches, the research process was started with a comprehensive review of the classical theories and recent researches related to the topic. The literature review mainly covered four areas: project overheads, estimation methods (both classical and contemporary methods), cost estimating accuracy and factors affecting project overheads.

1.4.1.2 Phase 2 – Review the Practice of Project Overheads Estimation

This part of study is to establish the information regarding the general nature and characteristics of project overheads and estimating practice in the industry. Two surveys, together with cost data collection, was devised in this phase. In order to ascertain the estimating practice of project overheads amongst the large contractors in Hong Kong, survey by questionnaire (Stage 1 survey) was developed to collect information in this regard. Survey results were then analyzed by descriptive analysis to summarize the general methods adopted by the estimators in Hong Kong to estimate project overheads. Follow-up interviews were also designed to collect more detailed information e.g. perceived important project overheads from the respondents.

An opinion survey (Stage 2 survey) was designed to collect views on the likelihood of inaccurate estimation of each project overhead item. Descriptive analysis of the results could provide relative accuracy indices of each project overhead item so as to rank the extent of estimation inaccuracy accordingly. As a result, the items that possessed a higher likelihood of over- and under-estimation could be found.

To analyze the cost distribution of different project overhead items, project cost data was collected from large contractors in Hong Kong. Unlike the past study on project overheads distribution which only analyzed the tender allowance in contract, actual expenses were collected here. Thus, the relative percentages of actual expenses on each project overhead items could be assessed. As different companies have different cost accounting system, a common cost breakdown structure for the project overhead items had to be developed for data collection. Then, percentage expenditure of each project overhead item could be analyzed.

1.4.2 Development of Estimation Model

1.4.2.1 Phase 3 - Develop the List of Principal Factors Affecting Project Overheads

To build a cost estimating model, important factors affecting the cost have to be ascertained first. Therefore, possible factors affecting project overheads were consolidated based on the existing literature. Nevertheless, in order to ensure all possible factors being included in the preliminary list, a focus group discussion inviting experienced quantity surveyors was designed to appraise the preliminary list.

Then, an opinion survey was conducted with the Quantity Surveying Managers of the large contractors in Hong Kong to scrutinize the preliminary list of project overhead items. The survey results were analyzed by exploratory factor analysis to extract the principal factors for use as input variables in the estimating model.

1.4.2.2 Phase 4 – Design Estimating Model

The selection of modelling technique for the estimating model was made across three methods: fuzzy logic, artificial neural network and multiple linear regression. Design of the model was made upon the qualitative analysis of the three methods. Details of the model architecture were then developed.

1.4.2.3 Phase 5 – Train and Validate the Model

With the input variables (i.e. principal factors) identified by exploratory factor analysis and the model structure designed, measurement level of project data in relation to these input variables and the total project overhead cost was determined with reference to the previous work of other researchers and guidelines of the modelling software. Sampling method adopted to collect the project data was also considered.

The collected data were then fed into the ANN model using GMDH algorithm and trained sequentially with six selection criteria ((Full Complexity Prediction Squared Error (FCPSE), Prediction Squared Error (PSE), Minimal Description Length (MDL), Generalized Cross Validation (GCV), Final Prediction Error (FPE) and Regularity). Non-significant variables were removed from the inputs and re-trained again until all the non-significant variables were removed with the best model producing satisfactory accuracy. The model that produced the lowest error and gave the simplest formula was the desired solution.

A series of validations were designed to test the reliability of the model. Cross-validation was developed to verify the reliability of model under small sample size condition. Testing by other ANN networks (including multi-layered feedforward and general regression neural networks) and linear regression were used to verify the validity of the proposed model and the significant factors.

1.4.2.4 Phase 6 –Verify the Model

Validation results were analyzed to verify the reliability of the proposed model. If satisfactory results were obtained, the variables could be regarded as significant factors affecting project overheads and the proposed model could be described as reliable.

Before recommendation, the validated model was presented to a focus group of senior quantity surveyors to appraise whether the model was a friendly tool for estimating project overheads; and its likelihood of applying in the industry. This is to ensure that the research has practical as well as academic value.

1.5 Review of Work Done

The whole study was directed towards the achievement of the objectives as described in the last section. Therefore, five main research outcomes were produced. Altogether four papers were written to disseminate the associated data, including one of them under the re-review process. Two papers were international conference papers and the other two were academic journal papers.

1.5.1 1st Research Outcome : Estimating of Project Overheads in Practice

The first stage survey was aimed at collecting information for the 1st research outcome to achieve the first objective (to explore the estimating practice of project overheads). Forty-nine responses were successfully received out of the 119 sample. Results of the survey proved that majority of the contractors prepared the project overheads estimate in detailed, relied on estimators' professional judgment. The results were presented in the 18th ARCOM conference and were briefly described below.

1.5.1.1 18th ARCOM paper : Estimation of Project Overheads : A Contractor's Perspective

This paper was presented in the 18th Annual Conference ARCOM conference held in Newcastle Upon Tyne. It presented the findings of the survey followed by interviews within forty-nine large contractors in Hong Kong. The results confirmed that detailed estimation based on professional judgment was the most common method used by the estimators to estimate project overheads. Besides, a strong misalignment was observed between their perceived importance of project overheads estimation and the level of resources allocated to project overheads estimation by the contractors. Only 9% of the interviewees thought that project overhead estimation was the most important but all of them allocated the most expertise to this part of estimate. Project overhead items like site management salaries, protection to works, and site cleaning were commented by the contractors as being significant in terms of inaccurate estimation. A framework of twenty-three items was generated as a representative list of important project overhead items that had to be priced (in other words, having financial impact). The 23-item list provided a basis for the development of the questionnaire on the likelihood of inaccurate estimation of project overheads in the next stage of study.

1.5.2 2nd Research Outcome : Likelihood of Inaccurate Estimation of Project Overheads

In the Stage 2 survey, a questionnaire was designed to study the likelihood of estimation inaccuracy among the twenty-three project overhead items. Forty responses were received from 119 sampled contractors. The survey evidenced that inaccurate estimation was likely, especially for items like site cleaning, protection of finished works. The second objective, to identify critical project overhead items that possess higher risk of genuine under- or overestimation by tenderers in Hong Kong, was achieved satisfactorily. From the Stage 1 and Stage 2 surveys, the first hypothesis was evidenced: inaccurate estimation of project overheads by the large contractors in Hong Kong is substantial. Details of the analysis were published in the journal paper as elaborated below.

1.5.2.1 Journal Paper : An Analysis for the Degree of Accuracy in Construction Project Indirect Costs

This paper was published in the Journal of Cost Analysis and Management, Summer, 2004. It presented the findings of a survey conducted with forty large contractors in Hong Kong. The data collected were analyzed by descriptive statistics which revealed that under-estimation was more common than over-estimation. The items of *protecting finished works* and *cleaning and removal of rubbish* were found to be most difficult to estimate accurately, possessing the highest likelihood of both under-estimation and over-estimation. Furthermore, the likelihood of inaccuracy found in different types of company was quite different. This implied the differing project overhead requirement in different types of projects (e.g. civil engineering vs building projects). Around 40% of the respondents felt that their project overhead estimates were not accurate and that a shortfall was not easy to recover when there was under-estimation.

1.5.3 3rd Research Outcome : Project Overheads Expenses Pattern

The third research objective was to identify the overall pattern of project overhead expenses amongst projects in Hong Kong, and thus the data collection process was developed accordingly. Project overhead expenses breakdown of completed projects was collected from the large contractors in Hong Kong. The cost data was extremely sensitive as it was the internal cost data of the project. In the end, twenty project cases were collected. Although the sample size was small, the data provided indicative information about the distribution of project overheads which had never been explored in the past.

The findings of this data collection process were presented in the PMI Research Conference held by the Project Management Institute in July 2006. The major findings are summarized in the following section.

1.5.3.1 PMI Research Conference Paper : A Decision-Making Matrix Model for Construction Project Overheads Estimation

This paper was presented in the PMI Research Conference held in Montreal in July 2006. It covered the findings on the project data collected from twenty large projects in Hong Kong in order to establish the general cost breakdown pattern of project overheads. Besides, a decision-making model for top management to direct the estimating resources in order to achieve the highest efficiency was also recommended. From the project data, the total project overhead cost accounted 11% to 19% of the total project cost, which was quite in line with the literature (not more than 20% of the total project cost). Seven items were identified to account for more than 80% of the total project overhead cost, including 1) site management; 2) mechanical plants; 3) cleaning and removal of rubbish; 4) insurances and surety bond; 5) setting out; 6) temporary works; and 7) scaffolding. Site management remained as the most significant project overhead item, contributing an average of 36% of the total project overhead cost. The second was mechanical plants, which contributed 12.4% on average. The third item was cleaning and removal of rubbish which accounted around 10%. These three items already accounted 58.3% of the total cost. A matrix model was developed to differentiate project overhead items into four categories according to their potential cost and the level of estimation accuracy (as found in Stage 2 survey). By doing so, the limited estimating resources can be allocated more efficiently. The four levels of decision for estimating resources allocation were: 1. apply full capacity; 2. exert reasonable effort; 3. don't ignore them; and 4. estimate with professional judgment. This matrix model serves as a good guideline for contractors who are required to estimate itemized project overhead cost.

1.5.4 Development of Cost Estimation Model

1.5.4.1 Extraction of Principal Factors for Inputs of Model

To construct the ANN model, principal factors affecting project overheads had to be determined. Through the opinion survey, the preliminary list of factors affecting project overheads (consolidated from the literature review and focus group discussion) was scrutinized by exploratory factor analysis. Seventy-nine responses were received from the contractors. Fourteen factors were finally extracted from the original list of forty-eight factors using exploratory factor analysis. The findings of this analysis were detailed in the ASCE paper as described below.

1.5.4.1.1 ASCE Paper : Identifying the Principal Factors Affecting Project Overhead Expenses : An Exploratory Factor Analysis Approach

The paper was submitted to the Journal of Construction Engineering and Management and is undergoing the re-review process. This paper presented the interim results of identification of significant factors affecting project overheads. Before developing the ANN estimation model, identification of principal factors affecting project overheads expenditure had to be done for designing the model architecture. After compiling a comprehensive list of variables from the literature and focus group discussion, questionnaires detailing forty-nine variables were sent to Quantity Surveying Managers of large contractors in Hong Kong. Seventy-nine responses were received and analyzed by exploratory factor analysis using SPSS. From the results, eight external factors and six internal factors were extracted and interpreted with their relative importance.

The 8 external factors extracted were: 1) Design and variations, 2) Economic uncertainty, 3) Bonds and Warranties, 4) Project complexity, 5) Procurement nature, 6) Site and building shape, 7) Location and 8) Project size. The 6 internal factors were: 1) Strength of site management team, 2) External relationship of contractor, 3) Extent of staff and plants assigned to project, 4) Financial strength of sub-contractors, 5) Sub-contractors performance and 6) Contractor is client's subsidiary. These 14 factors were then used for the development of ANN model to estimate project overheads.

1.5.5 4th Research Outcome : Develop ANN Model to Estimate Project Overheads

Seventy-one project data sets associated with the fourteen principal factors together with the final project overhead cost were collected from the large contractors in Hong Kong to develop the estimating model. The GMDH network was first trained with the fourteen factors as inputs. After three levels of re-training to remove the non-significant factors from the inputs, the best model using five inputs were produced by the FCPSE selection criterion. The models were also tested with production data set that was not used in the training process after each trial. The R squared of the best model using training set and production set was 0.8818 and 0.9754 respectively, showing very good predictive performance.

The model was then cross-validated to increase the reliability of it as the sample size was considerably small. Satisfactory predictions were received in all trials with the GMDH network using FCPSE selection criterion. The best formula was then compared with the multiple regression model. The GMDH network using FCPSE selection criterion outperformed the multiple linear regression model. The regression equation produced by SPSS was:

OUT_COST = -23,337,799.14 +532.4383 (SIZE) + 15,488,602.1262 (STA_PLT)

producing an R squared value of 0.7961.

The cross-validation and the comparison with linear regression modelling proved that ANN using GMDH algorithm was a suitable and reliable method to predict project overheads. In other words, the second hypothesis: "Project overhead cost can be estimated more accurately by artificial intelligent modelling than conventional methods such as multiple regression." was verified. Therefore, the best model (produced by ANN) was concluded as follow:

Best model : GMDH algorithm using five input variables and FCPSE as selection criterion

$Y = -0.92*X_3 + 0.21*X_4 + 0.17*X_5 + 3.7*X_3*X_5 - 0.44*X_1*X_3$

Where $Y = (OUT\_COST-427{,}000)/129{,}373{,}000$; $X_1 = (ECON-85.32)/26.43$; $X_3 = (SIZE-3{,}500)/179{,}500$; $X_4 = (RELAT-1)/4$ and $X_5 = (STA\_PLT-1)/4$

1.5.6 5th Research Outcome : Identify the Significant Attributes Affecting Project Overhead Cost

As mentioned in the Review of Methodology Section, GMDH network was chosen for its ability to produce the best formula with the identification of significant variables. Five variables were identified as significant after a series of training by the ANN model using GMDH algorithm. These five variables were: economic uncertainty, procurement nature, project size, external relationship of contractor and extent of staff and plants assigned to the project.

These five variables were also validated with other ANN models, including multi-layer feed-forward with backpropagation and general regression neural network, to verify their reliability to predict project overheads. Validation results were satisfactory, with R squared above 0.65 and 0.5 for MLFF and GRNN respectively (for all trials in training, testing and production). This indicated that the five variables were reliable to predict project overheads. Therefore, this research outcome also proves the 3rd hypothesis that project overheads can be estimated by simple parameters satisfactorily.

1.6 Review of Conclusions

The various stages of data collection and analyses in the exploratory study of project overheads confirmed the difficulty and accuracy involved in project overheads estimation, and the high level of estimating expertise allocated to this estimating task. This clearly justifies the need for this research study based on the dearth of current knowledge in project overheads. This research also identified the misconceptions regarding the nature of project overheads amongst the contractors in Hong Kong. These misconceptions included the importance of project overheads, the proper method to estimate project overheads, and a systematic allocation of estimation resources among various project overhead items which should be revisited by the estimators.

To estimate the overall project overhead cost, the ANN network with GMDH algorithm was proved to be a suitable and reliable method with satisfactory accuracy. Similar to other ANN studies, ANN outperformed conventional regression model. The proposed model can provide a useful indicative figure for the overall project overheads cost of a project using only five input variables. It is thus a time-saving and accurate method in lieu of the existing practice. The five attributes that were identified as reliable variables to predict project overhead cost included : *Economic uncertainty*, *Procurement nature*, *Project size*, *External relationship of contractor*, and *Extent of staff and plants assigned to project*. With these five significant factors successfully identified, more attention can be paid to them during estimating in order to arrive at an accurate estimate of project overheads. Besides, contractor can spend more effort to improve the management of those significant attributes that are within its control (External relationship of contractor and Extent of staff and plants assigned to project) in order to reduce the project overhead cost in the long run.

On the other hand, the analysis of project overhead cost breakdown provided a useful decision-making framework to allocate estimating resource more effectively. Project overhead items can be estimated with different levels of resources and expertise according to their respective likelihood of estimating inaccuracy and cost impact.

1.7 Guide to the Thesis

The thesis was divided into eleven chapters, including this one. The rationale of study, objectives and hypotheses of the research and an overview of the whole thesis is presented in this Chapter. The next chapter is the Methodology Chapter. It gives a broad framework to illustrate how the research process is developed. It illustrates how the whole research project is divided into six phases leading to exploratory study of project overheads and development of project overheads estimation model. Details regarding development of the data collection and sampling method in each phase are explained with clear elaborations on the rationale made.

Following the Methodology Chapter is the Literature Review. All the relevant literature regarding project overheads is reviewed. Besides, the theories related to the research model and artificial neural network are also examined. The factors affecting project overheads and other related domains (project cost and time, project risk) are also studied.

The design of research instruments like questionnaire and data analysis tools including the modelling technique is exemplified in three chapters:
Chapter 4 - Design of Exploratory Study of Project Overheads,
Chapter 5 - Development of Inputs and Outputs to Cost Model, and
Chapter 6 - Design of ANN Cost Estimating Model.

The Design of Exploratory Study of Project Overheads Chapter covers the design of questionnaires used in the surveys and the approach to collect project overheads cost. Special features regarding the data analysis are also discussed. The Development of Inputs and Output to Cost Model Chapter mainly illustrates the development of inputs and output for the model. Details related to the questionnaires, factor analysis of the survey results, approach to collect input and output information are fully explained. The cost estimating model development is portrayed in the Design of ANN Cost Estimating Model Chapter. The design of the model and choice of validation methods are all illustrated with the rationale behind the choice explained. References to the relevant theories and principles are also made.

The Data Analysis Chapter presents the data and findings of the various stages. The response rate and data collected in each step is shown. Since the findings of one stage of study often leads to the research method of the next stage (e.g. views from a survey are used to design the questionnaire of the next stage), the data analysis and findings of each stage are examined thoroughly before moving to the next stage. Therefore, the link between each stage can be visualized clearly.

Before proceeding to the conclusion of the study, the Analysis and Validation of ANN Model Chapter is incorporated to present some areas of concern that may be interested to the researchers in similar subject or to practitioners who are involved in the estimating process. Limitations of the current research are also discussed. The Conclusion and Recommendations Chapter helps to summarize the whole research study in a more summative manner. Main ideas or findings in each chapter are captured as an overview. Recommendations for further research are also drawn. The References and Bibliography Chapter lists out all the references cited or based upon when doing this research study. The last chapter is the Appendix Chapter. It contains all the questionnaires, structured survey form and elaboration on calculations. Furthermore, detailed training and validation results generated from the ANN models are tabulated in this Chapter as well.

CHAPTER 2
Methodology

2. Methodology

2.1 Introduction

Empirical study of project overheads estimation is rather limited, especially those related to the practice in Hong Kong. Therefore, the purpose of this research is to reveal the estimating practice of project overheads in contractors and the characteristics of project overhead expenses. As a result, significant factors affecting project overhead can be identified; and a reliable statistical model can be devised to estimate project overheads in a more efficient and accurate way. This Chapter is to elaborate systematically the research process that had been gone through in order to achieve the said objectives. The process covered gathering of theoretical information and empirical data, as well as various approaches to develop the estimating model.

2.2 Overview of Methodology Adopted

The methodology applied in this research involved the exploratory study of project overheads and the development of an estimating model to estimate them accurately. As portrayed in the flow diagram in Figure 2-1, there were six phases in the whole process. The first two phases, literature review and empirical study of industrial practice, were implemented to explore the project overheads in greater detail. The subsequent phases from 3 to 6 were designed to construct and validate the estimating model for project overheads. Details of the development of each phase are explained in the latter sections. Table 2-1 summarizes the research methods involved in this study.

2.3 Exploratory Study of Project Overheads

Before building up a cost estimation model, a thorough understanding of the subject matter was essential. Details regarding the estimation practices and respective accuracy, factors affecting the cost and project cost data should be collected. Therefore, the first part of the research process was to gather existing information by literature review and to study the current practices through empirical study.

Table 2-1 Summary of research methods adopted in various phases of work

Stage of work	Data Collection Method / Decision method	Sampling Method	Sample size	No. of Response	Remarks
Exploratory study of project overheads cost					
Phase 1 Literature review	From library, electronic database, internet, interest groups	N/A	N/A	N/A	
Phase 2 Stage 1 survey	Questionnaire survey by post / fax	Sample all Group C contractors in H.K.	119	49 (41.1%)	
Follow-up interview after Stage 1 survey	Structured interview invited by phone calls	Sample the respondents in stage 1 survey	49	22 (44.9%)	
Stage 2 Survey	Questionnaire survey by post / fax	Sampled all Group C contractors in H.K.	119	40 (33.6%)	
Project overheads cost data collection	Data collection through invitations by letters	Sample all Group C contractors in H.K.	119	20	
Development of estimating model for project overheads					
Phase 3 Consolidate preliminary factor list affecting project overheads	1. Literature review 2. Focus group discussion	N/A Random sampling of Group C contractors	N/A 119	N/A 8	
Opinion survey on impact of factors affecting project overheads	Questionnaire survey by post / fax	Sample all Group C contractors in H.K.	109	79 (72.5%)	
Phase 4 Design estimating model	Qualitative analysis of different methods	N/A	N/A	N/A	Analyzed 3 methods : regression, fuzzy logic and ANN
Phase 5 Train the model	Data collection through invitations by letters	Sampled all Group C contractors in H.K.	N/A	71 cases	Trained 63 cases; Trained 6 models using different selection criteria; 3 rounds of re-training to reduce input factors
Test the model	1. Testing with separate set of real project data 2. Validate with other alternative methods	N/A N/A	N/A	N/A	8 production cases; Tested 6 models using different selection criteria ; 8 rounds of cross-validation; Validated with 2 other ANN algorithms and regression
Phase 6 Analyze the training and testing results	1. Compare validation results 2. Focus group review	N/A Invitation by phone to those who joined the focus group discussion in phase 3	N/A 8	N/A 7	

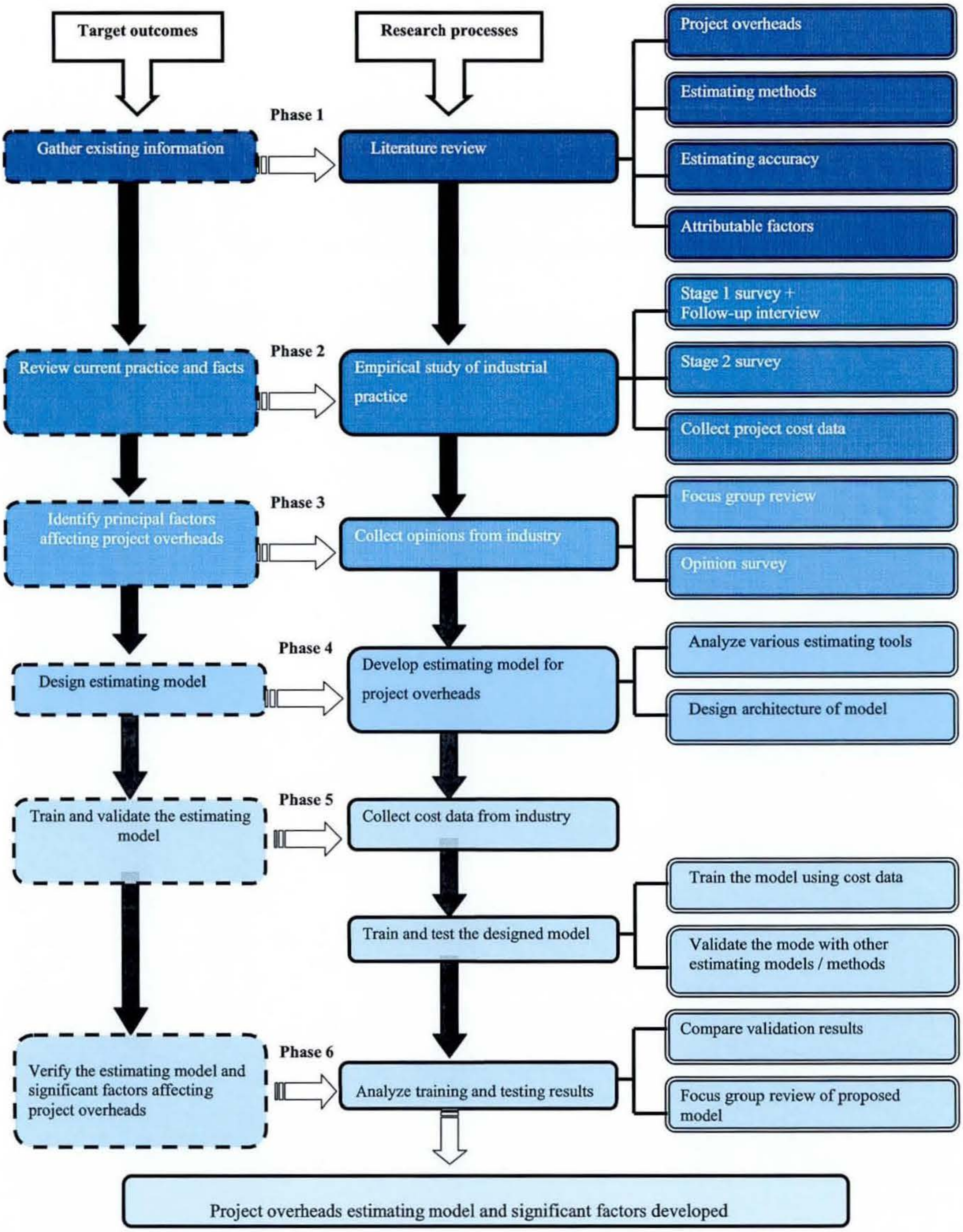

Figure 2-1 Flow Diagram of Research Activities

2.3.1 Phase 1 – Gather Existing Information by Literature Review

2.3.1.1 Subjects to Search and Review

At the start of the research process, compilation of information related to the research topic is fundamental. It is to gain sufficient understanding of the topic, and to identify any deficiency in the subject matter. As my research area is estimation of project overheads, literature review should therefore covered four primary aspects:

1. characteristics and estimation of project overheads,
2. estimating methods in construction industry,
3. contemporary cost estimating techniques, and
4. factors affecting project overheads.

Literature regarding definitions, nature and coverage of project overheads should all be reviewed. Estimating methods used in the construction industry was the second aspect to study. Classical methods are well-documented in most estimating text books and published guidelines of the professional bodies. These should be reviewed to establish the knowledge of the "standard" practice. With a comprehensive understanding of the standard methods used in the industry, it is easier to appraise the strengths and weaknesses of individual method. Furthermore, the estimating accuracy of project overheads should also be investigated to identify any room for improvement. In order to ensure an effective estimating model to be designed for the estimating task in question, contemporary estimating techniques or tools were also reviewed. In this way, any gap between the classical approach and contemporary methods can be identified more easily.

Based on the fact that all estimating or forecasting models have to be built upon the independent factors, identification of the critical factors attributable is a must. Therefore, factors affecting project overheads were also reviewed. To enhance the reliability and accuracy of the model, a broader spectrum of literature was examined, to include all the likely attributable factors as well.

As a result, besides studying the factors affecting project overheads, factors affecting project cost and duration were also reviewed as these two attributes are often used to estimate project overhead cost (Solomon, 1995; Geddes, 1996; Taylor, 1994 and Mansfield, 1983). Since the factors affecting project duration and cost are closely related to the project time and cost overruns, factors affecting the latter two were also studied. Finally, as cost estimating accuracy is highly influenced by the uncertainty involved (Ogunlana and Thrope, 1991), uncertainties involved in construction projects were studied in order to identify the factors leading to such. Therefore, four main groups of factors which might affect project overheads were examined and consolidated from the literature review:

1. factors affecting project overheads;
2. factors affecting project cost and duration;
3. factors affecting project time and cost overruns;
4. factors leading to construction risks.

In this way, the necessary information relating to project overheads and cost estimation were collected for the subsequent study.

2.3.1.2 Sources of Literature

Having identified the subject areas to review, sources of literature search were considered. With the facilitation of the world-wide-web, a much wider range of sources can be browsed across physical boundaries. There were four sources used to gather the information and are further described below.

First, published books, reports and dissertations were searched through the on-line catalogue system of the City University of Hong Kong. The University contains more than 844,200 volumes of books and 179,800 volumes of bound periodicals[3]. In addition, new books are recommended by each faculty or department annually (an approximate amount of HK$500,000 was assigned to my department every year to purchase the latest editions or new publications related to construction discipline). The holdings provide a good range of references related to construction management and economics.

Secondly, electronic databases were also browsed and reviewed within the huge electronic database platform of the City University of Hong Kong. The University subscribes a number of construction- or engineering-related databases which provide full-text references. The applicable databases included:

- Compendex,

[3] Information available from http://www.cityu.edu.hk/lib/about/intro.htm, website of Run Run Shaw Library of City University of Hong Kong.

- Science Citation Index Expanded,
- IngentaConnect, and
- ScienceDirect

Thirdly, internet was also surfed to find the other useful information related to the topic. Websites of professional bodies like the RICS, CIOB (Chartered Institute of Building), AACE International (Association for the Advancement of Cost Engineering), HKIS (Hong Kong Institute of Surveyors) and research organizations like the Hong Kong Productivity Council, CIB (International Council for Research and Innovation in Building and Construction), UK-ci.net, etc. were visited to look for the publications and projects funded by them. The researches conducted by them are usually applied researches in nature which closely align with the industrial needs.

Fourthly, interest groups on the internet were joined to exchange ideas of contemporary researches. Hong Kong academics interest group and the cnbr-group were joined. Useful information, particularly on the research methods and approaches, were shared amongst the members and thus enriched the pool of literature.

Since the research process lasted for several years, updating of the literature review was made at least once every year during the summer. This is to ensure that any new finding or development in the industry was included for consideration during the research process.

2.3.2 Phase 2 – Review Current Practice through Empirical Study of Industrial Practice

The main objective of this phase was to establish the current practice and associated facts related to project overheads estimation of the local construction industry. Although some of the information on current practice may be obtained from the secondary data like existing reports or researches conducted by other organizations, the data may not be taken in Hong Kong and may not aligned entirely with the research objectives of this study. Therefore, it is decided to conduct various surveys to collect primary data related to the local practice in project overheads estimation.

2.3.2.1 Information to be Collected

Similar to literature review, information regarding to the project overhead estimates, estimation accuracy and the estimating practice in Hong Kong was collected. Therefore, three categories of information was collected from the industry to develop a general understanding on the current practice:

1. Nature of project overheads estimation, e.g. estimation practice
2. Accuracy of project overheads estimation
3. Project overheads data from real projects.

The first and second categories of data are to provide the foundation for the development of the estimation model, and also to justify the importance and need of research in this topic. The project overheads cost data in the last category can provide the necessary information to understand the expenses pattern of each project overhead cost item.

2.3.2.2 Data Collection Methods

There are various methods to collect primary data, like survey, interview and focus group discussion. Since the information to be collected was aimed at reflecting the general practice of the industry, questionnaire survey was chosen to collect information on the nature of the estimating practice and the estimation accuracy of project overheads. Two separate surveys were implemented to give a better focus, one on overall estimating methods / practice and the other on estimating accuracy.

Whilst the estimating method adopted by the companies may be diversified, a follow-up interview after the survey on estimating practice was incorporated. Structured interviews with open-ended questions would be adopted. This is to allow more interactions made with the respondents when they answered the questions. Discussions about the participants' responses can be initiated to acquire more in-depth information; which is particularly useful for studies in an exploratory nature (Yegidis and Weinbach, 1996). On the other hand, separating the interview from the survey could simplify the questionnaire in Stage 1 survey so as to encourage more responses, especially when the sample was a group of experienced professionals who held senior positions in the company.

Therefore, two stages of survey and 1 follow-up interview were suggested as follows:

1. Stage 1 survey – collect information on estimating methods used amongst the construction companies
2. Follow-up interview after Stage 1 survey – explore the estimation methods used by respondents in greater detail

3. Stage 2 survey – collect information on the likelihood of estimating error in project overheads estimation.

For the project overheads cost data collection, invitations would be sent to the construction companies to explain the purpose of the research. This helped to relieve the concern of companies as cost data is often confidential. Project overhead cost breakdown from real projects were collected for analysis of distribution pattern.

2.3.2.3 Sampling Method

To have an accurate picture of the project overheads expenses pattern to build an estimation model, the most reliable approach is to obtain the information from the contractors. Thus, the cost data and accuracy opinions were collected from the contractors. In Hong Kong, the Works Bureau of the Hong Kong Special Administrative Region Government maintains a list of Approved Contractors and Suppliers for carrying out the public works. The Approved Contractors are classified into three groups: A, B and C. Among them, only Group C contractors are allowed to tender for public works of HK$50 million or more. This group covers all the large, local and international contractors in Hong Kong. Since the government is the largest developer in the construction industry, almost all the eligible contractors are registered as one of the approved contractors in the government list.

In typical projects with reasonable size, only the large contractors can be the main contractors to provide and manage the overall site facilitates; and have full control of them. The group A and B contractors are the small and medium-sized contractors. They are normally subcontractors in large projects, or the main contractors of smaller jobs like interior decoration of shops, refurbishment of commercial arcades, etc. Besides, the set up of these smaller contractors is usually loose with less formal documentation and accounting system. Therefore, having considered the purpose of the research and the characteristic of the industry, only the large contractors in Group C were surveyed or interviewed to collect information for the exploratory study.

The number of Group C contractors was quite stable, maintained around 100 - 110 throughout 2001 to 2005 (Works Bureau, Hong Kong Special Administrative Region). Since the group size was not large, all contractors within this group were sampled. As a result, the problem of non-random sampling did not exist. Since Stage 1 and 2 surveys were to explore the estimating method used and the respective accuracy amongst the large contractors, the Head of the Estimating Department was the most appropriate person to sample as he/she was in charged of the whole estimating process. They should have the best knowledge about the estimating methods used in their company and the overall estimating accuracy. For the later collection of project overheads cost data after Stage 1 and 2 surveys, consent had to be obtained from the Head of the Production Department as the required data (project overheads expenditure breakdown) is confidential and sensitive. Therefore, Production Department Heads were sampled at this point. The questionnaires were sent to the Department Heads of the Estimating Department or Production Department of the companies by post or by fax after enquiring of the company's receptionist the suitable method. For example, in the Stage 1 survey where general estimating practice of project overheads was studied, the questionnaires were sent to the Chief Estimators. Follow-up interviews were also arranged with them. In Stage 2 survey where the likelihood of inaccurate estimation of project overheads was studied, the questionnaires were also sent to the Chief Estimators of the contractors. When collecting the project overheads cost data, invitation letters were sent to the contract managers of the companies as the data involved were usually monitored by the project team rather than the estimators.

2.3.2.4 Responses

One hundred and nineteen Group C contractors were available in April 2002 when the stage 1 survey was conducted. Forty-nine valid responses were received, representing 41.1% response rate. In the follow-up interview after the stage 1 survey, twenty two participants are willing to join. Among the one hundred and nineteen Group C contractors in Stage 2 survey, forty responses were received, representing a response rate of 33.6%.

In the project overhead cost collection, cost breakdown from twenty projects were collected from the Group C contractors.

2.4 Development of Project Overheads Estimating Model

After collecting the existing information from literature review and latest facts from the industry, the research could proceed to the next part: development of the project overheads estimating model. According to Li (1995), three basic phases should be involved in developing cost estimating model: design phase, training phase and operation phase. Based on Li's suggested framework, the development of project overheads estimating model was divided into a four phase-model, with the design phase split into a two-tier approach : i) to identify the principal factors affecting project overheads, and ii) to develop the estimating model.

Therefore, the four phases involved were:

1) collect opinions from industry to identify the principal factors affecting project overheads
2) develop estimating model for project overheads
3) collect cost data from industry, train and validate the estimating model
4) analyze the results to make recommendations on the suitability of the proposed model.

2.4.1 Phase 3 – Identify Principal Factors Affecting Project Overheads

No matter which technique is used in developing a cost estimating model, identification of the attributable factors leading to the estimate must be ascertained. In view of the broad range of factors reviewed in the Literature Review (phase 1), scrutinization of them is necessary to ensure that the only the principal factors were included in the model. This was achieved by collecting views from the practitioners, followed by statistical analysis to generate a representative list of principal factors to predict project overheads. Figure 2-2 below illustrates the work involved.

2.4.1.1 Data to be Collected

The preliminary list of factors affecting project overheads was compiled from literature review. These factors included factors affecting project overheads, project cost and duration, project cost and duration overruns, and project risk. It was quite certain that some of the factors compiled might not directly affecting project overheads. On the other hand, the compiled list might have excluded some factors that were applicable to Hong Kong context only. Therefore, comments on the completeness and representativeness of the factor list were the prime information to be collected in this phase.

2.4.1.2 Data Collection Method

Although the preliminary list generated from literature review included most of the likely factors affecting project overheads, it was not ready for direct use in questionnaire. It should be reviewed first to screen out the unnecessary items and to check if there was missing of any possible factor. Therefore, two stages of work were involved here – a preliminary review and an opinion survey.

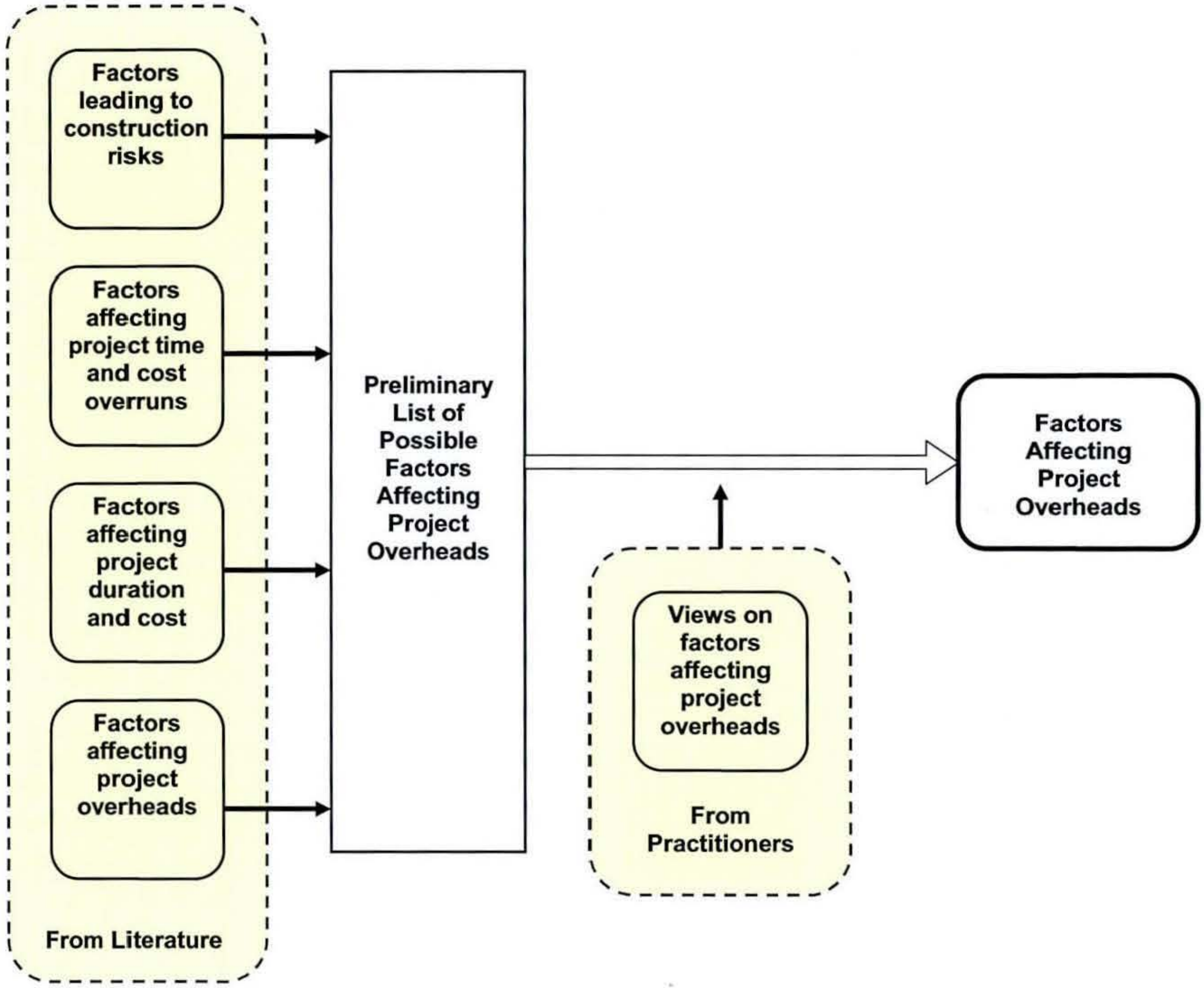

Figure 2-2 Development of Principal Factors Affecting Project Overheads

A review in the form of focus group discussion was chosen as it can allow a free exchange of comments amongst the participants. It would be suitable to review and make criticism on the completeness of the preliminary list. Whilst the number of participants in focus group is rather limited and there was a need to extract only the principal factors for the model, an opinion survey was designed to collect opinions regarding the perceived significance of each factor. As in Boussabaine and Elhag (1999) and others' work, only the meaningful / critical factors should be included in the estimation model. Therefore, the opinions collected from the survey would then be scrutinized by statistical analysis to extract the critical factors from the factor list. By doing so, principal factors identified could be used as the input factors in the subsequent estimating model.

2.4.1.2.1 Focus Group

Since the purpose of the focus group discussion was to compile a comprehensive list of possible factors affecting project overheads expenditure, involving people with a thorough understanding and experience in overseeing project overhead expenditure is crucial. As estimators only prepare estimates during the tender stage in the head office, their exposure to realistic project overheads spending is much less than the quantity surveyors who execute cost control during the construction stage. Quantity surveyors should know better how and why the project overheads are incurred in their projects. Therefore, experienced quantity surveyors (not estimators) working for the Group C contractors were invited as their cost control duties provide the suitable exposure to appreciate the critical factors affecting project overheads.

2.4.1.2.2 Opinion Survey

The main objective of this survey was to derive the principal factors affecting the expenditure of project overheads in construction projects. Based on the focus group discussion results, a questionnaire containing the screened factors was designed to survey the impacts of factors on project overhead expenses.

2.4.1.3 Sampling Method

2.4.1.3.1 Focus Group

As described before, the number of Group C contractors was only around one hundred and ten. Therefore, invitations were sent by post to all Managers of the Quantity Surveying Department of the Group C contractors. This can avoid the problem of non-random sampling. However, since focus group discussion requires more time from the participants, it was expected that the rate of acceptance would be very low. This was one of the reasons why an opinion survey was designed as the next step to collect generalized view.

2.4.1.3.2 Opinion Survey

Similar to the previous surveys done in the exploratory study, all the large Hong Kong contractors belonging to Group C were surveyed. This could ensure that the sample was familiar with the nature and impact of project overheads. Since this opinion survey was to collect views on the financial impact of each possible factor affecting project overheads, the sample should possess solid knowledge and experience regarding the accumulation of project overheads cost in a real project situation. Although estimators prepare many project overhead estimates, they have little experience in monitoring and controlling of the overheads expenditure. However, quantity surveyors are involved in the whole construction process to oversee the expenditure of different project overhead items. Only a close monitoring of the project cost can allow visualization of the extent of impact that each factor will bring to the final expenditure. Furthermore, personal bias might be difficult to detect or remove if the survey was conducted with the estimators as they may understate the level of inaccuracy. Therefore, unlike the previous surveys, the questionnaires were sent to the Quantity Surveying Manager / Chief Quantity Surveyor of these companies (not the Chief Estimator).

2.4.1.4 Responses

In the focus group discussion, eight senior quantity surveyors (with more than twenty years industrial experience) from the Group C contractors were willing to share their views on the factors affecting project overheads. The discussion lasted for two and a half hour.

When carried out the opinion survey on factors affecting project overheads, there were one hundred and nine Group C contractors in the List. Altogether seventy-nine responses were received, representing a response rate of 72.48%.

2.4.2 Phase 4 – Design Estimating Model

In this phase, the technique used to build the estimating model for project overheads was determined. The various estimating methods reviewed in the Literature Review phase could now provide a basis of choice. Different techniques have their own merits and drawbacks. Therefore, the choice of model was decided upon a qualitative analysis of the past performance and characteristics of each model.

After determining the technique used to build the model, the model architecture was designed accordingly. The principal factors affecting project overheads as found in the last phase would be incorporated into the model as the independent variables, whereas the project overhead cost was the dependent variable. Details of the model features will be discussed further in the Design of ANN Cost Estimating Model Chapter.

2.4.2.1 Qualitative Analysis Process

Before determining the use of artificial neural network as the modelling technique in this study, three techniques were analyzed qualitatively, including the fuzzy logic, artificial neural network and multiple regression. The first two are contemporary estimating / forecasting techniques widely used in recent years. The third method is a conventional statistical tool to predict the value dependent variable, based on the independent variables (George and Mallery, 1999).

Four basic criteria were used to analyze the suitability of the model for this research, including:

1) Characteristics of estimate (output)
2) Priori information required for the model
3) Accuracy of the model
4) Other special features of the model.

2.4.3 Phase 5 – Train and Validate the Estimating Model

Once the model was available, the data could be fed into it for training. Therefore, two major processes were involved in this phase: 1) collection of data from the industry and 2) training and testing the model.

2.4.3.1 Data Collection from Industry

After the type and architecture of the estimating model was defined, data related to the independent variables and dependent variable could be collected. In phase 3, the principal factors affecting project overheads were already identified. Data related to these factors was collected to serve as the independent variables or inputs of the model. The project overhead cost data was then used as the dependent variable or output.

2.4.3.1.1 Data to be Collected

There are two sources of data for the cost estimating model. The first is to collect real data measuring the input factors and the final cost. This method provides greater reliability and is adopted by most of the researchers (Vojinovic and Kecman, 2001; Emsley et al, 2002; Moselhi and Siqueira, 1998; Mosilhi, 1998; etc.). However, collection of real life data, especially cost data, is often sensitive and difficult. The second source is to use artificially constructed data for training and testing (Zhou et al, 1996). This approach can solve the problem of difficulty to obtain data, and is particularly applicable if knowledge regarding the function of cost estimation is available to the researchers (Zhou et al, 1996). However, the weakness of poor empirical base cannot be solved and thus not used by the majority of researchers. Despite the anticipated difficulty to collect cost data from the contractors, attempt was made to collect real project data for the estimating model.

Having decided to collect real data for the estimating model, the kind of data, or more appropriately, the measurement method of data, was established to enable all the data fitted well into the model. According to Weinbach and Grinnell (2001), there are four levels of measurement:

1. Nominal – value categories are discrete, e.g. measurement of gender, races, occupation.
2. Ordinal – value categories have distinct quantitative meaning and thus possible to rank orders. Example of an ordinal level of measurement: considerable, some, little, none. Likert scale also falls into this category.
3. Interval – value categories preserve rank order and unit differences. The categories are set on an equally spaced continuum. However, interval-level measurement does not have an absolute zero point, e.g. IQ score (0 score does not mean no intelligence).
4. Ratio – value categories preserve rank order, unit differences and have fixed zero point. Therefore, no negative values can exist, e.g. number of site visits per month, number of complaints over a 6-month period.

For the variables that were scrutinized from the literature, reference to the work of past researchers could be made to determine the way to measure in order to increase the validity. Nevertheless, no matter whether such a reference was available or not, the rules associated with the selected modelling technique must be adhered to. In other words, the measurement level of all the required data was determined according to the requirement of the estimating technique.

2.4.3.1.2 Sampling Method

Similar to the surveys done in the earlier stages, all the large Hong Kong contractors belonging to Group C were surveyed. This was to ensure that the project overhead cost of the sampled projects was under the full control of the respondents. Here, the required project data involved two categories of information: 1. data that are available or foreseeable in the tender stage e.g. gross floor area, contractor's past relationship with project parties; and 2. data that are unforeseeable in the tender stage, i.e. final project overhead cost. In this regard, estimators cannot provide the complete set of necessary information for model building purpose. Therefore, Quantity Surveying Managers of the Group C contractors were sampled. Invitations were sent to them by post to ask for project information. Telephone calls were made to those who had joined the opinion survey before, hoping to have a higher chance of response.

2.4.3.1.3 Responses

The data collection process spanned across two years from 2004 to 2006, successfully obtaining seventy-one project cases from the Group C contractors for training and testing.

2.4.3.2 Training and Testing the Model

Project data collected from contractors were presented to the model for training. Training rules and criteria were set with reference to the guidelines as set out in the manual of the modelling software. Upon completion of training, error of the model was examined to check whether the prediction was satisfactory.

To affirm the satisfactory performance of the proposed model, it must be validated. Several validation methods were proposed for this study, including:

1. Separate set of project data which was not used in any training process was set aside for validation of the model after training.
2. Cross-validation of the model by rotating the testing data with the training data.
3. Present the training data to other alternative models including conventional regression model.

The first method which used separate set of project data for testing is a standard method employed in a lot of modelling exercises, including ANN modelling (Ward Systems Group, Inc., 1995; Adeli and Wu, 1998; etc.). The errors from the training and testing sets are used as indicators to show the accuracy of the model. The second method, cross-validation, was adopted by Tam et al (2005) in particular to solve the problem of small sample size in ANN modelling. By rotating the training sets and production set of data, the proposed model was validated by a series of cross-validation trials to see whether similar results were produced. This method was used in this study to enhance the reliability of the model. The third validation method is commonly used by other researchers like Boussabaine and Elhag (1999) or Moselhi and Siqueira (1998). This is to check whether the proposed model is a better estimating tool than the other methods. In additional to some other ANN models, conventional regression equation would be built for validation as well.

2.4.3.2.1 Number of trials in training and validation tests

The 71 sets of data collected were split into 2 groups, one for training and the other for testing (called production). Six selection criteria were trained with the training set of data to obtain the best model. Besides conducting separate test for the model using the production set of data (for all 6 selection criteria), 8 rounds of cross-validations were also carried out to verify the reliability of the model. Furthermore, 3 other tests were carried out by feeding the data into multi-layered feedforward neural network, general regression neural network and multiple regression method.

2.4.4 Phase 6 –Verify the Estimating Model and Significant Factors Affecting Project Overheads

After the validations made in the previous phase, comparison of the validation results were carried out to check the suitability and accuracy of the proposed model. If higher accuracy was observed in the proposed model, it would be considered as a successful design for the estimating task. A final check was made by inviting professionals in the industry, who were potential users of the model, to appraise the applicability of the model. The model was ready for use if the comments were satisfactory.

2.4.4.1 Data to be Collected

The main objective of the appraisal was to ensure that the model was a practical solution to the estimation problem irrespective to the accuracy or reliability concern. Evaluations on the practicality of the model were thus to be collected.

2.4.4.2 Data Collection Method

Since the industry was unlikely to be familiar with the proposed model, focus group discussion was chosen to allow interactive explanation to be made to the participants. A group of experienced quantity surveyors would be invited to appraise the applicability of the model in lieu of the existing estimating practice (which was based heavily on professional judgment). It is understood that the estimating function within a construction company is handled by estimators, not quantity surveyors. However, if estimators were asked for their opinions regarding renouncing their current estimating practice due to their unsatisfactory accuracy, biased comments are likely to be received. Therefore, to reduce the likelihood of personal bias, quantity surveyors were invited as they are more impartial about the subject matter. Besides, all of them possess professional knowledge in an estimating context which enables professional comments to be made.

2.4.4.3 Sampling Method

The eight senior quantity surveyors who had joined the previous focus group discussion on the factors affecting project overheads were invited first to provide consistency of comments. Invitations were made by phone calls.

2.4.4.4 Responses

Seven out of eight quantity surveyors were willing to join the focus group review. The overall discussion lasted for one hour and fifteen minutes.

2.5 Summary

In this Chapter, the overall methodology of the research is illustrated in Figure 2-1. A 6-phase process was developed, with the first two phases aimed at exploring the nature and accuracy of project overheads estimation. Literature review on a broad range of topics was carried out to obtain the basic knowledge of the subject. Questionnaire surveys were then developed as the major tool to collect data from the industry, so as to enable a more in-depth understanding of the estimating practice of project overheads in the local industry.

Phases 3 to 6 were designed to develop the estimating model for project overheads. Preliminary list of factors affecting project overheads were consolidated through analysis of past literature, followed by focus group discussion with experienced quantity surveyors to check for any possible factor missing. Opinion survey was then carried out with quantity surveying managers in contractors to collect views on their perceived impact of the factors affecting project overheads. The collected data could be analyzed statistically to extract the principal factors as inputs to the estimating model.

A suitable modelling technique was chosen through a qualitative analysis of the alternative methods in phase 4. After determining the technique used, the model was designed accordingly. Moving to phase 5, the nature of data, in terms of the level of measurement, was decided before proceeding to the data collection stage. The required data related to the principal factors (independent variables) and the project overheads cost (dependent variable) was then collected from the industry. The collected data was subsequently fed into the proposed model for training and testing. Various validation methods including testing with a separate set of real data, cross-validations and comparison with other estimating models were implemented. In the last phase, results from the validations were reviewed. The accuracy of alternative models was compared with those produced by the proposed model. To ensure applicability of the model, the model would be appraised by practitioners in the industry in a focus group discussion.

CHAPTER 3

Literature Review

3 Literature Review

3.1 Introduction

Construction cost estimation refers to the technical process of predicting costs of construction (CIOB, 1997). Project overheads belong to the cost category that covers the site cost of administering a project and providing general plant, site staff, facilities, site-based services and other items not included in all-in-rates (CIOB, 1997). In the cost estimation process, project overheads have to be ascertained in order to arrive at the overall cost of a project.

Having established the rationale of study and research hypothesis in the Introduction Chapter, the past literature related to cost estimating techniques and project overheads estimation is reviewed in this Chapter. Firstly, the cost estimating theories and classical techniques are examined. Following is a detailed examination of the nature of project overheads including its characteristics and risks involved. Then, the principle methods of project overheads estimation and the estimating methods in practice are presented.

In the past, cost estimation was mainly a mathematical exercise. However, with the advancement in artificial intelligence and statistical modelling, a lot of researches are conducted aiming to improve the accuracy and efficiency in cost estimation. Therefore, after going through the general estimating methods, contemporary cost estimation models will be examined in depth. The purpose is to explore the possibility to apply artificial intelligent techniques to the estimation of project overheads. Since the development of a new statistical model for estimating project overheads must require the definition of input variables at the outset, literature on the possible factors affecting project overheads are explored at last.

3.2 Classical Construction Cost Estimating Theories

At different stage of the project life-cycle, cost estimating may be requested for different purposes by various parties. For instance, at the inception stage, the project owner may require a preliminary estimate of the overall cost of the proposed development. Contrasted with the preliminary estimate is the detailed estimate prepared by the consultant quantity surveyor upon the full set of design information available. Ostwald (2001) illustrated the nature of estimate precisely according to five characteristics:

- purpose
- accuracy
- timeliness
- effort
- accountability.

Based on different estimating purpose and accuracy requirement, timeframe for the estimating exercise, estimating effort available and accountability of the estimate; the estimating method used or available will be varied. For example, estimates can be prepared for a contractor's bid, or for a developer's budget / project funding purpose. In terms of accuracy, different firms may require different levels of estimating accuracy as performance indicator. Besides, the estimating time and effort available to the estimators or consultant quantity surveyors may also dictate the estimating method chosen, as more accurate estimate normally requires more time and effort. Finally, estimates can be prepared by the project owner, contractor, subcontractor, etc. Due to the difference in accountability, the estimating method used will be different to allow modifications to the estimating process.

Nevertheless, irrespective to the purpose and accuracy of estimate prepared, the method of preparing cost estimates can be summarized into four main types Ostwald (2001):

- unit rate approach
- comparison approach
- opinion approach
- probabilistic approach

3.2.1 Unit Rate Approach

This method is by far; the most popular and traditional method introduced by text books or established standard procedures. A lot of books (e.g. Brook, 1998; Geddes, 1996) detailing the method of estimating the unit rates are in fact, based on (or similar to) the Code of Estimating Practice (CIOB, 1997). According to the Code of Estimating Practice, 1997 (p. 79 - 105) issued by the Chartered Institute of Building (CIOB), preparing an estimate would involve three stages: establishment of the all-in-rates (for labour, plant and materials), the incorporation of rates from specialists and sub-contractors, and calculation of the net unit rates. Although the Americans do not adopt the SMM, they have similar approach to breakdown the construction works / costs, either based on the UNIFORMAT II[4] or the MasterFormat (developed by the Construction Specifications Institute (U.S.)) (Holm et al, 2005). Similar to the U.K. Code of Estimating Practice, the construction cost is built up by calculating the unit rates of associated elements (e.g. walls, doors) which is called the "assemblies approach" (R.S. Means, 2001; Pratt, 2004). Therefore, the method of cost estimating recommended is the same as the CIOB Code.

To explain further, the unit rate approach is to establish the rates for different factors of production. In construction project, different factors of production are required, including labour, plant, materials, sub-contractors and specialists. In short, as depicted in equation (3-1), the relevant all-in-rates of labour, plant, and material, together with the sub-contractors' quoted unit rates (if that item of work involves sub-contracted portion), overheads and profit allowance will be added to form the net unit rate for a particular item of work.

$$R_i = R_l + R_p + R_m + R_{sc} + O_s + O_{ho} + p \qquad \textbf{(3-1)}$$

Where R_i = net unit rate of work item i, dollars per unit of work

R_l = all-in rate for labour (per unit of work)

R_p = all-in-rate for plant (per unit of work)

R_m = all-in-rate for materials (per unit of work)

R_{sc} = cost of hiring subcontractor per unit of work

O_s = site overheads allowance per unit of work (where applicable)

O_{ho} = head office overheads allowance per unit of work (where applicable)

p = profit allowance per unit of work

$$C = \sum R_i \times Q_i \qquad \textbf{(3-2)}$$

Where C = total cost of project

Q_i = quantity of work item i

Depending on the scale and complexity of project work, overheads can be allowed separately. The estimate developed from this method is detailed in nature and is most widely used by contractors when preparing their bids for tenders.

[4] Original version developed by the U.S. General Services Administration, current version enhanced by American Society for Testing Materials.

3.2.2 Comparison Approach

This is another approach commonly described in literature due to its reasonable level of accuracy and lesser requirement of complete design information (Ostwald, P., 2001, CIOB, 1997). Especially in the early stage of project development when complete design information is not available, this estimating technique provides a useful tool for estimators to set budgets and assess the feasibility of the scheme, or to plan for necessary funding requirement for the developers. In other words, this method is more commonly used to set developers' budgets, rather than to prepare contractors' tender bids.

This method is often regarded as "approximate estimating" as the level of accuracy is generally lower than the unit rate approach. By utilizing historical data of earlier comparable projects, the project cost is estimated by adopting the cost of the earlier schemes. Approximate estimate in three main types are developed (CIOB, 1997) based on the historical data available:

1. superficial area;
2. functional unit;
3. elemental cost.

For superficial area method, cost per floor area of earlier comparable scheme is used to estimate the cost of the current project. This is the most popular method due to its convenience and simplicity to use. Cost per floor area of past comparable scheme is used to multiply the total floor area of the current project to arrive at the total cost estimate. Where differences should be found between the earlier scheme and current project, adjustments will be made to take into account the significant differences. Typical adjustments include location, specification, degree of complexity, size, shape of building, ground conditions, number of storeys, etc. Cost of different building types can be conveniently stored in company's database for future use.

In functional unit method, historical data is available in the form of "cost per functional unit". The functional unit, for instance, is a room in hotel development, a bed in hospital, or a car park space in multi-storey carpark. The total cost estimate can be found by multiplying the unit cost of earlier scheme with the total number of functional units in the current project. As with other approximate estimating methods, adjustments will be made to take into account the differences between earlier scheme and current project.

In the elemental cost method, cost per floor area of each building element e.g. external walls, wall finishes, fire services system, etc. is developed as the basis of calculation. Compared with the other two approximate estimating techniques mentioned earlier, this method requires a more detailed breakdown of the total cost (at least to the various building elemental costs). Due to this requirement, the level of detail regarding the project and cost information is higher; and the accuracy can also be enhanced. Cost adjustment can be made reasonably easy by comparing any difference in the design and specifications between the earlier and current projects.

3.2.3 Opinion Approach

This method is not highlighted by the popular text books, but it is indeed one of the common methods used in practice by estimators (Akintoye and Fitzgerald, 2000). Ostwald, 2001 incorporated this method as one of the methods in estimating construction cost and acknowledged its flexibility. This estimating method relies on the estimator's opinion and professional judgment. Although the relative accuracy may be lower, this method may be the only option especially in the absence of historical cost data and shortage of time.

3.2.4 Probabilistic Approaches

Probabilistic approaches are alternative methods to estimate cost and have gained recent support by some researchers (Akintoye and Fitzgerald, 2000). According to Ostwald's book (2001), probabilistic estimating requires the estimator to give a probability estimate of each uncertain event. If the event is certain, the associated probability will be equaled to 1. Sometimes, two events are mutually exclusive; and the sum of the probability of the two events will be 1. The techniques used to judge / estimate the probabilities are usually based on professional judgment (which is subjective) or analysis of historical data to give relative-frequency interpretation (which is objective). More simple and common probabilistic methods are briefly explained below.

3.2.4.1 Range Estimating

There are probable errors in the estimate. By extending the single-valued estimate to an estimate range provides more information for decision-makers to know the chance of exceeding (or below) the most likely cost. This is particularly the case when contractor estimate for submitting their bids.

The principle behind is every cost estimate is build up from a number of cost element which are independent. Each cost element is assumed to correspond to a beta distribution which can be skewed or normal. Skipping the complex mathematics of the beta probability density here, the equations to calculate the expected cost are in fact quite simple. Figure 3-1 below is used to illustrate the equations.

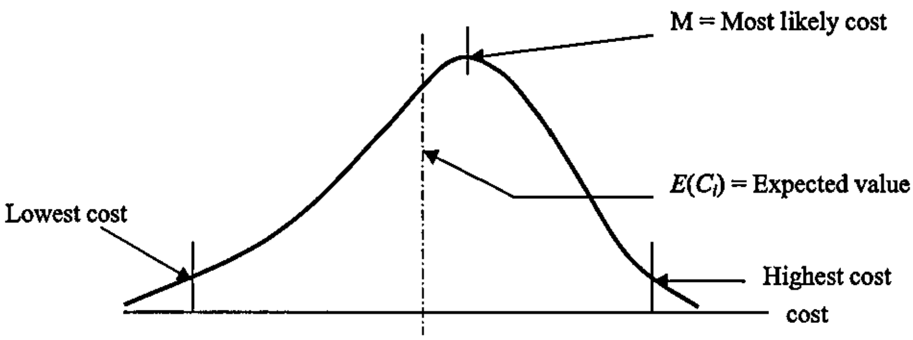

Figure 3-1 Location of Estimates for Range Estimating Method

According to Ostwald (2001), the mean and variance of the single cost element are :

$$E(C_i) = \frac{L + 4M + H}{6} \qquad (3\text{-}3)$$

$$Var(C_i) = \left(\frac{H - L}{6}\right)^2 \qquad (3\text{-}4)$$

Where $E(C_i)$ = expected cost of distribution, i = 1,2,3, ...,n

L = lowest cost, or best-case estimate of cost distribution, i

M = modal value, or most likely estimate of cost distribution, i

H = highest cost, or worst-case estimate of cost distribution, i

$Var(C_i)$ = variance of cost distribution i, i = 1,2, 3, ...,n

If all the cost elements are estimated in the same way and added together, the new distribution of the total cost will be approximately normal (following the central limit theorem). The mean of the sum is the sum of the individual means, and the variance is the sum of the variances.

$$E(C_r) = E(C_1) + E(C_2) + \ldots + E(C_n) \qquad (3\text{-}5)$$
$$Var(C_r) = Var(C_1) + Var(C_2) + \ldots + Var(C_n) \qquad (3\text{-}6)$$

Management can then assess the chance or risk of exceeding the expected value of the total cost by the following equation :

$$Z = \frac{UL - E(C_r)}{[Var(C_r)]^{1/2}} \qquad (3\text{-}7)$$

Where Z = Z value of the standard normal distribution

UL = upper limit of cost, selected by decision-makers

For instance, $Z = 0$ if upper limit is 50% above the expected cost.

This method allows management to have more understanding of the risk of estimates. More information regarding the bids and cost information is available, thereby improving the accuracy of the estimate.

3.2.4.2 Monte Carlo Simulation

Monte Carlo simulation selects random numbers that correspond to a probability distribution, assuming the cost variables are independent to each other. Such distribution can be developed from historical data. It has the advantage of objectivity though the historical data is not always available. Therefore, an alternative means is to choose from some known patterns like normal, beta, rectangular and triangular (Smith et al, 1999). Before choosing the shape of distribution, careful consideration should always be made to the likely result of that variable.

By fitting the random number to the predetermined cost function of the project, the cost can then be found. Each of this process is called iteration. The distribution of the cost values generated from the iterations may be a jerky histogram or a normal distribution, depending on the number of iterations conducted. Hence, it is common to carry out 1,000 iterations for an average project to ensure that the results are free from most statistical biases (Smith et al, 1999). By applying the central limit theory in a normal distribution, the mean cost value, standard deviation and confidence interval, and other statistical properties can be found.

3.3 Overview of Project Overheads

After going through the traditional cost estimating theories developed for the construction industry, the estimation method of project overheads can be studied. However, before appraising the suitability and efficiency of the existing method used in estimating project overheads, an in-depth understanding of the literature about nature of project overheads is necessary. Therefore, the following section will portray the characteristics of project overhead costs, followed by the theoretical estimating method for them.

3.3.1 General Nature and Characteristics of Project Overheads

Project overheads belong to the cost category that covers the site cost of administering a project and providing general plant, site staff, facilities, site-based services and other items not included in all-in-rates (CIOB, 1997). They are the costs specific to a project, but not specific to a trade or work item (Assaf et al 1999), for instance, insurance, site cleaning, and hoisting equipment. Other names for project overheads include "preliminaries" (Brook 1998, Geddes 1996, Kwakye 1994, Cooke 1984), "job overheads" (Ostwald 2001, Dagostino and Feigenbaum 2003), "project indirect cost" (Bartholomew 2000, Mansfield 1983), and "general expenses" (Pratt 2004). No matter which name is being used, the understanding of the project overheads coverage is consistent across the world.

3.3.1.1 Typical Project Overhead Items

Although the exact requirements of project overheads differ amongst projects in practice, it is not difficult to cite some typical items covered in project overheads, e.g. staffing, plants, site accommodations and facilities, and other contract requirements. The Standard Method of Measurement Ed. 7 (SMM7) provides a good reference list for these items (CIOB 1997). Most of the text books follow or expand from the SMM7's list. It is also worth to note that the MasterFormat developed in the U.S. (standard format used in most construction contracts in the U.S.) does not include a clear breakdown of the project overheads. Only some items are included in the Division 0 (Bidding & Contract Requirements) and Division 1 (General Requirements) of the MasterFormat, but most of them are not found anywhere in the specifications. Nonetheless, it is understood that the project overheads have to be allowed by the contractors when pricing even they are not explicitly identified in the MasterFormat (Pratt, 2004).

Typical project overhead items mentioned by the text books are summarized as follows:

Staffing and Plants

- Site staff: provision of site management staff for the supervision of the project.
- Plants and equipment: rental, mobilization, running and maintenance of plants and equipment, e.g. tower crane and hoist.
- Non-mechanical plants: provision and mobilization of non-mechanical plants like scaffold, theodolites, bar benders, etc.

Temporary site accommodation

- Temporary site offices, sheds, workshops: provision and future dismantling of the site offices for contractor, resident designers and sub-contractors, workshops and storage area for materials.
- Temporary site enclosures: provision and future dismantling of the hoardings, fencings and the like.
- Temporary access and structures: provision of temporary access roads, bridges and the like where necessary.
- Latrines: provision and removal of temporary sanitary facilities.

Temporary utilities

- Temporary water and power: connection and supply of temporary water and power.
- Site communications: connection and supply of telephone, fax and internet service.

Contractor's obligations

Items to be covered vary with different projects and the followings are only some common items mentioned in the references.

- Insurances and bonds: provision of workers' compensation, contractor's all risk insurance, third party insurance, as well as the surety bonds as required by the contract.
- Statutory fees: payment of levies and fees as stipulated by local regulations.
- Safety equipment: provision of safety equipment.
- Samples, mock ups: submission of samples and mock ups for designers' approval.
- Testing: laboratory and field tests as required by the contract conditions.
- Site cleaning: progressive cleaning and final cleanup of the site.
- Protection: protection to finished works and adjacent buildings.

- Site security: provision of watchmen, personal I.D. cards, security checks and the like to safeguard the works, materials and plants against damage and theft.
- Progress photos: provision of progress photos and report for project owner and designers.

3.3.1.2 Project Overhead Items in Hong Kong Construction Contracts (according to the Hong Kong Standard Method of Measurement)

Whilst there is a common understanding of the nature of project overheads all over the world, the exact items found in contacts of different countries can vary due to difference in for example, legal, climatic, geographical, and cultural settings. Although there is no formal literature on estimation of project overheads in Hong Kong, the professional bodies in Hong Kong have published the Hong Kong Standard Method of Measurement (HKSMM) for use in all building projects. The HKSMM thus provides a good reference to the typical project overhead items found in construction contacts in Hong Kong. Table 3-1 below shows the summary of project overhead items found in the Preliminaries Section of the HKSMM (1979).

Table 3-1 Project Overhead Items in the Preliminaries Section of the HKSMM

Preliminary Particulars		17	Injury to persons and damage to property
1	Employer, Architect and Quantity Surveyor	18	Insurances, etc.
		19	Provisional and Prime Cost Sums
2	Description of works	20	Conditions of payment
3	Site and inspection	21	Surety or bond
4	Division of work into sections	22	Watching
Conditions of Contract		23	Protection
5	Form of Contract	24	Treasure trove, coins, etc.
6	Particulars to be inserted in Appendix to Schedule of Conditions	25	Variations and methods of measuring and Valuing
General Matters		26	Samples
7	Working hours, rates of wages, etc.	27	Testing of materials
8	Plant, tools, sheds, etc.	28	Water
9	Notices and fees	29	Lighting and power
10	Safety precautions	30	Attendance
11	Industrial Training Levy	31	Hoardings, etc.
12	Setting out	32	Works by Public Authorities
13	Foreman	33	Temporary roads
14	Protection of public property, etc.	34	Drying the building
15	Sub-letting	35	Removal of rubbish
16	Artists or Tradesmen not Sub-contractors	36	Defects after completion

There are some differences found between the HKSMM and the SMM7 (U.K.), in terms of the items included in project overheads. For instance, there is Industrial training levy in Hong Kong (item 11). On the contrary, the professional fees are not paid by the contractors in Hong Kong, and there is no special overheads incurred for winter construction as well. Furthermore, land acquisition is not done by contractors in Hong Kong and thus there is no such provision in the project overheads.

3.3.2 Expenses on Project Overheads

Although the project overheads provision may vary from project to project, the identification of its approximate percentage contribution to the overall project cost can provide a good reference to the estimators. However, literature gives diverse recommendation and is difficult to conclude an indicative figure in this aspect.

According to most text books, project overheads are normally not more than 15% of the total project cost (Pratt 2004, p.323, Brook 1998 p. 189, Adrian 1982); typically in a range of 6% to 15% (Mansfield 1983, p. 187) or 7% to 15% (Geddes 1996, p.36). However, Assaf et al (1999) found that 52% of the Saudi Arabian contractors had their project overhead cost higher than 15% in their survey. The average ratio of project overheads to total project cost found from the 61 contractors was 16.8%. In another survey done by Hegazy and Moselhi (1995) with 90 Canadian contractors, the project overheads cost ranged from 10% to 30%.

From the different percentages suggested in text books and surveys, it is evidenced that there is a substantial deviation between the actual range of project overheads percentage and our classical thinking. Furthermore, the lack of empirical study in this area also casts doubts to the reliability of such indicators. This can create great trouble in practical estimation to some contractors when only a mark-up is applied to the total cost to allow for project overheads.

3.3.3 Distribution of Project Overhead Costs

Very little existing literature goes into detail of the project overhead cost distribution. Even some of the text books attempted to give an indicative ratio of the project overhead cost to total project cost, none of them discussed about the distribution pattern of project overhead costs. Amongst the very few studies conducted, an early work of Solomon (1993) analyzed the project data of a large QS consultant firm in the U.K.. He identified four major items in the project overhead costs, according to their percentage distribution: staffing (26.5%), mechanical plant (22.3%), access & scaffolding (18%) and site office and the like (11.8%). These four items accounted an average of 78.6% of the total project overhead costs. Other items included:

- power (6.3%),
- site cleaning and clearance (4.8%),
- telephone / fax (2.1%),
- temporary roads (1.8%),
- hoardings and signs (1.7%),
- watching and security (1.7%),
- insurances (1.7%), and
- all other items (1.3%).

In a latter survey done by Assaf et al (1999) with the Saudi Arabian contractors, similar results are obtained and prioritized according to their percentage contribution to the total project overhead costs in descending order:

1. staffing cost,
2. plant and equipment cost,
3. temporary construction,
4. insurance, taxes, and bonds.

Although the average percentages of distribution are not stated by Assaf et al, the first three major items identified in Assaf's survey are the same as those highlighted by Solomon (1993). This gives an indication that staffing cost, plant and site accommodations are the major items that account for a large proportion of project overhead costs.

3.4 Estimation Theory on Project Overheads

From the previous sections, it can be seen that up to this point of time, there is no definite answer to the overall percentage contribution of the project overheads to the total project cost. Furthermore, knowledge about the distribution of project overhead costs is also limited. Nevertheless, most of the text books on construction cost estimation provided a detailed explanation on the principle method to estimate project overheads. The general method is introduced below based on the CIOB Code of Estimating Practice (1997), supplemented with other text books where appropriate.

The theory behind is that every project overhead item is made up of time-related costs e.g. rental charges, salary and/or fixed costs e.g. installation, dismantling costs. The fixed cost can be built up by adding all the costs concerned, just the same as estimating the unit rates of direct works. Some fixed costs are related to the total volume of work e.g. insurance premium, and hence a simple percentage of the total cost of project can be applied instead. Since the theory suggests that a lot of project overhead items are time-related, a project summary schedule (in the form of Gantt chart) becomes an integral step before estimation in order to identify the time-related costs outlay (Holm et al, 2005; Pratt, 2004; Dagostino and Feigenbaum, 2003; Brook, 1998; CIOB, 1997; Mansfield, 1983).

The theoretical approach to calculate each typical project overhead item is summarized in Table 3-2. Most of the text books suggested very similar methodology of estimating project overhead items. Hence, the CIOB Code is adopted as the basic reference for explaining the principle method of estimation.

Table 3-2 Summary of estimation method for typical project overhead items (based on CIOB Code 1997)

	Project overhead items	Time-related	Volume-related	Fixed cost
1	Staffing	✔		
2	Plant and equipment	✔		Mobilizing cost, erection cost, dismantling cost
3	Non-mechanical plants	✔		✔
4	Site offices and storage	✔		Mobilizing cost, erection cost, dismantling cost
5	Temporary site enclosures			✔
6	Temporary access & structures			✔
7	Latrines	✔		Mobilizing cost, erection cost, dismantling cost
8	Temporary power and water	✔		Installation cost
9	Site communications	✔		Installation cost
10	Insurances & bonds		% of total project cost	
11	Statutory fees			✔
12	Safety equipment			✔
13	Samples, mock ups			✔
14	Testing			✔
15	Site cleaning			✔
16	Protection			✔
17	Site security	✔		✔
18	Progress photos*			✔

*Note: the item is not highlighted in the CIOB Code but covered in most other text books only.

1. Staffing
   - All of the books suggested that staff salary including all costs associated with employing the staff (e.g. pension schemes, allowances, etc.) should be estimated. The duration of the supervision requirement is used to extend the total cost of this item.
2. Plant and equipment
   - The all-in rate of hiring the plants (including the plant operators) has to be estimated. Large mechanical plants like tower crane and material hoist involves mobilization, erection and dismantling costs as well and these costs should be included.
3. Non-mechanical plants
   - Rental charges are incurred in most non-mechanical plants like theodolites. Some items like scaffolding can be a sublet item and the estimate can be calculated from sub-contractor's quotation. Other ancillary cost like mobilization cost may also involve. However, small hand tools can be estimated as a percentage of total labour price.
4. Site offices and storage
   - The site offices may be hired, purchased or available from contractor's own stock. Therefore, hiring cost on a time basis, or transportation, erection, fitting out and decoration, subsequent dismantling and reinstatement cost may be incurred. Except for the rental charges, the rest of the cost will be in the form of fixed cost.
5. Temporary site enclosures
   - The cost of erecting and removal of site enclosures; and latter reinstatement of the site has to be estimated.
6. Temporary access and structures
   - Temporary access and hardstandings has to be carefully planned to make the best use of the existing and future permanent roads. Cost has to be allowed for fixing and future reinstatement or making up levels when the works is completed.
7. Latrines
   - Similar to the site offices and sheds, cost has to be allowed for hiring of portable toilets (which is on time basis), as well as transportation, erection and

removal of them. Furthermore, waste hauling from portable toilets has to be estimated as a fixed cost.

8. Temporary power and water
   - Temporary power and water consumption is usually estimated as a time-related cost plus an estimate of installation fixed cost. Dagostino and Feigenbaum (2003) and Pratt (2004) suggested that a historical monthly allowance for the temporary power was an acceptable practice. However, Geddes (1996) suggested that the cost of temporary water supply could be estimated by a percentage to the total cost of work (volume-related) which was an acceptable and usual method.
9. Site communications
   - Historical cost on monthly phone charges is usually estimated, together with the installation cost of the system. The internet connection can also be estimated in the similar manner.
10. Insurances and bonds
    - It is usually recalculated after the final review meeting as a percentage of the contact value.
11. Statutory fees
    - Exact method to calculate the statutory fees is not described in the CIOB Code. However, a fixed sum is suggested to be estimated by Geddes (1996).
12. Safety equipment
    - The extent of safety equipment, including the number and type of safety equipment required, should be estimated and priced.
13. Samples, mock ups
    - The number of samples and mock ups should be estimated with reference to the contract requirement and priced accordingly.
14. Testing
    - The charges for tests are usually estimated from testing labs' quotations.
15. Site cleaning
    - Method of estimating this item is not clearly explained in the CIOB Code. According to Pratt (2004), fixed cost like constructing and removal of the refuse chutes should be incurred. While daily cleaning may be a sublet item, it can be estimated based on sub-contactor's quotation. Alternatively, Dagostino and

Feigenbaum (2003) suggested to estimate the number of loads per day. Having the quotation for each haul, the total cost could be estimated. A lot of authors suggested that final cleaning was usually estimated on a basis of a price per unit gross floor area (Pratt, 2004; Dagostino and Feigenbaum, 2003; Mansfield, 1983).

16. Protection
    - Protection to adjacent buildings is critical and can be measured as a fixed cost. For protection to new works, fixed cost has to be estimated as no one will admit to damaging the works in most cases.
17. Site security
    - This item is basically covered in the watchmen item (under site supervision) in the CIOB Code and hence can be estimated as a time-related cost. Other books recommended site security to cover security system other than watchmen like personal I.D. cards, security checks, and the like. The set up cost and running cost for such has to be estimated according to the sub-contractor's quotation (Pratt 2004).
18. Progress photos
    - Progress photographs are not specifically highlighted in the CIOB Code but are mentioned by some other text books as typical requirements in the project (Dagostino and Feigenbaum, 2003). Although digital cameras are common and convenient, allowance has to be made for developing selected photographs for submission purpose. This is usually estimated as a fixed cost. However, Mansfield (1983) suggested a monthly cost to be estimated for the progress photographs as they were usually required at a certain prescribed interval e.g. monthly.

On the contrary to the detailed estimation of project overheads as suggested above, some authors also recommended a simple approach to estimation of project overheads – estimate as a percentage of total project cost based on historical data (R.S. Means, 2001; Holm et al, 2005). As stated in the CIOB Code of Estimating Practice (1997, p. 146), this method of applying a fixed percentage to the total value for project overheads allowance was particularly common in case of small-scaled, repetitive works. However, this may result in under-estimation as many project overhead items bear no linear relationship to the value of works.

Therefore, it is clear that most of the text books recommended a detailed approach to estimate project overheads. They also acknowledged the importance of detailed estimation of project overheads (Dagostino and Feigenbaum, 2003). Taylor (1994) put forward an interesting scenario that if contractors sublet all or most of the work to sub-contractors, two similar companies would obtain equal and off-setting bids from sub-contractors. Then, the only difference would be their markups for overheads and profit. Pratt (2004) stated clearly that even though the project overhead items were not listed clearly in the CSI MasterFormat (U.S.), these unlisted items were indispensable and careful assessment could be critical to the competitiveness and financial success of the project. Ostwald (2001) also emphasized that undersestimating or overestimating of overheads was serious in view of the proportion to total job cost. He explained that while the ratio of fixed cost to variable cost had risen in recent decades (like the ratio of equipment to worker), distribution of overheads by a single rate was misleading. Park (1992) also argued that estimating project overheads as percentage of direct job costs could be satisfactory only when the contactor performed the same type of work all the time, and who maintained a stable work load.

3.5 Cost Estimation in Practice

Sections 3.2 and 3.4 detailed the classical theories in estimating general construction cost and project overheads. In practice, theories may not be followed. This section is to explore the empirical studies related to estimating methods adopted by the practitioners to estimate general construction cost and project overheads. Besides, it is understood that accuracy is an important indicator of the performance of a cost estimation system. Without studying this aspect, the analysis of estimation methods will be meaningless. Therefore, studies related to the estimating accuracy of construction cost and project overheads are also reviewed.

3.5.1 Construction Cost Estimation in Practice

3.5.1.1 Overall Cost Estimation Methods Used by Practitioners

Although all of the literature recommended detailed estimation like unit rate approach should be used in contractors' bid preparation, empirical findings from a lot of researches prove that the theoretical methods are not followed entirely. In a survey done by Hegazy and Moselhi (1995) with ninety Canadian contractors, 92% of them estimated the direct cost in a detailed manner. 4% estimated labour cost only and added a percentage for materials and equipment. The remaining 4% only made a rough estimate based on their experience. So, there is an overall of 8% contractors using a more experience-based approach to estimate their bids. However, in a survey done by Skitmore and Wilcock (1994) with small contractors in the U.K., up to a 50% of the bill items were priced by non-detailed methods based on their own judgment.

In a more recent survey done by Akintoye and Fitzgerald (2000) with some U.K. contractors, the most popular methods used was the "estimating standard procedures" (mean = 4.869 measured by a 5-point likert scale of the extent of use). The second and third common methods were "comparison with similar projects based on personal experience" (mean = 3.919) and "comparison with similar past projects based on documented facts" (mean = 3.643). The "usage of probabilistic techniques" ranked the eleventh (mean = 1.651) and the "use of complex statistical formula" ranked the last (fourteenth). The results indicate that although unit rate approach is the most popular method, the comparison approach seemingly receives wide acceptance amongst the contractors as well. The probabilistic approach, though suggested to be alternative methods with high accuracy, remains unpopular due to the lack of knowledge and confidence to use the method. Furthermore, the survey also found that although most contractors prepared detailed estimates using unit rate approach, they had a higher tendency to prepare detailed cost estimates for large projects than small projects.

It is evidenced that some contractors devise their own methods to prepare estimates and bids based heavily on their experiences; results in a large percentage of business failures and high potential of claims (Hegazy and Moselhi, 1995). It is therefore reasonable to see contractors unanimously agreeing on the importance of estimators' experience and expertise within the estimating and tendering process (Lowe and Skitmore, 1994; Oteifa and Baldwin, 1991).

3.5.1.2 Accuracy of Contractor's Total Cost Estimates

Unfortunately, most of the studies regarding construction cost estimation were based on the estimates prepared by consultant quantity surveyors. The reason behind is understandable. It is easier to collect project cost estimates from the consultant QS firms; and the analysis of accuracy can be done by comparing the QS' estimates against the contact sums as signed. If the purpose of study is to examine the accuracy of a contractor's estimates, comparison has to be made with its internal costs during construction. In practice, however, it is difficult to obtain a contractor's internal cost data due to confidentiality. Although comparison of the contractor's estimate with its competitors' bids seems to be an alternative approach of study (as some researchers did), the bidding strategy of each bidder is unknown. Therefore, study related to the precise measurement of estimating accuracy is not available.

Among the limited studies in this subject, Hegazy and Moselhi (1995) claimed that 5% of variations existed between the contractor's bid and the second-lowest bid (in case when the contractor won) or winner's bid (in case when the contractor lost) when they conducted the survey with the American contractors. There are some other surveys done on the QS' estimates. In Singapore, Cheong (1991) reported that quantity surveyors perceived their level of estimating accuracy to be 5% to 10%. However, one of the quantity surveying firms was found to have a deviation of 33.79% overestimation to 31.3% underestimation when compared with the contract sums signed. In a more recent survey done by Gunner and Skitmore (1999), the deviation was around 10% and similar result (9% deviation) was found in Ling and Boo's study (2001). However, when asking the respondents about the acceptable level of deviation, Ling and Boo found that the figure was much lower, only around 6.3% overestimation. As commented by Ling and Boo, the significant difference between the expected accuracy and the actual accuracy of estimates demanded more effort to be made to make estimates more accurate.

Although the above studies cannot reflect the complete picture of estimating accuracy, they did expose the problem of high inaccuracy level in the estimating practice among the contractors and consultants in different countries.

3.5.2 Project Overheads Estimation in Practice

3.5.2.1 Project Overheads Estimation Methods Used by Practitioners

As discussed in Section 3.4 (Estimation Theory on Project Overheads), a lot of literature emphasized the importance of project overheads estimation in the tendering process. Similar view is observed among the practitioners, e.g. from the results in Oteifa and Baldwin's survey (1991) with the U.K. contractors. In that study, respondents agreed that calculating project overheads was an important task within estimating and tendering (ranked 5th out of fourteen tasks).

However, the estimating methods used by practitioners seem not to align with their perceived importance of project overheads. A survey of general estimating practice in the U.S.A. (Hegazy and Moselhi 1995) revealed 83% of responding contractors estimated project overheads in a detailed manner; 14% of them estimated project overheads as a percentage of the direct cost while the remaining 3% did not estimate them at all. Similar results obtained in a survey by Fayek et al (1999) amongst the Canadian contractors identified only 79% of them accounted project overheads separately. The rest of them either allowed the project overheads in the measured work or in the markup percentage. Another survey of large foreign contractors in Saudi Arabia (Assaf *et al* 1999) revealed only 71% (sample size = 61) estimated the project overheads with reference to the contract requirements laid out in the tender document. The rest of the contractors just applied a percentage against the direct costs of measured works for the project overheads allowance.

Similar to the American and Saudi Arabian practice, the British and Australian contractors also exhibit comparable behaviour. From interviews conducted with the U.K. contractors by Skitmore and Wilcock (1994), pricing of the project overheads by them was based on their perceived expectations of the Architect, not the specifications laid down in the contract. Other interviews (Tah *et al,* 1994) showed contractors either priced the project overhead costs as a percentage of the measured works or as a lump sum allowance. In the survey done by Fayek et al (1999) among the Australian contractors, more than 90% of the projects examined used profit margin to cover the overheads. All these studies indicate that estimation of project overheads is quite dependent on both the individual estimator's decision and contractor's own practice. Some textbooks also claimed to observe similar practice among the contractors. Steward *et al* (1995, p. 373) mentioned that estimating the project overhead cost was a time-consuming and inexact task, and hence contractors often applied a percentage of direct costs as an estimate for project overheads. Peurifoy and Oberlender (2002, p.14) also criticized that some contractors multiplied the direct cost by a certain percentage to get the overhead cost was a quick method but might not be sufficiently accurate for most estimates.

Unfortunately, there is no formal study conducted in Hong Kong regarding the estimation method used by the contractors. Hence, comparison can not be made here.

3.5.2.2 Accuracy of Project Overhead Estimates

There is always a concern of resources constraint in the estimation process. Therefore, approximate estimates sometimes do have their advantage over detailed estimates. However, in view of estimating project overheads in the tendering process, the abovementioned literature had already suggested that an accurate estimation of project overheads is essential.

However, there appears to be a dearth of literature relating to the accuracy of estimating project overheads and only Hegazy and Moselhi's work (1995) is available for reference. From their study, the estimation of project indirect costs exhibited much more variation (more than 15%) whereas the estimation of the direct cost exhibited the least variation (around 5%). Higher variability among the estimates of project indirect costs meant a greater non-uniformity among the competing bidders, which implied a larger inaccuracy among the estimates. Hegazy and Moselhi further suggested that the possible reason for the high variability might be the difficulty to assess the qualitative factors involved in indirect costs. The concern of difficulty to assess project requirements will be further discussed and examined in the last section of this Chapter when discussing the attributes affecting estimating accuracy.

3.6 Contemporary Cost Estimation Methods

With the advancement in information technology in recent years, increasing amount of researchers attempted to find ways to estimate costs in a more reliable and efficient manner. Sophisticated estimation models using artificial neural network, fuzzy logic, and simulation models are proposed as being effective tools for cost estimation.

In the following sections, the cost estimation models developed for the purpose of estimating general project costs; or cost of other production processes are reviewed. For the sake of clarity, they are categorized into three main types:

- Simple parametric estimation models
- Statistical models
- Artificial intelligent models

3.6.1 Simple Parametric Estimation Models

A parameter is defined as a distinct attribute of a project (Ahmad and Ommi, 1996). Examples of a parameter are gross floor area, structural frame and the like. By referring to the historical cost data of similar projects, cost of the project in question can be projected. Amongst the various studies in this subject, which parameters to be used to forecast the project cost is the main research theme. Some authors also attempted to build a parametric estimation model using programming techniques so as to improve the accuracy.

A lot of researches advocated the power of parametric estimation in terms of speed and reliability (Popham, 1996; Davis, 1998; Akintoye and Fitzgerald, 2000). Ahmad and Ommi (1996) proposed that parametric estimation could be applied to construction cost estimating exercise. The suggested parameters could be broadly divided into two types: quantitative and qualitative. Examples of the parameters used were listed in Table 3-3.

Table 3-3 Examples of Parameters for Cost Estimation (Ahmad and Ommi, 1996)

Quantitative parameters	Qualitative parameters
Gross floor area	Type of structural frame
Number of floors above ground	Type of roof
Perimeter of building	Exterior-closure type
Average height of floor	

Davis (1998) also developed a parametric model for cost estimation of a "generic building". Similar to the model of Ahmad and Ommi (1996), the parameters used were all project-based, e.g. gross floor area, parking area, housing plan, building components, etc. Popham (1996) suggested similar proposal but the exact parameters used to project the cost was not clearly spelt out. By collecting sufficient cost information and activity characteristics of the projects, the likely cost could be estimated by comparison with appropriate adjustments e.g. changes in scope of work, learning curves being made (Popham, 1996).

From the illustrations in the above studies, parametric method is somehow a replicate of the comparison approach used by a lot of Professional Quantity Surveyors when prepared the pre-tender estimate. The past researches were just trying to highlight the careful selection of unconventional parameters (like the qualitative parameters suggested by Ahmad and Ommi, 1996) to fit into their estimation models. In general, the power of this method is the ability to provide the estimate when data of similar projects is available. However, the drawback is that a very substantial amount of in-depth information must be available for providing enough cost data for each estimation exercise (Popham, 1996).

3.6.2 Statistical Cost Estimation Models

To a certain extent, statistical cost estimation models belong to the same family of parametric estimation. In simple parametric method, the cost is projected by comparison of historical data with adjustment or simple graphical curve fitting. In statistical cost models, regression analysis is used to find the underlying relationship between the variables and the cost (Garza and Rouhana 1995). Al Khalil et al (1999) proposed a conceptual model for estimating water reservoir projects. Based on multiple regression analysis, they identified that storage capacity of the reservoir, project duration and access were the most significant factors affecting the overall cost of the projects. The regression equation, with 95% confidence level was suggested as:

$$\text{Cost} = -2.107 \times 10^5 + 64.8022\ \text{CAPACITY} - 8.09423\ e^{\text{DURATION}} + 28219\ \text{ACCESS}$$

where CAPACITY is the storage capacity of the reservoir (in m^3); DURATION is the duration of the project (in months); and ACCESS is the weighting transformed by combining the two variables – vicinity of site and distance between the contractor's headquarter and the site.

Similar to Al Khalil et al, Phaobunjong and Popescu (2003) collected project cost data from one hundred and thirty-nine projects and developed a multiple regression model to predict project cost. The final regression equation was:

$$\text{Cost per building gross floor area} = 202.245 + 15.740\text{TFLR} - 126.196\text{RATIO}$$

where TFLR = *ln*(number of floor levels); and RATIO is the space usage ratio

The regression model was subsequently tested with validation data set, with predictions ranging from underestimating by -0.8% to overestimating 13.5%.

In another study done by Nelson, Powell and Federle (1998), the authors selected more than ten cost categories to build the regression model of office buildings, continuous care facilities, and grocery stores. Details of the twelve cost categories are tabulated in Table 3-4.

Table 3-4 Job Cost Categories in Historical Database by Nelson, Powell and Federle (1998)

1. General conditions	5. Exterior enclosure	9. Mechanical
2. Site development	6. Interior finishes	10. Fire protection
3. Foundation and slab-on-grade	7. Equipment and special construction	11. Electrical
4. Structure	8. Conveying systems	
12. Total		

Then, a series of variables (as shown in Table 3-5, e.g. no. of stories, location, site area, percentage of skin in glass, etc.) were input to the regression model to predict the cost of each cost category. The authors emphasized that different variables could be input to the model by users in order to enhanced the reliability of the model for different types of buildings or projects.

Table 3-5 Variables Used in Regression Models for Three Types of Buildings by Nelson, Powell and Federle (1998)

Office buildings	**Continuous care facilities**	**Grocery stores**
Time index	Time index	Time index
Location index	Location index	Location index
AIA gross square footage	AIA gross square footage	AIA gross square footage
Area of exterior wall divided by building square footage	Area of exterior wall divided by building square footage	Area of exterior wall divided by building square footage
Site acreage	Site acreage	Area of site receiving paving
Percent acoustical ceiling	No. of apartments	Percent acoustical ceiling
Percent skin in glass	No. of parking stalls within the building structure, not in open lot	Percent skin in glass
Partition density	Partition density	Floor covering area
Percent receiving floor coverings	Common area square foot	Percent receiving floor coverings
Are deep foundations required?	Are deep foundations required?	Sales area
No. of stories	No. of stories	Back area
Is frame precast?	Is frame wood?	Dock area
Is skin precast?	Are windows wood?	Ceiling area
Is interior built-to-suit?	No. of elevator stops	Skin area
	Are there stand alone kitchens?	Plate area
	Is cooking allowed in apartments?	Glass area
	Are units low-income?	
	Are utilities branched of existing facility?	
	Are through-wall mechanical units used?	
	Is there a nurse call system?	
	Footprint-to-GSF ratio	
	Percent exterior in brick	
	Is exterior insulation used?	

Garza and Rouhana (1995) applied multiple regression technique in a similar manner to estimate the cost of piping jobs. There are some other similar studies like using regression models to estimate cost of water supply projects (Ulug, 1993), road projects (Jensen, 1993), etc. Although different variables may be chosen by the authors to predict the cost in each case, the methodology of estimation used in these studies is the same. Whilst the concept of the model seems to be rather simple, the general mathematical form of the relationships (e.g. linear / hyperbolic / step / impulse / discontinuous function) and the independent variables have to be identified before modelling (Garza and Rouhana, 1995).

3.6.3 Artificial Intelligent Estimation Models

Cost estimating using artificial intelligence has became a more user-friendly technique than before as a result of the wide application of computers and the vast development of commercial software packages. Major techniques developed for the cost estimating exercise include fuzzy logic and neural networks.

Nevertheless, it is worth to note that although these AI models are advocated by a lot of researchers, they are rarely used in reality (Akintoye and Fitzgerald, 2000; Elhag and Boussabaine, 1998). Most of the contractors only apply computerization to the extent of cost data management and spreadsheet programme (Peurifoy and Oberlender, 2002).

3.6.3.1 Fuzzy Cost Estimation Models

Fuzzy theory approach is developed by Zadeh (1965) and soon gains popularity in various applications that include management, economics and engineering (Zadeh, 1994). Although the actual adoption in practice is low, it serves as an alternative way to estimate the construction cost (Ostwald, 2001). A variable in fuzzy logic has sets of values which are characterized by linguistic expression, e.g. high complexity, moderate complexity, low complexity, etc. These linguistic expressions are represented numerically by fuzzy sets. Every fuzzy set is characterized by a membership function, which varies from 0 to 1. Fuzzy sets have a distinct feature of allowing partial membership. As a result, an element can be a member of the fuzzy set, with degree of membership varying from 0 (non-member) to 1 (full member) (Ostwald, 2001). This is in contrast to conventional sets, where an element can either be or not be the set (in other words, either 0 or 1). Therefore, fuzzy technologies can be used to model decision processes for which mathematical precision is impossible or impractical.

There are many ways to assign membership values to fuzzy variables, which can be intuitive or based on algorithmic or logical operations. For example, as illustrated by Mason and Kahn (1997), cost drivers (independent factors) were first identified and the memberships were defined. Fuzzy logic was used to estimate the cost of excavation for residential development by basically two cost drivers: ground water level and political stability. Altogether nine fuzzy sets of cost drivers combination were used in the system (as there were three fuzzy sets describing each cost driver). Table 3-6 below shows the details of the fuzzy sets and the respective implied estimation cost.

Table 3-6 Fuzzy Sets and the Implied Cost for the Excavation Cost Estimation Model (Mason and Kahn, 1997)

Political stability	**Ground water level**		
	High	Moderate	Low
Unstable	**High** ground water level, **unstable** political stability = **HIGH cost**	**Moderate** ground water level, **unstable** political stability = **AVERAGE cost**	**Low** ground water level, **unstable** political stability = **AVERAGE cost**
Moderate	**High** ground water level, **moderate** political stability = **HIGH cost**	**Moderate** ground water level, **moderate** political stability = **AVERAGE cost**	**Low** ground water level, **moderate** political stability = **LOW cost**
Stable	**High** ground water level, **stable** political stability = **AVERAGE cost**	**Moderate** ground water level, **stable** political stability = **LOW cost**	**Low** ground water level, **stable** political stability = **LOW cost**

The final cost was then calculated by "defuzzify" a conclusion; which was the centroid of the membership function. The model was proved to be useful when only linguistic or qualitative assertions about the relationships between the cost and the project attributes were available.

Although the model appears simple in terms of the construction, Manson and Kahn (1997) addressed that one of the major problem in using fuzzy logic to estimate cost was to anticipate the "cost surface" (i.e. the implied cost under different fuzzy set). The problem would be more profound when more than two cost drivers were present, due to the complexity of forecasting the combined effect of various cost drivers. Nevertheless, the attractiveness of such application was that one could use linguistic expressions to describe the relationships between costs and project attributes, instead of using quantitative data to estimate the project attributes.

3.6.3.2 Artificial Neural Network Estimation Models

Recently, there are abundant studies on forecasting using artificial neural networks (ANN). Application of ANN in construction management and costing is also examined by a lot of researchers (Tam et al, 2005; Boussabaine, 1996; Li, 1995; Moselhi et al, 1991). The model is commonly used in decision-making, forecasting and optimization (Bhokha and Ogunlana, 1999). However, similar to fuzzy models, ANN is not commonly used among professional quantity surveyors or estimators (as discussed in Section 3.5) due to the lack of familiarity and knowledge. Before the review of various studies in using ANN to estimate construction cost, the general principles of ANN are discussed first.

3.6.3.2.1 Artificial Neural Networks - Principles

Artificial Neural Network (ANN) is a technique imitating the way of learning, thinking, storage and retrieval of information in human brains. It was evolved from Rosenblatt (1958)'s perception learning algorithm which was then further worked on by Minsky and Papert (1969). In simple terms, a neuron, when receives an input signal, will process and transmit an output signal to other interconnected neurons. The ANN is to simulate the working principles of these neurons. Boussabaine (1996) stated seven major concepts of ANN which assists a better understanding of an ANN model:

1. A set of processing neurons.
2. A state of activation for each neuron.
3. A pattern of connectivity among the neurons.
4. A propagation method to propagate the activities of the neurons through the network.
5. An activation rule to update the activities of each node.
6. An external environment that provides information to the network and interacts with it.
7. A learning method to modify the pattern of connectivity by using information provided by the external environment.

Figure 3-2 illustrates a simple ANN model. According to Boussabine (1996), a multilayered ANN consists of a number of nodes with each of the nodes linked to each of the nodes of another layer. The nodes in the input layer receive inputs x_1, x_2 and x_3 respectively with corresponding weight factors w_1, w_2 and w_3. The node then calculates and delivers an output obtained from activation or transfer functions. The output is delivered to the nodes of the next layer, where a similar computation takes place. The middle layer, which is hidden, gives a critical computational ability to the system. The final value is received in the output layer.

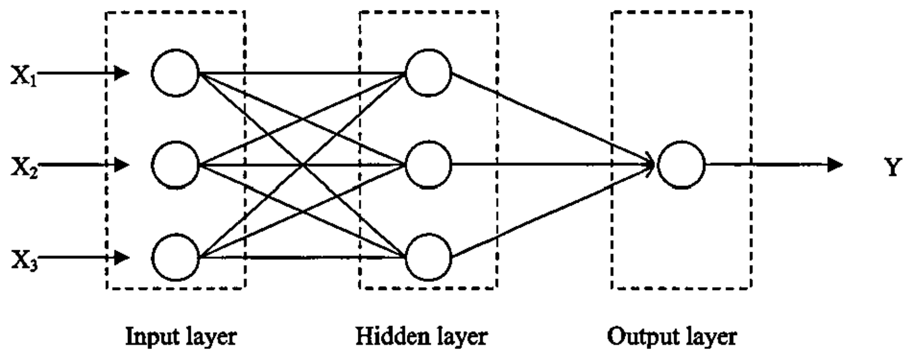

Figure 3-2 Example of a three-layered ANN architecture

- Input layer (3 layers were used in Figure 3-2 example), input = x_1, x_2, x_3
- Hidden layer (3 layers were used in Figure 3-2)
- Output layer (1 layer was used in Figure 3-2); output = Y.

The pattern of network architecture specifies how each node is connected to the other units in the network. The strength of each connection is controlled by the weight factors. As the network learns (or trains), the numerical values of the weights may change according to the new information that is circulating in the network. A learning (or training) method is used to change the weight of the network and other adaptable parameters. Therefore, unlike expert systems and traditional modelling methods, where knowledge is made explicit in form of rules, neural networks generate their own rules by learning from examples (Gallant, 1993).

There is a diverse range of ANN models in terms of architectures. More common ANN architectures include : Backpropagation; Kohonen; Probabilistic; Regression and Group Method of Data Handling (also called Polynomial). Each one of them will be described briefly below with reference to the NeuroShell 2 User's manual (Ward Systems Group, Inc., 1995).

3.6.3.2.1.1 Backpropagation Network

Among the various ANN models developed in the past years, the most commonly used algorithm in construction management is Backpropagation (Moselhi, 1998, Boussabaine and Kaka, 1998, Hegazy and Ayed, 1998). Back Propagation Algorithm is a supervised learning process by which, using a defined number of iterations of forward and backward feeding of the signals (as shown in Figure 3-3) between the input layer and output layer through the hidden layer. From the input layer, a scaling function is applied to map the input values to a fixed range from 1 to 0 or -1 to 1. The hidden layers produce outputs based on the sum of weighted values passed to them, apply the transfer function and map this sum to the output value. The deviations from actual outputs and calculated outputs (correction signals) are propagated through the network each time during the learning. As a result, the amount of estimation error can be minimized. The learning process will continue until the error rate at the output unit is reduced to a satisfactory level. Garza and Rouhana (1995) used an analogy of tuning a guitar to illustrate the back propagation theory, "we keep on adjusting string tensions until we get as close as possible to the desired sound, a tuned guitar is ready to play".

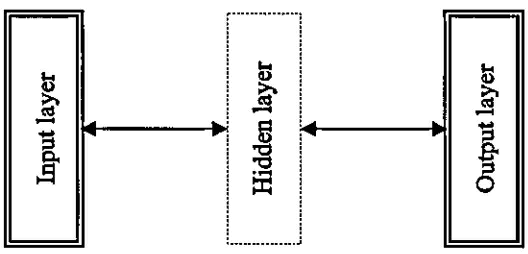

Figure 3-3 Standard 3-layered Backpropagation Neural Network Structure

Therefore, in backpropagation network, the choice of a suitable transfer function is essential. Common transfer functions provided in NeuroShell 2 (developed by Ward Systems Group, Inc. , include linear, tanh, Gaussian and logistic. The equations of these transfer functions are:

Linear : $f(x) = x$

Tanh : $f(x) = \tanh(x)$

Gaussian : $f(x) = \exp(-x^2)$

Logistic : $f(x) = 1/(1 + \exp(-x))$

where x is the input to transfer function of hidden layer

Assuming scaling function is chosen from -1 to 1, linear function will not transfer the values further and remain to range from -1 to 1. Tanh function maps the values from -0.7616 to 0.76156 in an S-curve through the origin. For Gaussian function, the mapped values will range from 1 to 0.3679 in the form of a bell-shaped curve. The logistic function maps values into the range from 0.2689 to 0.7311. Through the transfer function, non-linearity is introduced into the network design. In case of multi-layered architecture, different transfer functions can be applied to each hidden layer. For example, one hidden slab uses a Gaussian function to ascertain the features in the mid-range of the data and the other hidden slab applies a logistic function. In this case, a wider range of data and relationship pattern can be introduced.

3.6.3.2.1.2 Kohonen Network

The Kohonen Self Organizing Map network is a type of unsupervised network, which has the ability to learn without being shown correct outputs in sample patterns. These networks can separate data into a specified number of categories.

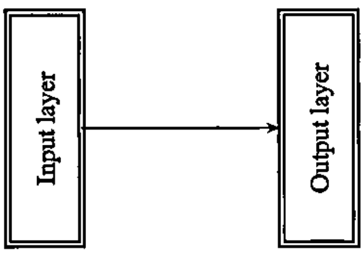

Figure 3-4 Typical Kohonen Neural Network Structure

As shown in Figure 3-4, there are only two layers: an input layer and an output layer which has one neuron for each output category. The training pattern will be propagated from the input layer to the output layer to produce a single neuron as the "winner". This process is repeated for all training patterns for a number of predefined epochs (i.e. a complete pass through of the entire network).

3.6.3.2.1.3 Probabilistic Neural Network (PNN)

The characteristic of Probabilistic Neural Networks is that it can train quickly on sparse data sets. It is a three-layer supervised network (as shown in Figure 3-5) for the purpose to separate outputs into a specified number of categories. Therefore, there will be more than one neuron in the output layer, each neuron for one possible category. The number of neurons in the hidden layer must be at least the same number as the number of training patterns. A smoothing factor is applied to all the links in the network. Higher smoothing factors cause more relaxed surface fits through the data. The network produces activations in the output layer according to the probability density function estimate for that category. The most probable category is then represented by the highest output.

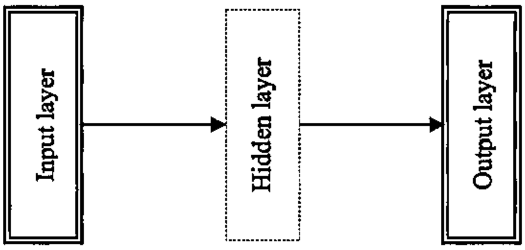

Figure 3-5 Typical Probabilistic Neural Network Structure

3.6.3.2.1.4 General Regression Neural Network (GRNN)

Like Probabilistic Networks, General Regression Neural Networks can also train quickly on sparse data sets. As in Figure 3-6, it is also a three-layer supervised network. However, rather than categorizing the data (as in Probabilistic Network), Regression Network will produce continuous valued outputs. Therefore, it is especially useful for continuous function approximation. Similar to the PNN, the number of neurons in the hidden layer has to be at least the number of training patterns. Smoothing factor is also applied to each link for modelling.

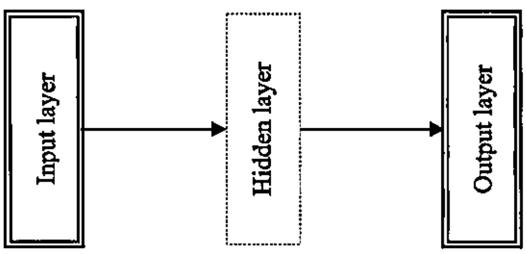

Figure 3-6 Typical Regression Neural Network Structure

3.6.3.2.1.5 Group Method of Data Handling (GMDH) Neural Network

The technique is also called Polynomial neural network. It was invented by A. G. Ivakhnenko in the former Soviet Union (Ward Systems Group, Inc., 1995), but enhanced by others, including A. R. Barron. It is a self-organized network involves successive layers with complex connections that are the individual terms of a polynomial.

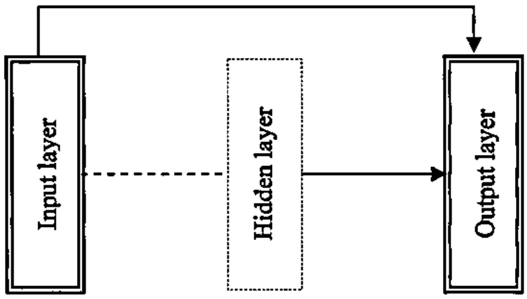

Figure 3-7 Typical Group Method of Data Handling Neural Network Structure

As shown in Figure 3-7, the initial layer is the input layer. The powers of input variables are chosen along with their co-variants and tri-variants as terms of the polynomial $\{x_1, x_2, x_3, \ldots, x_1^2, x_2^2, x_3^2, \ldots, x_1x_2, x_1x_3, \ldots, x_{n-1}x_n, x_1x_2x_3, \ldots\}$.

The first hidden layer is created by computing a linear combination of all of the polynomial terms with variable coefficients like the following:

$c_1 + c_2x_1 + c_3x_2 + c_4x_3 + c_5x_1x_2 + c_6x_1x_3 + c_7x_2x_3 + c_8x_1^2 + c_9x_2^2 + c_{10}x_3^2$

where c_1 is a constant; $c_2, c_3, \ldots$ are coefficients; x_1, x_2, and x_3 are the input variables.

As above, each neuron is a polynomial of some inputs (up to three inputs in each polynomial as allowable in NeuroShell 2). The algorithm determines values of the coefficients and then choosing the best ones (the ones which have a lower selection criterion value). Similarly, the second hidden layer is created by computing regressions of the values in the first hidden layer along with the input variables, and only the best ones (called survivors) are chosen by the algorithm. Similarly, successive layers take inputs from either the original inputs or the polynomials from the immediately preceding layer. This process continues until the network stops getting better according to a predetermined selection criterion.

The selection criterion is regarded as the "objective function" which determines which survivor should go to the next slab. There are six selection criteria available in the NeuroShell 2, including Full Complexity Prediction Squared Error (FCPSE), Prediction Squared Error (PSE), Minimal Description Length (MDL), Generalized Cross Validation (GCV), Final Prediction Error (FPE) and Regularity.

1. FCPSE (Full Complexity Prediction Squared Error) = Norm. MSE + CC * var(a) * C/N

   Where Norm. MSE = mean squared error of a model on the training set; CC = Criterion Coefficient; var(a) = variance of the actual output variable; N = Number of patterns and C = overall model complexity which takes into account the complexity of each term in the model. The criterion is actually the sum of two terms: Norm. MSE and an over-fitting penalty. The algorithm for calculating overall complexity is a Ward Systems Group proprietary method and hence cannot be elaborated further here.

2. PSE (Prediction Squared Error) = Norm. MSE + CC * var(a) * k/N

   It is similar to PSE except the overall C is changed to k, the number of coefficient in the model. The over-fitting penalty is the difference between the Best Criterion Value and the Norm. MSE.

3. MDL (Minimal Description Length) = Norm. MSE + CC * var(a) * k/N * ln(N)

   This is a criterion similar to PSE, but it incorporates a stronger penalty for model complexity.

4. GCV (Generalized Cross Validation) = Norm. MSE / $(1 - CC * k/N)^2$

   This is another method of introducing an over-fitting penalty.

5. FPE (Final Prediction Error) = Norm. MSE * (N+k)/(N-k)

   This is the minimum variance unbiased estimator of the mean-squared error of prediction. The Criterion Coefficient is not used for FPE, so the value set in the Training Criteria module is ignored.

6. Regularity

   This is the average squared error of the model on the test set. The application of this criterion is very much like calibration, i.e. to optimize the network by applying it to an independent test set during training. Calibration finds the optimum network for the data in the test set by computing the mean squared error between actual and predicted for all outputs over all patterns in test set.

In some respects, GMDH algorithm is very similar to the working principle of regression analysis, but it is far more powerful than regression analysis. GMDH Neural Network can build very complex models while avoiding overfitting problems. Furthermore, with GMDH network, the polynomial function which produces the minimum error can be identified when training terminates. In other words, the problem of "black-box" operation in other ANN architectures can be eliminated in the GMDH network.

3.6.3.2.2 Artificial Neural Networks – Applications in Cost Estimation

As ANN is well known for its learning and generalization capabilities, it is ideal for the development of decision-support tools and forecasting modules (Moselhi and Siqueira, 1998). In the last decade, there are numerous studies on cost estimation using ANN as the principal method. Specifically, Li (1995) portrayed three basic phases in the development of neural network-based cost estimation model:

1. Design phase :
   a. Identify the variables that characterize the cost estimation task;
   b. identify the types of variables (i.e. whether the variables have numerical values, like company size, or nominal values);
   c. determine the configuration of the neural network : number of input variables, output variables, hidden layers and number of nodes on hidden layers (though it is still be seen as a trial-and-error process);
   d. select training criteria (i.e. the permissible error and training rate).
2. Training phase :
   a. Present the network with cost examples and allow it to modify its connection weights, so that the error is reduced to a predetermined level.
3. Operation phase :
   a. Test the neural network model to see how well it performs; ready for use if the results is satisfactory.

Emsley et al (2002) used neural network to predict the total construction cost. Different parameters were tried, including changing the number of layers, the learning method, and the number of variables used in the model. A three-layered Generalised Regression Neural Network (GRNN) with forty-one variables was found to be the best, yielding the highest R^2 value (0.789) and lowest mean absolute percentage error (16.6%). Although Emsley et al admitted that the results were not much more accurate than the current cost estimation practices, the nonlinearities in the data indicated that neural networks were suitable for such modelling purpose.

On the other hand, an extensive study was done by Boussabaine and Elhag (1999) on the estimation of tender price using ANN. Besides using ANN to estimate the tender price, the authors tested the accuracy of prediction by changing the ANN architectures, transfer functions, learning rule and learning rates; and further compared with a neurofuzzy model and a multiple regression model. The findings reinforced a lot of principles regarding the design of ANN models, and affirmed the suitability of using ANN in cost estimation.

Moselhi and Siqueira (1998) used NeuroShell (a commercial neural network software package) to work out an ANN model for estimating structural steel buildings. Backpropagation was used to train the model, with seven nodes in the hidden layer exhibited the highest R^2 value (close to 0.9). Input variables used for the model included: area, height, joist span, and vertical loads. Based on this architecture, the minimum average error for the test set of data was 0.00002798. A linear regression was also performed using 15% of the project data extracted randomly by the system, and neural network was found outperformed regression. However, details of the comparative study were not elaborated in the publication.

Hegazy and Ayed (1998) had also conducted a study of estimating highway project cost using NeuroShell to construct the ANN model. Ten input variables were chosen in their study, including:

1. project type : bridge (1), highway (2), and others (3)
2. project scope : new (1), rehab. (2), and others (3)
3. year
4. construction season : winter (1), summer (2), and fall (3)
5. location : St John's (1), St John's suburbs (2), and Avalon region (3)
6. duration (months)
7. size (length in km)
8. capacity : 2-lanes (1), and 2-lanes divided (2)
9. water body : no (1), and yes (2)

10. soil condition : class 0 (0) to class 9 (9)

A three-layered backpropagation network was adopted as the training method. However, there was no detail on the selection process of architecture and training criteria for the network. The minimum weighted error achieved (by optimizing the amount of training to arrive at the least average error) was 10.34%, with 1.43% on training cases and 19.33% on test cases respectively. The results were compared with those generated by simplex optimization and genetic algorithm. It was found that the best overall model (for both training and test cases) was the one produced by simplex optimization. Although neural network produced smaller errors in training cases, it behaved relatively poor in the test cases. Nevertheless, in terms of the overall error, Hegazy and Ayed commented both simplex optimization and neural network were suitable models to predict the total cost.

In a study done by Adeli and Wu (1998), regularization neural network was formulated to estimate the cost of construction projects. Based on the regularization theory by Tikhonov and Arsenin (1977), the overfitting problem of ANN could be compensated by adding a regularization term to the standard error term. Only two variables were applied, quantity and thickness (of pavement), to estimate the road work projects using Matlab. As highlighted by Adeli and Wu, the estimation from regularization neural network would only depend on the training examples, but not the architecture of the neural network (e.g. number of nodes in the hidden layer) and the number of iterations. Hence, a more objective estimate could be arrived at.

In another example done by Garza and Rouhana (1995), ANN was used to estimate the overall cost of steel pipes by three input variables: pipe diameter, number of elbows, and flange rating. The network used backpropagation to train ten sets of data. The mean squared error (MSE) of using 31,000 learning cycles was 2.47. The model results were also compared with those from a parametric-based multiple regression equation. It was observed that the neural network technique outperformed the regression model, with a lower mean squared error (MSE) (2.47) compared with multiple regression (7.64).

There are other examples of using neural network in construction cost estimation. For instance, Moselhi (1998) used the technique to estimate the cost of change orders, Ahmad and Rahman (1994) applied it to estimate the contingency allowance, etc. Most of them concluded that ANN was an effective tool for cost estimation. Table 3-7 summarizes the major features and findings of these ANN applications.

Table 3-7 Summary of Studies Using Neural Network in Cost Estimation Models

Authors	Application of ANN	Best model					Comparison with other model	Better model
		ANN architecture	No. of variables	No. of hidden nodes	Train data set	Test data set		
Kim and Han, 2003	Activity-based costing	Back-propagation	3	7	125 (mean absolute error (%)=0.809)	42	Hybrid model (GA + ANN)	Hybrid model
Emsley et al, 2002	Total construction cost prediction	GRNN	41	Not available	N/A (R^2 = 0.789)	N/A	Linear regression	ANN
Boussabaine and Elhag, 1999	Tender price estimation	Back-propagation	2	5	27 Mean absolute error = 6.6%	9 Mean absolute error = 9.1%	Fuzzy and Multiple regression	ANN
Moselhi and Siqueira, 1998	Structural steel building cost estimation	Back-propagation	4	7	22 (R^2 = 0.9)	7	Linear regression	ANN
Moselhi, 1998	Cost of change orders estimation	Back-propagation	4	Not available	27	7 (R^2 = 0.862)	Linear regression	ANN
Hegazy and Ayed, 1998	Highway project cost estimation	Back-propagation	10	5	14 (min. weighted error = 1.43%)	4 (min. weighted error = 19.33%)	GA and simplex optimization	Simplex optimiza t-ion
Adeli and Wu, 1998	Total construction cost estimation	Regularization	2	Not available	121	121 (average error = 0.05)	Nil	-
Gara and Rouhana, 1995	Cost estimation of pipework	Back-propagation	3	4	10	6 (MSE = 2.47)	Linear regression	ANN
Ahmad and Rahman, 1994	Cost estimation of contingency	Back-propagation	5	4	10	5 (max. error % = 10.5%)	Nil	-
Hanna and Chao, 1994	Forecasting of cost escalation	Multi-layered feed-forward	4	20	15 (max. error = 0.66%)	10 (max. error = 1.85%)	Nil	-

From Table 3-7, several features can be observed related to the build up of neural network structure in these cost estimation studies:

1. Backpropagation is the most commonly used architecture in construction cost estimation.
2. As discussed by Li (1995) and Moselhi and Siqueira (1998), the number of input variables and the number of neurons in the hidden layer which could work out the best prediction was often a trial and error exercise. It is evidenced in the studies that

the best model had to be identified by trying different number of input variables (Emsley et al, 2002; Zhou et al, 1996) and different number of neurons in hidden layer (Moselhi and Siqueira, 1998; Boussabaine and Elhag, 1999)

3. Closely related to the number of input variables, the number of training epochs has to be carefully tuned to avoid the model being "overtrained" (Ward Systems Group Ltd., Inc.) It happens when too many hidden neurons, or a huge amount of noisy data (like financial data) are used. The model will memorize the patterns instead of smoothly interpolate between the patterns. The noise from the conflicting data will also be learnt by the model if it is overtrained. Again, different combinations have to be tried. As illustrated by Garza and Rouhana (1995) and Boussabaine and Elhag (1999), different number of training cycles were tried to find the best solution.
4. The proportion of test set and training set of data also varies amongst the studies. According to NeuroShell 2 User's Manual (p.118), the test should be approximately 10 to 40% the size of the training set of data. In the detailed study by Boussabaine and Elhag (1999), the best ratio of training to test sets is 70/30% for training and 90/10% for testing; which fits the principle stated in NeuroShell 2 User's manual.
5. ANN is a suitable method for cost estimation, especially when compared its performance with traditional parametric method. As summarized in Table 3-7, ANN outperformed traditional methods like regression and parametric estimation. Details on the performance of ANN will be further discussed in the following section.

3.6.3.3 Comparison of ANN with Other Methods of Estimation

To prove the reliability and effectiveness of ANN, several researchers had conducted comparative study on the performance of ANN with other estimation methods like regression, genetic algorithms (GA) and even sensitivity analysis. Summary of some studies are included in Table 3-7. Results of ANN performed better than conventional parametric methods for cost estimation are presented in Emsley et al (2002), Boussabaine and Elhag (1999), Hegazy and Ayed (1998), Garza and Rouhana (1995), Moselhi and Siqueira (1998), and Smith and Mason (1997). As in the study of Emsley et al (2002), ANN produced the smallest mean absolute percentage error of 16.6% whereas the best regression model gave 19.3%. In Moslhi and Siqueira's study on cost estimation of structural steel buildings (1998), ANN produced smaller errors on both the training and test cases (overall mean absolute error is 10.77% from ANN and 14.76% from regression) for estimating the highway projects cost. In the study by Hegazy and Ayed (1998), genetic algorithms failed to produce better results than ANN. The overall error of GA was much higher (21.8%) when compared with ANN's results (10.34%). In another comparison done by Garza and Rouhana (1995), ANN was found better than multiple regression in estimating the pipe cost; based on the MSE generated (linear multiple regression: 11.205, nonlinear multiple regression: 7.64, ANN: 2.47). In the remarks made by Smith and Mason (1997), ANN would be a good substitute for regression since ANN was capable to accept a large number of cost drivers than regression and could accommodate multicollinearity readily.

On the other side, there are reports which do not fully favour ANNs and are presented in Wright and Williams (2001), Kim and Han (2003) and Chan and Genovese (2001). A common problem of ANN as found from the studies is that there is no single type of neural network model that will be adequate for all real life problems and more researches have to be done to enable neural network to become a standard tool for industrial applications (Vojinovic and Kecman, 2001; Hegazy and Ayed, 1998). A lot of researchers committed that ANN inherited some problems that were difficult to resolve, including:

1. Massive amount of data was required for the training of ANN in order to secure the reliability of the model (Li, 1995; Zhou et al, 1996; Moselhi and Siqueira, 1998).
2. Difficult to explain a particular conclusion as the model itself is not transparent. (Li, 1995; Zhou et al, 1996; Moselhi, 1998; Moselhi and Siqueira, 1998).
3. Design of the network is more of an art than science and requires some sort of trial and error (Garza and Rouhana ,1995; Boussabaine and Elhag, 1999; Ward Systems Group, Inc.). Whether sufficient trials have been made requires patience and experience of the researchers. However, on the other hand, the model is sensitive to the configurations (Moselhi and Siqueira, 1998) and thus care has to be taken when using the technique.

Nevertheless, ANN does possess a lot of advantages that cannot be found in traditional estimation method. They are:

1. The adaptivity of ANN models make them more suitable for cost estimation prediction in real environment (Boussabaine and Elhag, 1999). It is not necessary to have in-depth knowledge of the parameters, and/or how they affect the output (Zhou et al, 1996; Garza and Rouhana, 1995). Even if the underlying mathematical

equations representing is unknown, ANN can model the factors contributing to the costs satisfactorily (Boussabaine and Elhag, 1999).

2. The non-linearity characteristic of ANN is valuable to cost estimation as the cost attributes have known and unknown non-linear relationships (Boussabaine and Elhag, 1999). As a result, no priori assumptions were made as to the relationships between the input factors and output factor (Li (1995)).
3. ANN can account for complex cases that required a large number of parameters to be considered in parallel (Moselhi and Siqueira, 1998; Boussabaine and Elhag, 1999).

3.6.3.4 Comparison of Different ANN Architectures

Among the various types of ANN architectures, most of them are supervised networks. These include backpropagation, probabilistic neural network and general regression neural network in NeuroShell 2. In supervised networks, parameters like number of layers, number of hidden neurons, transfer functions, etc. have to be set or tried by the users when design the network architecture. As discussed, backpropagation was the most common architecture used by researches to build cost estimating models due to its popularity in academia. Nevertheless, other algorithm like GMDH gains increasing recognition in a lot of domains like finance and economics systems, medical diagnostics, military systems, acoustic and seismic analysis, and a lot more (Ivaknenko and Cerda 2002).

GMDH network is transparent and self-organized in nature. Neural network using GMDH algorithm can minimize the instructions from the users and the priori information of the model is also reduced to the minimal (Ivakhnenko and Ivakhnenko, 1995). Besides, the time-consuming setting of parameters in the "trial-and-error" approach can be avoided as the structure is optimized automatically, with the non-significant variables automatically excluded in the model. As commented by Kim and Park (2006), GMDH could provide an automatic selection of essential inputs without using prior information on the relationship among the inputs and output. Kondo and Pandya (2000) proved that the estimating of large-spatial air pollution patterns using GMDH neural network was accurate and commented that the GMDH network was easy to apply to complex practical problem. Tam et al (2004) used various network to model hook times of mobile crane, and found that GMDH worked well compared with other networks like multi-layered feedforward and general regression neural nets.

Although GMDH algorithm can provide a systematic design procedure to generate an optimal model structure, Oh and Pedrycz (2002) highlighted its drawback of the tendency to produce an overly complex network when the systems were non-linear.

3.7 Cost Attributes

From the literature review of various cost estimation methods / models, it is obvious that the identification of input variables or attributes affecting the estimate is the most fundamental step before any estimation can take place. Therefore, a thorough review on the existing literature regarding the factors affecting the project overheads is done in this section. Furthermore, the factors affecting construction project duration and cost; time and cost over-runs; and construction risks are also reviewed due to their close relationship with project overheads.

3.7.1 Factors Affecting Project Overheads

It is apparent that time is a significant factor affecting project overheads. The Code of Estimating Practice had put a great emphasis on the time factor (i.e. the duration of the project). Advocates to this idea include Geddes (1996), Taylor (1994) and Mansfield (1983). In fact, one can easily identify items that are time-related like lighting and power, telephone tariff, staff salaries, etc.

While project duration is agreed to be a major factor by most of the theorists, Cooke (1984) particularly highlighted location of the site would affect the price of a number of project overhead items including: travelling expenses, provision of site transport, access, temporary hardstandings for plants, importation allowance, and protection of public property. Nevertheless, similar to most text books, there is no empirical ground on the above statement raised by Cooke.

In an analysis done by Solomon (1993), the four major items contributing to the project overhead costs (amounting to about 80% of the overall project overheads) are: staffing, mechanical plant, access, and site accommodation. The factors affecting their respective expenditure are summarized in Table 3-8.

Table 3-8 Summary of Factors Affecting Project Overhead Cost (by Solomon, 1993)

Majors items contributing to project overhead costs	Distribution of project overhead costs	Factors affecting the item expenditure
Staffing	20 – 30%	- scale of project* - complexity of project* - type of project - extent of sub-letting*
Mechanical plant	17 – 26%	- nature of project - site conditions - equipment available to contractor
Access*	15 – 22%	- site layout - shape of structure to be constructed - nature of structure frame
Site accommodation	8 – 14%	- staff and labour size - size of mechanical plant fleet - shape and nature of site - client's requirement

Besides the abovementioned factors, Assaf et al (1999) surveyed sixty-one contractors and identified eight factors that affected project overhead costs. Details are listed in Table 3-9 according to their order of significance.

Table 3-9 Summary of Factors Affecting Project Overhead Costs (by Assaf et al, 1999)

Rank	Factors affecting project overhead costs	Index % (a weighted average of response, weightings from 5 to 1 representing strongly agree to strongly disagree)
1	Project complexity, location and size*	83.6
2	Percentage of sub-contracted work*	79.6
3	Payment schedule	64.6
4	The need for work	61.0
5	The client's strictness in supervision	56.4
6	Type of contract	53.2
7	Number of competitors	50.2
8	How much cash the contractor has available	43.2

Although the survey was done in Saudi Arabia, the contractors selected for the survey were all large foreign contractors who could bid for projects worth up to US$13.3 million. The findings, therefore, can be a good reference for this research.

Based on the empirical findings obtained from the surveys by Solomon (1993) and Assaf et al (1999), some common factors affecting project overhead costs can be identified (as marked with asterisks in Table 3-8 and Table 3-9). They are namely:

- site access,
- size of project,
- complexity of project, and
- extent of sub-setting.

3.7.2 Factors Affecting Project Cost and Duration

As mentioned in the last Section, a lot of text books suggested that project overheads were affected by duration of the project. On the other hand, the analysis on project overheads (as discussed earlier) indicated "size of project" is one of the critical factors affecting its expenses. Since the research in factors affecting project overheads is very limited but ample researches are done to investigate the factors affecting the project duration and cost, the studies in the latter is reviewed here to get further insight into the study.

From the EPSRC Research by Elhag and Boussabaine (1998), sixty-seven factors were selected for a survey with one hundred and eighteen quantity surveyors in the U.K. to study the factors affecting project cost and duration. The factors were related to client characteristics, consultant and design characteristics, contractor characteristics, project characteristics, contract procedures and procurement methods, and external factors and market characteristics. A 3-point likert scale (not significant, moderately significantly and highly significant) was used for each question / factor. Results showed that "extent of alterations and late changes to design" was the most significant single factor whereas the consultant and design characteristics was the most significant category of factors affecting the project duration and cost. Details of the significance of each factor are listed in Table 3-10.

Based on the findings above, Elhag and Boussabaine (1998) used twenty-one factors to predict project cost as shown in Table 3-11. Using the cost model developed by ANN, prediction of project cost accounted 76.4% accuracy in the training set of data, and 41.8% accuracy in the testing set of data. They had also used neurofuzzy modelling (Elhag and Boussabaine, 1999) to predict project duration. The accuracy of prediction (by the best model) was 90%, using the factors as shown in Table 3-11.

Table 3-10 Summary of Results from the EPSRC Research on Factors Affecting Cost and Duration of Construction Projects by Elhag and Boussabaine, 1998

Rank	Factors	Severity index	Rank	Factors	Severity index
1	Absence of alterations and late changes to design	94.43	35	Productivity effects : (managerial, organisational, labour, technology) (of contractor)	71.57
2	Management team (suitability, experience, performance)*	93.53	36	Construction method / technology	71.50
3	Priority on construction time / deadline requirements	92.17	37.5	Type of foundations (pile / raft / pad / ... etc)	71.43
4.5	Completeness & timeliness of project information (design, drawings, specifications)	91.67	37.5	Access to site	71.43
4.5	Variation orders & additional works (magnitude, timing, interference level)	91.67	39	Plant costs / availability / supply / condition / performance	71.37
6	Intensity / complexity of building services*	91.20	40	Current work load (of contractor)	70.50
7	Quality of design & specifications	88.90	41.5	Partnering arrangements (of client)	69.80
8	Complexity (of project)*	87.87	41.5	(Contractor) present claims (size & quantity)	69.80
9	Level of competition & level of construction activity	86.73	43	Location (regions / rural, urban) (inner city / outskirts) (of project)	69.63
10	(Client) certainty on project brief	86.17	44.5	(Project) type / function (residential, commercial, industrial, offices)	68.63
11	Client requirements on quality	82.87	44.5	Type of structures (steel, concrete, brick, timber, mansory)	68.63
12	No. of basement levels (of project)	82.30	46	Type of contract / use of standard form of contract	68.60
13	Experience on similar projects*	81.97	47	Quality of finishing (of project)	68.00
14	Level of uncertainty of soil conditions	81.37	49	Level of communications within the contractor organisation	67.63
15	Phasing requirements (areas to be handed over first or initial non-availability)	80.97	49	(Client) experience of procuring construction	67.63
16	Type of client (public / private / developer)	80.53	49	Number of bidders on competitive projects	67.63
17	Materials prices / availability / supply / quality / imports	80.00	51	Type of cladding & external walls (brick, double glazing, ... etc)	66.67
18	Buildability of design	79.60	52	Off-site prefabrication	66.60
19	Labour costs / availability / supply / performance / productivity	79.03	53	Financial ability / payment record (of client)	65.73
20	Planning capability & level of resource deployment / utilization / optimization (of contractor)	78.43	54	Interest rate / inflation rate	65.70
21	Method of procurement (traditional, design & build, project management, ... etc)	78.07	55	Claims & disputes resolution methods (litigation / arbitration / others) (contract procedure)	62.83
22	Management (of contractor) / labour relationships & confidence in work force	93.53	56	Weather condition	61.90
23	(contractor's) previous claims record i.e. assessment of 'low tender' – 'high claims' performance	77.47	57	Interviewing of selected prospective contractors	61.83
24	Tender selection method (open, selected, negotiation, single or two stage, etc...)	77.20	58	Inspection, testing & approval of completed works (toughness / requirements)	61.10
25	Submission of early proposals for costing / cost planning (by consultant)	76.87	59	Estimation method & cost control technique (of contractor) (accuracy & reliability)	60.80
26	Site conditions / site topography	76.47	60	Government regulations / policies (health & safety, fire, CDM, ...etc)	60.00
27	Height / no. of stories (of project)	75.50	61.5	% of main contractor direct work & % of subcontracted work	58.83
28.5	Project finance method / appropriate funding in place on time (of client)	75.07	61.5	(Contractor) record of payments to sub-contractors	58.53
28.5	Working relationships with client / contractors / other design team consultants (previous / present)	75.07	63	(Contractor's) bond / warranty arrangement	58.10
30	(project) size / gross floor area	74.53	64	Number of subcontractors	54.23
31	Spread of risk between construction parties (client / consultant / contractors)	73.27	65	(Contractor's) CDM regulation awareness	53.33
32	Payment modalities (fixed price, cost plus, BOT, PFI-DBFO, etc ...)	72.53	66	(Contractor) markup policies & % (general & project wise) (special or normal conditions applied)	50.97
33	Stability of market conditions	72.37	67	(Contractor's) accidents on sites record	48.03
34	Financial capability (of contractor)	72.20			

Table 3-11 Selected Attributes for Projection of Project Cost (Elhag & Boussabaine, 1998) and Modelling of Project Duration (Elhag & Boussabaine, 1998)

No.	Factors	Used to forecast project cost	Used to forecast project duration
1	Type of client	✔	
2	Type of building	✔	
3	Deadline	✔	
4	No. of storey	✔	
5	No. of basement	✔	
6	Frame structure	✔	
7	Location	✔	
8	Site conditions	✔	
9	Type of foundation	✔	
10	Access	✔	
11	Complexity of building services	✔	
12	Procurement method	✔	
13	Site topography / slope of site	✔	
14	Working space	✔	
15	Ground conditions	✔	
16	Type of soil	✔	
17	Mark-up %	✔	
18	Need for work (%)	✔	
19	Duration	✔	
20	Area	✔	✔
21	Tender selection method	✔	✔
22	Type of contract		✔
23	Fluctuation in prices		✔
24	Contract sum		✔

There are a few studies taken in Hong Kong that are related to factors affecting project cost and duration (Tam and Harris, 1996; Kumaraswamy and Chan, 1998; Dissanayaka and Kumaraswamy, 1999; Chan and Kumaraswamy, 1999; Khosrowshahi, 1997). In the study done by Tam and Harris (1996), twenty attributes were selected (as summarised in Table 3-12) to study impact of factors to cost and duration of projects in Hong Kong. Results found that six factors were significant ones affecting the cost and duration of projects. They were:

1. complexity of project,
2. percentage of professional qualified staff,
3. project leader's experience,
4. contractor's past performance or image,

5. origin of the company,
6. architect's or client's supervision and control of the quality of work and work progress.

Table 3-12 Summary of Attributes Selected to Study Factors Affecting Duration and Cost of Hong Kong Projects by Tam and Harris, 1996

No.	Factors	No.	Factors
1	(Contractor's) staff training programme	11	Origin of the company (contractor)
2	(Contractor's) plant ownership policy	12	Amount of direct employed labour (of the contractor in hand)
3	(Contractor's) size of company	13	(Contractor) listed on the stock market
4	(Contractor's) quality of management team – professional qualifications	14	(Contractor) decision making centralised in head office or de-centralised to site
5	(Contractor's) quality of management team – project leader's experience	15	Contractor is client's subsidiary firm
6	(Contractor's) past performance of the project manager	16	The architect / engineer (quality of drawings, variations, ...)
7	Contractor's experience in the type of job	17	Architect's or client's supervision and control of the quality of work and work progress
8	Contractor's workload	18	Punctuality of payment by the client
9	Contractor's past performance or image	19	Complexity of the project
10	Number of years in the business (contractor)	20	Profitability (of the project from the contractor's point)

Khosrowshahi (1997) concluded that nature of project was one of the factors significantly affecting the duration and cost of public housing projects, whereas Chan and Kumaraswamy (1999) found that height of building, labour availability and presence/absence of facades were critical factors affecting duration of public housing projects.

Based on the above Hong Kong studies, a summary of significant factors is shown in Table 3-13. Apparently, there are only a few common factors amongst the various studies. However, detailed examination highlights that some of the factors suggested by Tam and Harris are covered in the first thirteen factors (out of sixty-seven factors) as found by Elhag and Boussabaine (1998). Those factors with asterisks in Table 3-10 (Elhag and Boussabaine's study) represent comparable factors identified in Tam and Harris's work as being significant. Although the names of the factors are not the same in the two studies, they are measuring similar (if not identical) attributes.

Table 3-13 Summary of Significant Factors Affecting Project Cost and Duration by Different Researchers

No.	Significant factors affecting project cost and duration	Tam & Harris (1996)	Khosrowshahi (1997)	Chan and Kumaraswamy (1999)	Elhag and Boussabaine (1999)
1	Complexity of project	✔			✔ (rank 8th)
2	Percentage of professional qualified staff	✔			(Rank 2nd)
3	Project leader's experience	✔			(Rank 2nd)
4	Contractor's past performance or image	✔			(Rank 13th)
5	Origin of the company	✔			
6	Architect's or client's supervision	✔			
7	Project nature		✔		✔
8	Building height			✔	✔
9	Labour availability			✔	(Rank 19th)
10	Presence/absence of facades			✔	

Note : the ranking in bracket is the overall ranking found by Elhag and Boussabaine (1998) in the first stage study of the factors affecting project duration and cost. A total of 67 factors were studied.

3.7.3 Factors Affecting Project Time and Cost Over-Runs

There are two studies in Hong Kong that are related to factors affecting project cost and time over-runs. One study is done by Dissanayaka and Kumaraswamy (1999), investigating the significant factors affecting project time and cost over-runs of Hong Kong construction projects. The factors were first identified by linear regression. Five factors were found to be significantly affecting project time over-run in Hong Kong:

1. levels of design complexity
2. levels of construction complexity due to sub-contracting
3. change orders / variations
4. client type, levels of client confidence in the construction team
5. levels of project team motivation and goal orientation

Four factors were found to be significantly affecting project cost over-run in Hong Kong:

1. levels of client confidence in the construction team
2. risk retained by client for quantity variations
3. levels of construction complexity related to new technology
4. payment modality (e.g. whether 'lump sum fixed price' or 'remeasure').

The data were then trained by ANN using the significant factors as inputs. The results of ANN prediction using the above factors was satisfactory, with 2.18 and 2.57 for the mean absolute percentage error (MAPE) for training set and testing set respectively in the time index model. The cost index model also produced reasonable MAPE; 0.05 (for training set) and 11.05 (for the testing set).

The other study is an extensive survey done by Kumaraswamy and Chan (1998) on the contributors to construction project delays. One hundred and forty-seven responses were received from contractors in Hong Kong for analysis. Descriptive statistics identified six common significant factors contributing to project time over-runs:

1. unforeseen ground conditions,
2. poor site management and supervision,
3. low speed of decision making involving all project teams
4. client-initiated variations,
5. necessary variations of works, and
6. inadequate contractor experience.

Other overseas researches related to project time and cost over-runs include Rosenbaum (1997), Kaming (1997), and Nkado (1995). Rosnaum (1997) highlighted political instability, fluctuations in currency, corruption, interest rates and material availability were critical factors affecting cost over-runs in developing countries. Research done by Kaming (1997) in Indonesia concluded that inflation, project complexity, weather conditions, project location and local regulation were the main contributors to cost over-runs of projects. Ten most important factors affecting project cost over-runs in Africa as identified by Nkado (1997) were : client's specified completion sequence, contractor's programming of works, form of construction, priority on construction time (by client), priority on construction time (by designer), complexity of project, project location, buildability of design, availability of construction management team, and completeness and timeliness of project information. Table 3-14 summarizes the significant factors identified in these studies.

Table 3-14 Summary of Significant Factors Affecting Project Cost and Time Over-runs by Different Researchers

No.	Significant factors affecting project time and cost over-runs	Dissanayaka and Kumaraswamy (1999)	Kumaraswamy and Chan (1998)	Rosenbaum (1997)	Kaming (1997)	Nkado (1995)
1	design complexity	✔				
2	construction complexity due to sub-contracting	✔				
3	change orders / variations	✔	✔			
4	client type, levels of client confidence in the construction team	✔				
5	project team motivation and goal orientation	✔				
6	construction complexity related to new technology	✔		✔		✔
7	payment modality	✔				
8	unforeseen ground conditions		✔			
9	site management and supervision		✔			
10	speed of decision making involving all project teams		✔			
11	contractor experience		✔			
12	political stability			✔		
13	fluctuations in currency			✔		
14	Corruption			✔		
15	interest rates			✔		
16	material availability			✔		
17	Inflation				✔	
18	weather conditions				✔	
19	project location				✔	✔
20	local regulation				✔	
21	client's specified completion sequence					✔
22	contractor's programming of works					✔
23	form of construction					✔
24	priority on construction time (by client)					✔
25	priority on construction time (by designer)					✔
26	buildability of design					✔
27	availability of construction management team					✔
28	completeness and timeliness of project information					✔

As some of the above studies are conducted in Africa and Indonesia, some critical factors associated with political environment were suggested (e.g. political stability, corruption, local regulation). As shown in Table 3-14, the findings from these studies are quite diversified. Nevertheless, there are several factors that are commonly cited by researchers as critical factors affecting project time and cost over-runs: change order/variations;

construction complexity (related to new technology); and project location. They serve as good indicators for further studies.

3.8 Attributes Affecting Estimating Accuracy

In order to achieve improvements in cost estimation practices, cost estimating method as well as estimation accuracy must be investigated. Based on analysis of past studies, Ogunlana and Thorpe (1991) concluded that the level of uncertainty (in a project) determined the accuracy of cost estimates. In a survey done by Fayek et al (1998) with the Australian contractors, 64% of contractors would assign allowance in the cost estimates after assessment of project risks. As highlighted before in Section 3.5.2.2 Accuracy of Project Overhead Estimates, the probable reason for the high variance among the bidders' estimates on project overhead costs is the difficulty to assess a number of qualitative factors involved in project overheads (i.e. the risk items) (Hegazy and Mosehli, 1995). In other words, the propositions of these two studies matched well with each other.

As reviewed in Sections 3.5.1.2 and 3.5.2.2, the overall accuracy of project cost / project overheads estimates is apparently low. Since construction works necessarily involves uncertainty (Poon, 2003; Ogunlana and Thorpe, 1991), understanding of the factors leading to uncertainty and risks in construction projects is vital to improve the estimating accuracy.

3.8.1 Risk Factors in Construction Projects

Most of the text books about construction risks only dealt with risk management principles or at most listed out some risk areas as examples. For instance, Park (1992) said that there were a lot of risks faced by the contractor like errors, delays, strikes, disaster, etc. Most of the authors are not attempting to generalize the factors or reasons that lead to common risk items in construction projects. Therefore, this Section focuses on the studies or researches related to risk factors conducted in the recent years. Several studies related to overseas and Hong Kong projects were chosen to provide a basis of understanding.

Hall and Hulett (2002) claimed in their study that complexity was a large source of risk. Complex projects with a lot of parties involved (like sub-contractors and suppliers) would increase the potential of conflicts and co-ordination problems. Datta and Mukherjee (2001) also mentioned that subcontractor coordination was a potential risk as delay in one area would cause a ripple effect in other areas. Elkington and Smallman (2002) commented that different project would have different risk. In other words, nature of project should be one of the factors as well.

In the study of Australian contractors' estimating practices, Fayek et al (1998) identified a list of risk factors that were considered by contractors when preparing tenders. More than 30% of the respondents cited the following factors as risk allowance factors (out of forty-seven factors), in their order of frequency:

1. Likelihood of unexpected job site conditions (58%);
2. Likelihood of unexpected climatic conditions (56%);
3. *Likelihood of a schedule delay* (50%);
4. *Uncertainty in estimate for materials* (44%);
5. *Uncertainty in estimate for labour* (42%);
6. Duration of contract period (39%);
7. Amount of contractor coverage (e.g. latent conditions) (39%);
8. Degree of safety hazard on project (36%);
9. Complexity in construction methods (36%);
10. *Uncertainty in estimate for plant / equipment* (33%);
11. Amount of liquidated damages (31%).

Although the above findings is generated from empirical data, some of the items are not risk "factors" in fact (as highlighted in italics). These items only represent consequences of project uncertainties, e.g. uncertainty in estimate for materials. They are not "causes" of project risk.

Another study by Ghosh and Jintanapakanont (2004) in Thailand, nine risk factors were identified using factor analysis. One hundred and twenty-two responses were collected from project participants regarding their views on risk factors affecting a rail project. The findings were:

1. *Delay risk*
2. Financial and economic risk
3. Risk related to sub-contractors (e.g. financial stability, competence)
4. Risk related to contractual and legal
5. Design risk (quality of design information)
6. Force majeure risk

7. Safety and social risk
8. Physical risk (of site and ground conditions)
9. Operational risk (risk related to productivity and availability of labour / plant / materials).

Similar to Fayek et al's study (1998), some of the items listed by Ghosh and Jintanapakanont are not risk factors at all. "Delay risk" (which is marked in italics) is the consequence or result experienced by the contractor as a result of project risk.

Moselhi (1997) selected some project-specific uncertainties to study project contingency estimation. The uncertainty items included : project procurement system, project state of technology, project location, project complexity, site accessibility, extent of project information, quality of design, project schedule (extent of tightness), and project procurement plans and policies. He also recommended that inflation and items like force majeure should also be added but they belonged to other categories. However, there is no detailed explanation by the author on how the project-specific uncertainty items are selected.

Besides, there are a few studies done in Hong Kong that are related to project risk assessment. When Charlton reviewed the contracts in the East Asia (2001), he criticized that a lot of risks were actually due to the onerous and demanding terms in the contract. Having similar view, Wong et al (2002) developed twelve factors associated with contract risk in Hong Kong through interviews with professionals in the industry. They were:

1. scale of project;
2. complexity of project;
3. experience of contractor;
4. degree of safety hazard;
5. client identity (whether he/she is stringent or reasonable, friendly);
6. architect identity (whether he/she is stringent or reasonable, friendly);
7. nominated sub-contractor identity (cooperative / claim orientated / trouble maker);
8. time constraint;
9. extension of time clause (responsibility transferable or not);
10. amount of liquidated and ascertained damages;
11. time limitation for submission of claims for loss and expense;

12. payment terms.

Shen (1997) surveyed the contractors in Hong Kong and suggested eight risk factors according to their order to importance:

1. insufficient or incorrect design information
2. variations in ground and weather conditions
3. subcontractors' manpower shortage
4. shortage of materials / plant resources
5. *poor coordination with subcontractors*
6. *poor accuracy of project programme*
7. shortage of skills / techniques
8. *abortive work due to poor workmanship.*

Nevertheless, Shen also failed to differentiate the risk factors properly from the risk consequences. The items in italics, poor coordination with subcontractors; poor accuracy of project programme; and abortive work due to poor workmanship are not risk factors.

To facilitate the interpretation of existing literature findings, a summary of factors is prepared in Table 3-15. Those items that are not related to risk factors as discussed earlier are deleted to avoid confusion. If the authors' intended meaning is known (by referring to the content of the publication), the item may be rewritten for easier understanding. In order to trim the number of factors, similar factors are grouped together under the factor item with a wider meaning (e.g. "degree of safety hazard" is grouped under "safety and social risk".

From the summary, several risk factors are suggested in more than one study. They are:

1. project complexity
2. site and ground conditions;
3. risk related to sub-contractors' performance and strength;
4. quality of design information;
5. safety and social risk;
6. Operational risk (risk related to productivity and availability of labour / plant / materials);
7. Project schedule (time constraint); and
8. Contract terms (e.g. EOT clause, amount of LD, time limit to submit claims).

Table 3-15 Summary of important risk factors suggested by different researchers

No	Important risk factors in construction projects	Hall and Hulett (2002)	Datta and Mukherjee (2001)	Elkington and Smallman (2002)	Fayek et al (1998)	Ghosh and Jintanapak-anont (2004)	Moselhi (1997)	Charlton (2001)	Wong et al (2002)	Shen (1997)
1	Project complexity	✔			✔		✔		✔	
2	Extent of subletting		✔							
3	Nature of project			✔						
4	Site and ground conditions				✔	✔				✔
5	Climatic conditions				✔					
6	Contract duration				✔					
7	Amount of contractor coverage (e.g. latent conditions)									
8	Financial and economic risk					✔				
9	Risk related to sub-contractors' performance and strength					✔				✔
10	Quality of design information					✔	✔			✔
11	Force majeure					✔				
12	Safety and social risk				✔	✔			✔	
13	Operational risk (risk related to productivity and availability of labour / plant / materials)					✔				✔
14	Project procurement system						✔			
15	Project location						✔			
16	Site accessibility						✔			
17	Project schedule (time constraint)						✔		✔	
18	Project procurement plans and policies						✔			
19	Contract terms (e.g. EOT clause, amount of LD, time limit to submit claims)				✔			✔	✔	
20	Scale of project								✔	
21	Experience of contractor								✔	
22	Client identity (whether he/she is stringent or reasonable, friendly)								✔	
23	Architect identity (whether he/she is stringent or reasonable, friendly) Payment terms								✔	
24	Nominated sub-contractor identity (cooperative / claim orientated / trouble maker)								✔	

3.9 Summary

In this Section, existing literature that is useful for this thesis was reviewed. Cost estimating theories were first examined. A thorough study on "project overheads" was then undertaken, including its general nature, estimation method recommended by literature, and related researches that had been taken to broaden the understanding of it. The theories of cost estimation were then cross-compared with the practice as revealed from various empirical studies in the industry. It is obvious that a gap exists between the principles laid out in the text books and the practices adopted by professionals, which gives rise to the potential error in estimates and monetary loss in projects.

The literature review was continued with the examination on contemporary cost estimation methods, particularly the artificial intelligent approach, to pave way for the research methodology of this thesis. Different ANN architectures were examined and compared. Studies on the use of ANN models to build cost estimating model in the construction industry were also reviewed. Most of them proved that ANN outperformed conventional methods like parametric cost estimating method.

As identification of input variables is always an important step to any statistical estimation model, factors affecting project overheads were also reviewed in the last part of this Section. However, due to the limited study in the factors affecting project overheads, other studies in relation like factors affecting project cost and duration; factors leading to project time and cost overruns; and risk factors of construction projects were also studied. Common factors cited by different researches were identified.

CHAPTER 4

Design of Exploratory Study of Project Overheads

4 Design of Exploratory Study of Project Overheads

4.1 Introduction

In Chapter 2, the methodology adopted of this research was explained. As mentioned before, there were two main parts of research work: the exploratory study of project overheads and the development of project overheads estimation model. In both parts, collection of opinions and project data from the contractors as well as statistical analysis of the data was involved.

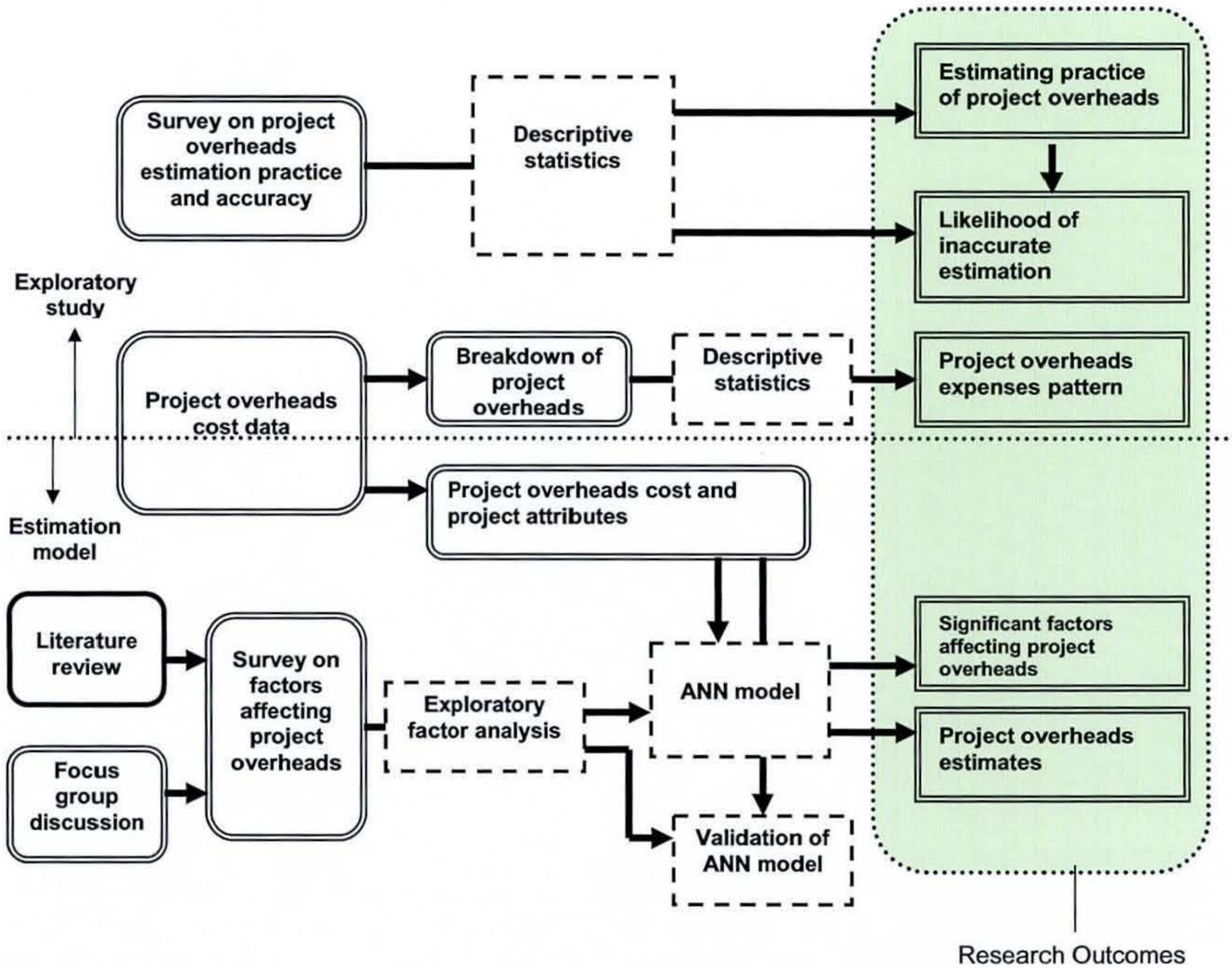

Figure 4-1 Alignment between Research Activities and Research Outcomes

As shown in Figure 4-1, the flow diagram illustrates the major research instruments and analysis tools employed; and how they were used to achieve the research outcomes.

To give a clear and smooth elaboration of the research process, the design of research instruments and data analysis tools in the various stages were grouped into three parts, each to be explained in one separate chapter.

Part 1: Exploratory Study of Project Overheads

1. 1^{st} stage survey and the follow-up interview to study project overheads estimating practice
2. 2^{nd} stage survey on estimating inaccuracy (in project overheads)
3. Cost data collection for the study of project overhead cost distribution.

Part 2: Development of Inputs and Outputs for Modelling

4. Focus group discussion on factors affecting project overheads
5. Opinion survey on the impact of factors affecting project overheads
6. Project and cost data collection for cost modelling.

Part 3: Development of ANN Cost Estimating Model

7. Design of ANN cost estimating model
8. Validation of the ANN model

To start with, in this Chapter, the details regarding the design of the research instruments (e.g. the design of questionnaire) in part 1 of the exploration study were examined and elaborated. Besides, design of the descriptive statistics used in this part was also discussed.

4.2 Exploratory Study of Project Overheads

Based on the limited study in project overheads, exploratory study of the actual practice of project overheads estimation was carried out first. The results from the study of industrial practice can justify the importance and need for a more accurate estimation model. As shown in Figure 4-2, after the literature review, surveys with the contractors in two different stages were designed to obtain the information related to the practice and difficulties of project overheads estimation. Besides, project overhead cost data was collected from contractors to have a complete understanding of project overheads expenses; in order to devise a realistic model to estimate project overheads.

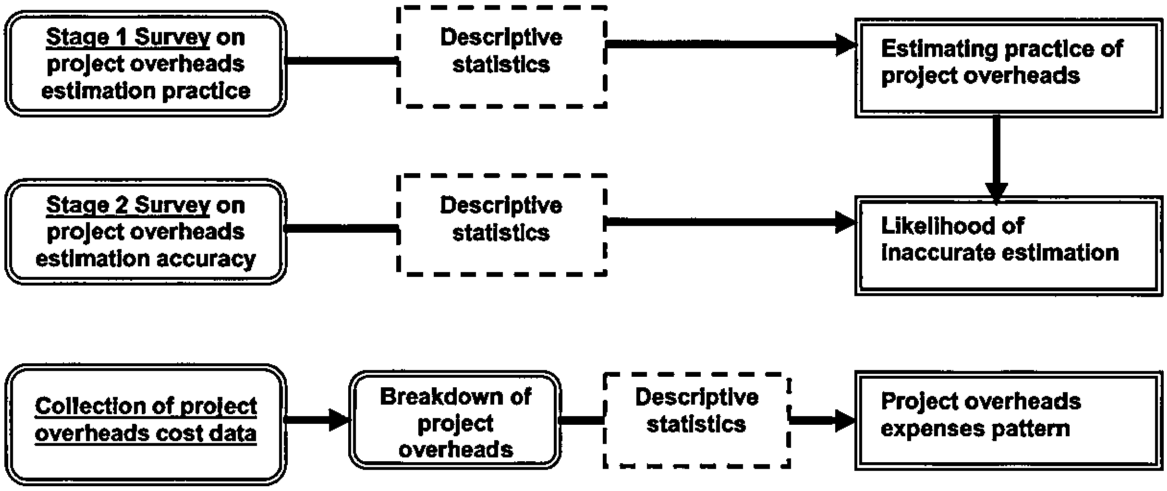

Figure 4-2 Schematic Work Flow in Exploratory Study of Project Overheads

4.2.1 Stage 1 Survey – Explore the General Practice of Project Overheads Estimation

Without much existing literature, the Stage 1 survey was aimed at exploring the general practices among the estimators in order to pave way for the next stage of survey. In view of this, a questionnaire was designed with the following specific objectives:

i) to identify the estimating methods (for project overheads) used in large contractors; and
ii) to measure the extent of these practices.

Since the information to be collected was rather straight-forward, the questionnaire was kept simple to collect the above information. This should encourage higher response rate.

4.2.1.1 Design of Questionnaire

To fulfill the objectives of studying the estimating practice of large contractors, the questionnaire was designed to collect information related to the following areas:

1. The method estimating project overheads used by the company.

   Here, contractors were asked about the general method they used to estimate the project overheads. Based on the past studies (Hegazy and Moselhi, 1995; Assaf *et al*, 1999; Tah *et al*, 1994 as mentioned in Literature Review Chapter), typical methods included estimation in detail with reference to the contract conditions, estimation as a percentage of the total value and estimation as a lump sum.

2. The method used to determine the allowance for individual project overhead items if the project overheads were estimated in detail.
   Since past studies as described earlier (Hegazy and Moselhi, 1995; Skitmore and Wilcock, 1994) revealed that estimators normally prepared the estimates based on their professional judgment, estimators who prepared the estimates of project overheads in detail were asked how their estimates were made, the use of experience or sophisticated estimation models were the two extreme practices. A third option was the build up of specific costs for each item as advocated by the CIOB Code of Estimating Practice. Respondents could also describe the specific method adopted by their companies if they did not use one of these methods.

3. The source of cost data used in determining the project overheads.
   Here, estimators were asked to identify the source of their cost data. Generally, cost data could be based upon:
   - Cost indices / schedule of rates published by the Hong Kong Government or large quantity surveying consultant firms in Hong Kong like Davis Langdon and Seah, and Levett & Bailey.
   - In-house cost databases
   - Quotations from subcontractors or suppliers for items that would be sublet or procured e.g. plant rental charges, scaffolding.

4. Feedback or review of estimation accuracy for the project overheads items.
   Respondents using in-house cost data were asked to state whether there was a feedback / review system on the estimating accuracy of project overheads in their companies. This was to see whether reliability of the in-house data was continuously reviewed by the companies.

A sample of the questionnaire is attached in Appendix A.

4.2.1.2 Stage 1 Survey - Follow-up Interviews

Following the Stage 1 survey, respondents were invited to a structured interview to collect more in-depth information regarding project overheads estimation practices. The interviews were conducted in Cantonese supplemented with the English terminology, allowing the senior estimators to express their views with greater freedom.

Since the interview was aimed at exploring the methodology of project overheads estimation used by the respondents, and to develop a list of significant project overhead items, the interview was designed with open-ended questions to cover three main areas as summarized in Table 4-1.

Table 4-1 Summary of the Content of Structured Interviews

Area of questions	Expected outcomes from the questions
1. Estimation method used by the company	- Identify:- Major component in estimating, - Staff involved in estimating project overheads, - Time spent in estimating, - Perceived risk transferability of estimates, - General accuracy of project overheads estimation
2. Project overhead items likely to be under- or over-estimated	- Identify at most 3 items likely to be under- or over-estimated by each interviewee
3. List of project overheads to be included for pricing/ study	- Identify the project overhead items that should be priced / studied

Details of the structured interview questions are listed in Appendix B.

4.2.1.3 Design of Descriptive Statistics

Stage 1 survey was a general study on estimating practice. Descriptive statistics like means and standard deviations are appropriate and sufficient to identify the facts related to estimating practice. Therefore, they were used as the basic method to analyze the data. Findings from structured interview were summarized to highlight the similarities or differences in the respondents' opinions.

4.2.2 Stage 2 Survey – Explore the Likelihood of Inaccurate Estimation

From the Stage 1 survey, general practice of estimating project overheads and the problem / importance of project overheads estimation should be verified. Besides, the typical project overhead items with cost implication were also identified. In Stage 2, a more in-depth examination of the likelihood of inaccurate estimation of project overheads was conducted. Based on the list of typical project overhead items generated in the Stage 1 survey, another survey was designed to find out the items that were more likely to be estimated inaccurately. As a result, the survey was designed to achieve the following objectives:

1. To explore the method of construction project overheads estimation used by large contractors in Hong Kong; and

2. To identify the critical construction project overhead items that possess a higher risk of genuine under- or over-estimation by the tenderers in Hong Kong.

4.2.2.1 Determination of Assessment Method of Estimating Accuracy

In terms of assessing the cost estimating accuracy to achieve the second objective, there are three main approaches used by the past researchers. The first approach is to measure the deviation from the lowest acceptable tender received in the competition (Morrison, 1984). The second approach is to measure the deviation between the pre-bid estimate and the contract sum signed (applied to consultant quantity surveyors' estimates only) (Ling and Boo, 2001; Gunner and Skitmore, 1999). The third one is to measure the perceived level of estimating accuracy from opinion survey of the estimators (Hegazy and Moselhi, 1995; Cheong, 1991; Jupp and McMillan, 1981). All the three methods have their drawbacks and merits, though they are not discussed by the users.

The first approach is mainly used to analyze the accuracy of pre-bid estimates, and hence the researchers only need to obtain the complete set of competitors' bids for a particular project from the consultant quantity surveyor. However, the whole comparison exercise is based on the assumption that the "lowest acceptable tender sum" is accurate which is of course, not necessarily true. Besides, tender bids always build in bidding strategy which is confidential. Therefore, it is difficult for the consultant quantity surveyor to assess the "accuracy" of the lowest bid without the true picture of the tenderers' bidding strategy. The second approach is in fact similar to the first approach which is usually applied to the analysis of pre-bid estimates. The only difference with the first method is the use of awarded contract sum instead of the lowest acceptable tender sum in the calculation. Comparing with the first method, this approach has the advantage of not requiring the complete set of tender bids from all tenderers. Sufficient data can be collected from those contractors who have the practice to keep track of the winners' bid amount. However, the problem of whether the winner's bid is accurate is not resolved either. In the third method, estimators (or respondents) are asked to indicate their under- or over-estimation level (as a percentage) in general. This method has the advantage of not requiring any confidential project data, thus allowing greater freedom to respondents. However, as criticized by some authors (Ashworth and Skitmore, 1982), this kind of opinion survey is likely to be over-optimistic as the respondents are usually biased to believe in their own ability. Some researchers found that there was a tendency of under-estimating the error level by the quantity surveyors when responding to opinion survey (Cheong, 1991).

Having considered all the drawbacks and merits of each method, the first and second methods were rejected. The main reason is there is no guarantee for the lowest acceptable tender sum or awarded contract sum to be free from estimating error. Using them as a benchmark for comparison contains inaccuracy. Besides, project overhead cost nature is very different from total project cost. It is meaningless to compare the project overhead prices of one bidder with the others because they can have entirely different site planning and resource allowance. Therefore, a mere comparison of the project overhead prices among different bidders is inadequate to conclude the level of accuracy in their estimates even if we ignore the bidding strategy of each bidder. As a result, the third method was chosen; to ask the estimators to indicate the likelihood of inaccurate estimation of each project overhead item. However, slight modification was adopted to reduce the effect of bias as criticized by Ashworth and Skitmore (1982). Instead of asking the respondents to give an exact percentage of the likely error in their own estimates, a 5-point likert scale was used to measure the likelihood of inaccuracy (1 = over-estimate to 5 = under-estimate). This also matched better with the purpose of the survey, which was to ask the "opinions" of the respondents about the likelihood of inaccurate estimation.

4.2.2.2 Design of Questionnaire

The questionnaire for this survey was to study the extent of estimation inaccuracy in project overhead items. It comprised of two main sections: the first part was to collect general information on the company nature and their methods of estimating project overheads. The second part was to measure the extent of under / over-estimation of different project overhead items. As discussed, the perceived level of estimating accuracy of project overhead items (from the point of view of estimators) was surveyed in the questionnaire. A 5-point likert scale was used to measure the likelihood of inaccuracy (1 = over-estimate to 5 = under-estimate).

To maintain consistency of response, the list of important project overhead items developed from the Stage 1 survey results was used here to measure the likelihood of estimate inaccuracy. The list of project overhead items covered the typical obligations of a contractor (e.g. testing, insurances, etc.) and the contractual arrangements (e.g. principles of measurement for bills of quantities, definitions of various contractual parties, contract conditions and amendments, etc.). Although contractors may not price explicitly against each item related to contract conditions in most cases, they will make provisions for these items somewhere (typically in the trade work items). For example, if the contractor has a poor working relationship with the contractual parties listed in the "definitions of various contractual parties" item, the estimator will make an extra allowance in the trade work, particularly in those items that require approvals from the parties concerned. Hence, in the second part of the questionnaire, the contractors were asked whether they were likely to over- or under-estimate the "financial impact" of the project overhead items.

The questionnaire used is attached in Appendix C.

4.2.2.3 Design of Descriptive Statistics

Stage 2 survey was an opinion survey to measure the level of estimating inaccuracy in project overhead items. Descriptive statistics like means and standard deviations were not enough to provide the full picture of the relative inaccuracy level amongst different project overhead items. Therefore, a Relative Accuracy Index (RAI) was used to rank the likelihood of under- and over-estimation among the various project overhead items. The Relative Accuracy Index (RAI) is adapted from the formula and methodology applied in a study done by Kumaraswamy and Chan (1998) to establish the relative rankings of the factors contributing to delay. In simple terms, an average score "relative" to the total highest score is calculated for each item in order to compute their relative rankings (by prioritising the items based on their average relative scores). Details of the formula of RAI are discussed below.

4.2.2.3.1 Analysis of Inaccuracy Level of Project Overheads Estimation

The practical consequence resulting from under-estimation is entirely different from that resulting from over-estimation. Under-estimation may give rise to a higher chance of success in the bid, but a future loss in the profit. Over-estimation may generate an uncompetitive bid. Hence, it is worth differentiating between items that are more likely to be over-estimated from those that are likely to be under-estimated. Since a 5-point scale ranging from over-estimation to under-estimation was used to measure the perceived estimation accuracy in the Stage 2 survey questionnaire, the mean scores in descriptive statistics could not help to identify separately the items that were most likely to be under-estimated nor those that that were likely to be over-estimated. To address this, the Relative Accuracy Indices that were adapted from the formula used by Kumaraswamy and Chan (1998) to calculate the relative rankings among the factors contributing to construction delay were applied here; so that the extent of over-estimation and under-estimation could be more clearly seen.

Three types of index were used, as follows,

- RAI_o (Relative Accuracy Index of over-estimation),
- RAI_u (Relative Accuracy Index of under-estimation), and
- RAI_i (Relative Accuracy Index of inaccurate estimation)

to differentiate the likelihood of over-estimation and under-estimation, and to give an overall indication of inaccurate estimation for each project overhead item.

4.2.2.3.1.1 Over-Estimated Project Overhead Items

According to the questionnaire, there are only two opportunities for an item to be considered as under-estimated. They are responses weighted in 4 and 5. In order to calculate the likelihood of over-estimation, responses that weighted 4 or 5 (representing under-estimation) were excluded when calculating the Relative Accuracy Index of over-estimation (RAI_o). In this way, the relative score (RAI_o) would only reflect the degree of over-estimation for each project overhead item.

The RAI $_o$, where the subscript '*o*' stands for over-estimation, is used to calculate the relative extent of over-estimation among the project overhead items.

$$RAI_o = \frac{\sum w_o}{A_o \times N_o} \qquad \textbf{(4-1)}$$

where RAI_o = Relative Accuracy Index of over-estimation;

w_o = weighting as assigned by each respondent in the range of 1 – 3; "1" representing over-estimation and "3" representing accurate estimation;

A_o = the highest weight (3);

N_o = total number of respondents assigning the weight in the range of 1 - 3.

Where two or more items had the same RAI_o, rank differentiation was achieved by the scores distribution of the items. If more respondents responded '1' in one item, the relative rank assigned to this item would be higher. In other words, the lower RAI $_o$ represented the higher likelihood of inaccurate estimation (i.e. higher ranking). The principle of ranking applied to the other RAIs as well.

4.2.2.3.1.2 Under-Estimated Project Overhead Items

In order to focus on the likelihood of under-estimation, the responses that weighted 1 and 2 were excluded when calculating the Relative Accuracy Index for under-estimation (RAI_u). Whilst maintaining consistency with the RAI_o for comparison, the responses in weights 4 and 5 were converted into 1 and 2 respectively to reflect the degree of under-estimation in the calculation of Relative Accuracy Index of under-estimation. Such conversion facilitated the direct comparison among the RAI_o and RAI_u values because a low value in both cases would imply a high likelihood of over- or under-estimation.

The RAI_u, where the subscript '*u*' stands for under-estimation, is used to calculate the relative extent of under-estimation among the project overhead items.

$$RAI_u = \frac{\sum w_u}{A_u \times N_u} \quad \textbf{(4-2)}$$

where RAI_u = Relative Accuracy Index of under-estimation;

w_u = converted weighting from each respondent in the range of 1 - 3, 1 represented under-estimation and 3 represented accurate estimation (assigned weight '4' was converted to 2 and assigned weight '5' was converted to 1, while assigned weight '3' remained unchanged);

A_u = the highest weight (3);

N_u = total number of respondents assigning the weight in the range of 3 - 5.

4.2.2.3.1.3 Inaccurately Estimated Project Overhead Items

To provide an overview of the inaccuracy in estimating project overheads, the Relative Accuracy Index of inaccurate estimation, RAI_i, was used to identify the degree of overall inaccuracy among the various items. In the questionnaire, all responses that were not '3' represented an inaccurate estimation. Hence, all responses to the items were included in the calculation of Relative Accuracy Index of inaccurate estimation. However, converted weightings must be applied to the responses that weighted 4 and 5 to reflect the same magnitude of inaccuracy contributed to over-estimation and under-estimation.

Hence, the RAI $_i$, where the subscript '*i*' stands for inaccurate estimation, is used to calculate the relative extent of inaccurate estimation among the project overhead items.

$$RAI_i = \frac{\sum w_i}{A_i \times N_i} \quad \textbf{(4-3)}$$

where RAI $_i$ = Relative Accuracy Index of inaccurate estimation;

w_i = converted weighting from each respondent in the range of 1 - 3, 1 represented highest tendency of inaccurate estimation and 3 represented accurate estimation (assigned weight '5' was converted to 1 and assigned weight '4' was converted to 2 to represent different degrees of inaccurate estimation, while assigned weights from 3 – 1 remained unchanged);

A_i = the highest weight (3);

N_i = total number of respondents = 40.

4.2.2.3.2 Cross-comparison of Relative Rankings Among Different Nature of Companies

As the sampled contractors were different in nature, including building contractors, civil engineering contractors, and contractors that engaged in both building and civil engineering works, further analysis was done to cross-compare their relative inaccuracies of project overheads estimation. A Rank Agreement Factor was used to measure the relative agreement in rankings between groups. The formula and methodology was developed by Okpala and Aniekwu (1988) and was used by Kumarswamy and Chan (1998) to compare the relative rankings between groups. The Rank Agreement Factor (RA) is the average absolute difference in the ranks of the items. With the RAs, the percentage agreement (PA) and the percentage disagreement (PD) can be computed.

The PD (Percentage Disagreement) is defined as:

$$PD = \frac{RA}{RA_{max}} \times 100\% \quad \textbf{(4-4)}$$

$$PD = \frac{\left(\sum_{i=1}^{N} | R_{i1} - R_{i2} |\right)}{N} \div \frac{\sum_{i=1}^{N} |R_{i1} - R_{j2}|}{N} \times 100\% \quad \textbf{(4-5)}$$

$$PD = \frac{\left(\sum_{i=1}^{N} | R_{i1} - R_{i2} |\right)}{\left(\sum_{i=1}^{N} | R_{i1} - R_{j2} |\right)} \times 100\% \quad \textbf{(4-6)}$$

where PD = Percentage Disagreement;

RA = Rank Agreement Factor;

RA_{max} = Maximum Rank Agreement Factor (see Appendix F for the calculation);

R_{i1} = relative rank of the i^{th} item in group 1;

R_{i2} = relative rank of the i^{th} item in group 2;

R_{j2} = relative rank of the $(N-i+1)^{th}$ item in group 2 (under the scenario that the two groups are in maximum rank difference);

N = total number of items for ranking = 23.

If looking at Equation (4-6), the numerator can be regarded as the summation of all the rank differences (in their absolute values). The denominator refers to the summation of the maximum rank difference among the available set of items.

Then, PA = 100% – PD **(4-7)**

where PA = Percentage Agreement;

PD = Percentage Disagreement.

4.2.3 Collection of Project Overhead Cost Data

In the Stages 1 and 2 surveys, opinions regarding project overhead estimating accuracy and methods were collected. In order to have a complete picture of the project overheads estimation, actual cost data on project overheads must be collected. The main purpose of this data collection process was to establish the cost distribution or cost pattern of project overhead expenses in construction projects.

4.2.3.1 Project Overheads Data to be Collected

Most of the construction contracts in Hong Kong contain a bill section with itemized list of the project overheads. This allows the tendering contractors to price each project overhead item so that project owners can know more precisely how the project overhead cost is, and to pay the contractors accordingly during the project period or prolongation period. On the other hand, large contractors in Hong Kong normally maintain a good cost breakdown structure to monitor their cash flow and profitability of the projects. Therefore, in terms of collecting project overhead cost data from the industry, there are three main sources to choose:

1. contract bill with price of various project overhead items;

2. contractor's internal estimate of various project overheads cost;
3. contractor's actual expense of various project overhead items.

The first type is comparatively the easiest one to collect, as this can be obtained from consultant quantity surveying firms or contractors. However, the use of such data is limited due to the difficulty to reveal the bidding strategy hidden in the prices. In other words, it is difficult to confirm whether the prices put down in the contract document are being the true estimates of the contractors, or the likely expenses of the items. Both types two and three data represent the "true" internal data of the contractors, and is more difficult to collect due to confidentiality. However, without the full calculation details and the resources allowance breakdown, the type two data alone is rather difficult to interpret and not very useful. Type three data, which is the actual expenses of the project overhead items, requires lesser supportive / explanatory documents. More importantly, the objective of this study is to develop an estimation model for project overheads, and hence knowing the actual expenses of project overheads will be very useful to enhance the accuracy of the model.

Therefore, balanced with the different nature of different data source, contractor's actual expense of project overhead items was collected for providing insights to the distribution pattern of the project overheads cost.

4.2.3.2 Design of Cost Breakdown Template to Collect Cost Data

Since the data collected would be used to identify the distribution pattern of project overhead costs, breakdown of project overhead cost upon completion of project had to be collected. To facilitate the future analysis of the cost data, a template was provided to the contractors to enter the project overhead itemized cost. The design of the cost breakdown structure was to a large extent based on the project overhead items list developed in the Stage 1 Survey. The list of significant project overhead items developed in Stage 1 survey was shown in Table 4-2.

Table 4-2 List of Significant Project Overhead Items

	Project overhead items
1	Definitions of various contractual parties*
2	Description of works*
3	Nature of site and site inspection*
4	Site possession and completion*
5	Site management, watchman and attendance to NSCs
6	Principles of measurement of the Bills of Quantities*
7	Drawings to be prepared by the Main Contractor
8	Contract conditions and amendments
9	Methods of measuring and valuing variations*
10	Fees and levies
11	Restrictions to noise and dust nuisance
12	Insurances
13	Protection of finished works
14	Protection of adjacent / existing works
15	Safety precautions
16	Facilities to be provided by Main Contractor : plants, scaffolding
17	Site offices, stores, latrines
18	Setting out
19	Samples, mock ups
20	Testing
21	Power and water supply
22	Temporary works e.g. hoardings, temporary roads, signboard
23	Cleaning and removal of rubbish

However, in normal circumstance, contractors will not account cost for the following items (those marked with asterisks Table 4-2) due to the nature of cost control purpose.

1. Definitions of various contractual parties,
2. Description of works,
3. Nature of site and site inspection,
4. Site possession and completion,
5. Principles of measurement of the Bills of Quantities, and
6. Methods of measuring and valuing variations.

Therefore, these items were deleted from the template provided to the contractors. Furthermore, some of the items in the above Table 4-2 were sub-divided into two items to enable comparison with Solomon's study (1993). For example, *Facilities to be provided by Main contractor* was broken down into "mechanical plants" and "scaffolding"; *Power and water supply* was also broken down into "power" and "water" and "*Site Management, watchman and attendance to NSCs*" was sub-divided into "site management" and "watchman". To ensure clear interpretation of the items, "telephone and fax" was also separated from the facilities provided by main contractor. The final list provided to the contractors for data collection was shown in Table 4-3.

Table 4-3 Template Provided to Contractors for Collection of Project Overhead Itemized Cost

	Project overhead items	HK$
1	Site management	
2	Watchman	
3	Drawings to be prepared by the Main contractor	
4	Contract conditions and amendments	
5	Fees and levies	
6	Restrictions to noise and dust nuisance	
7	Insurances and surety bond	
8	Protection of finished works	
9	Protection of adjacent / existing works	
10	Safety precautions	
11	Mechanical plants	
12	Scaffolding	
13	Site offices, stores, latrines	
14	Setting out	
15	Samples, mock ups	
16	Testing	
17	Power	
18	Water	
19	Telephone & fax	
20	Temporary works e.g. hoardings, temporary roads, signboard	
21	Cleaning and removal of rubbish	
	Total :	

4.2.3.3 Design of Descriptive Statistics

Due to the difficulty to collect a large number of samples, patterns of cost distribution across different items were only analyzed by descriptive statistics. Mean percentages of itemized cost were compared with past literature findings. Besides, cross-examination of the itemized cost data was made with the perceived inaccuracy level of individual item (as found in Stage 2 survey) to highlight clues for management attention.

4.3 Summary

In this Chapter, details of the research instruments used in the exploratory study of project overheads were examined. The design of questionnaires for the Stage 1 and Stage 2 surveys in the exploratory study were portrayed. The former one was a simple questionnaire whereas the second one was an opinion survey to measure the perceived level of accuracy in estimation. The respective data analysis techniques used were also explained and illustrated. The exploratory study of project overheads mainly involved descriptive statistics, except for the analysis of the estimating inaccuracy which involved the calculation of relative importance index and percentage agreement. Such calculation was based on other researchers' work and was explained thoroughly in the Chapter for better understanding. Regarding the analysis of cost distribution of project overheads, development of the breakdown structure of project overhead cost to be collected was also explained. It provided a standard framework for data collection which facilitated subsequent analysis. From the analysis, general cost distribution of the project overhead items could be ascertained.

CHAPTER 5

Development of Inputs and Output to Cost Model

5 Development of Inputs and Output to Cost Model

5.1 Introduction

In the exploratory study, the facts on the practice and accuracy of project overheads estimation in the industry were established. Proceeding from the exploratory study would come to the development of cost estimating model. In this Chapter, details of the development of the inputs and output of the estimating model were described. Besides, the details of project and cost data to be collected for model were also discussed. Figure 5-1 depicts the various processes in the development of inputs and output for the ANN cost estimating model. Details of the design of the ANN model will be explained in the next Chapter.

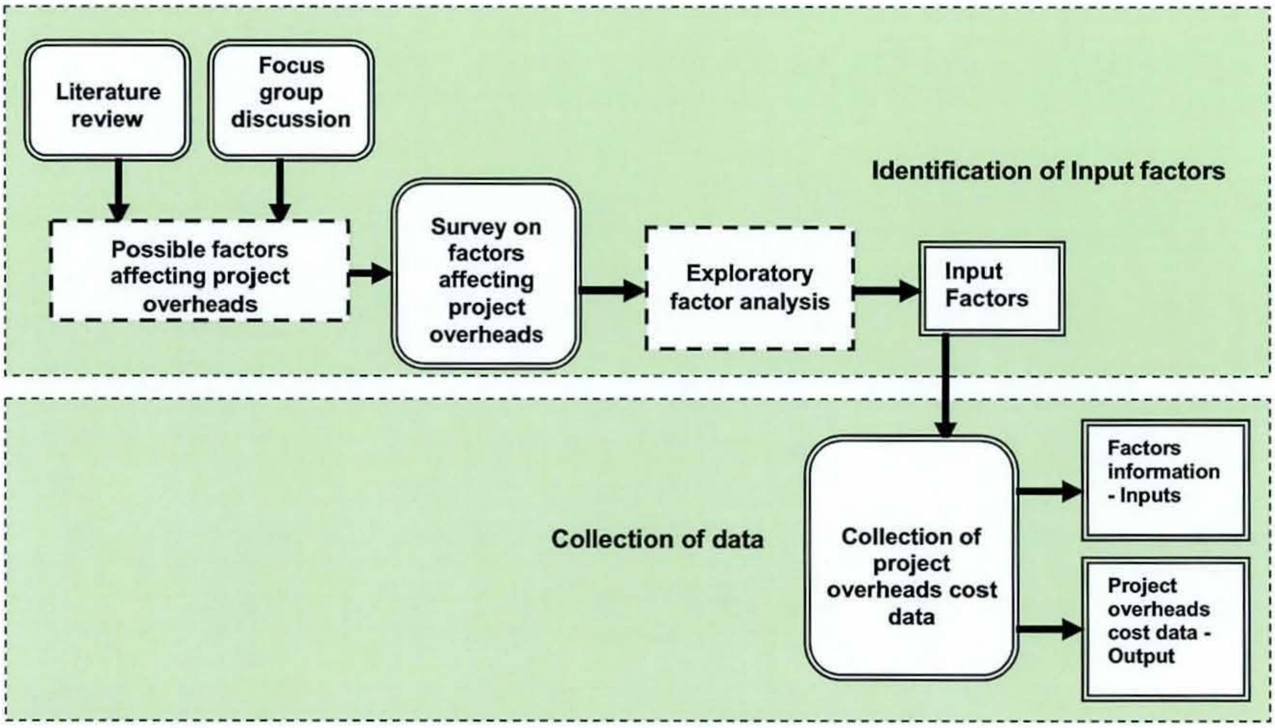

Figure 5-1 Schematic Work Flow in Development of Inputs and Output

5.2 Identification of Factors Affecting Project Overheads

Identification of the factors or variables is a fundamental step in the development of any statistical model. As examined in the Literature Review, the existing literature related to project overheads is very limited. There are only two studies directly related to this subject area, one by Assaf et al (1999) and the other one by Solomon (1993). The insufficient literature leads to the development of a 3-stage approach to compile a final list of factors for the estimating model. Firstly, a wider spectrum of literature was reviewed to generate a list of possible factors affecting project overheads. Besides the factors affecting project overheads suggested by the Assaf et al (1999) and Solomon (1993), those related to the factors affecting estimating accuracy; project time and cost; and overruns could be good references to develop the list of factors affecting project overheads. Secondly, views were collected from a focus group of practitioners about any other possible factors not identified. A focus group of senior quantity surveyors with industrial experience above twenty years were invited to advise on the possible factors affecting project overheads expenditure. This was to avoid any missing of possible critical factor. Thirdly, opinion survey was conducted to review the impact of the listed factors on project overheads. The results were then statistically analyzed to generate principal factors for the model.

5.2.1 Development of Preliminary List from Literature Review for Focus Group Discussion

First of all, studies on factors affecting project overheads were examined. The factors suggested by the two studies, Assaf et al (1999) and Solomon (1993), were included in the preliminary list. However, the "need of work" and the "number of competitors" (by Assaf et al, 1999) were excluded because the objective of this research was to develop a model to predict project overheads more accurately, not to help the contractor to win the project through strategic pricing of the project overheads. These two factors, nonetheless, only affected one's bidding strategy decision but not the actual expense on project overheads. Similarly, "how much cash the contractor has available" (suggested in Assaf et al (1999)'s study) affected mainly the pricing strategy (e.g. bid or not to bid, a higher or lower markup) but not exactly influenced the expenditure of project overheads. Hence, it was also excluded.

Besides, factors affecting estimating accuracy; project time and cost; and overruns were abstracted to compile the preliminary list of factors affecting project overheads. The common factors suggested by more than one study related to the above subjects were included in the preliminary list. The respective summaries are made in Tables 3-13, 3-14 and 3-15; and explained in Section 3.7. These common factors were summarized in Table 5-1 for easy reference. In the studies related to factors affecting project cost and duration, project complexity, percentage of professional qualified staff, project manager's experience, contractor's past performance / image, origin of company, architect's or client's supervision, project nature, and building height are cited by more than one studies as critical (Table

3-13). In the researches related to project time and cost overruns, changes change order/variations; construction complexity (related to new technology); and project location are cited by more than one studies as critical factors (Table 3-14). In terms of the construction risk studies, project complexity; site and ground conditions; risk related to sub-contractors' performance and strength; quality of design information; safety and social risk; operational risk (risk related to productivity and availability of labour / plant / materials); project schedule (time constraint); and contract terms (e.g. EOT clause, amount of LD, time limit to submit claims) are suggested by more than one studies as important (Table 3-15).

Table 5-1 Summary of Factors Affecting Project Overheads, Construction Risk, Project Cost and Duration Extracted from Past Literature

Project overheads variables	Variables affecting project time and cost	Variables affecting project time & cost overruns	Construction risk factors
project duration			
Location / site condition		project location	
type of project	project nature		
extent of sub-letting			
equipment available to contractor			
shape of structure to be constructed			operational risk (risk related to productivity and availability of labour / plant / materials)
nature of structure frame			
staff and labour size			
size of mechanical plant fleet			
shape of site			
client's requirement			
project size			
complexity of project	project complexity	construction complexity	project complexity
payment schedule			
the client's strictness in supervision			
type of contract			
site layout			
	percentage of professional qualified staff		
	project leader's experience		
	contractor's past performance or image		
	building height		
		change order/variations	
			ground conditions
			risk related to sub-contractors' performance and strength
			quality of design information
			safety and social risk
			project schedule (time constraint)
			contract terms (e.g. EOT clause, amount of LD, time limit to submit claims)

Some of the factors above were common in nature, e.g. "project complexity" and "construction complexity". Hence, a preliminary list containing twenty-eight factors was consolidated, as shown in Table 5-2.

Table 5-2 Preliminary List to Measure the Impact of Factors Affecting Project Overheads (Based on Literature Review)

Item	Preliminary Factor Description	Revised Factor Description
1	Duration of project	
2	Type of project	
3	Scale of project	
4	Building height	
5	*Complexity of project*	*Complexity of project (use of special techniques / plants / tradesman)*
6	Shape of site	
7	Shape of structure to be constructed	
8	Nature of structure frame	
9	*Location*	*Accessibility of site*
10	*Site layout*	*Site coverage of proposed building*
11	*Client's requirement*	*Required quality level of project*
12	Type of contract	
13	Payment schedule	
14	Quality of design information	
15	*Change order/variations*	
16	The client's strictness in supervision	
17	Contract terms (e.g. EOT clause, amount of LD, time limit to submit claims)	
18	Ground conditions	
19	Extent of sub-letting	
20	Risk related to sub-contractors' performance and strength	
21	Size of mechanical plant fleet	
22	Equipment available to contractor	
23	Project leader's experience	
24	Contractor's past performance or image	
25	Staff and labour size	
26	Percentage of professional qualified staff	
27	Safety and social risk	
28	Project schedule (time constraint)	

Some of the items were rephrased to suit the subject of the study better and to facilitate interpretation of the factor. The items to be rephrased were highlighted in italics in Table 5-2 and the revised item descriptions were shown next to the original description. The rationale of revision is explained as follows:

1. "Complexity of project" was expanded to "complexity of project (use of special techniques / plants / tradesman)". This item was cited by various researchers as important factors affecting project overheads (Solomon, 1993; Assaf et al, 1999), project cost and duration (Tam and Harris, 1996; Elhag and Boussabaine, 1999), project time and cost overruns (Dissanayaka and Kumaraswamy, 1999; Rosenaum, 1997; Nkado, 1995) and construction risk (Hall and Hulett, 2002; Fayek et al, 1998;

Moselhi, 1997; Wong et al, 2002). Although the term used in these studies varied, they all referred to the construction complexity in relation to use of new techniques or require special technicians to carry out the operations. Therefore, to differentiate from other possible complexities like complex procedures / documentation required by the client or complex procurement of sub-contractors; the expanded version helped to give a clearer picture to the readers.

2. "Location" was changed to "accessibility of site". According to the literature, location was included as one of the factors affecting project overheads as it affected accessibility of the site, which might give rise to additional temporary roads or hardstandings to be constructed for the site. However, "location" is a term leading to the address of the site. The implication of accessibility is not spelt out. Unless the readers are familiar with the site locations, they cannot immediately visualize the degree of site accessibility. Therefore, this item was rephrased into "Accessibility of site" to avoid difficulty in interpretation.
3. "Site layout" was changed to "site coverage of proposed building". Site layout was too general and difficult to illustrate which aspect was being referring to. Since the use of mechanical plants was included in the preliminary list already, the focus here was put on the available space for work and materials on the site. Hence, for simplicity, the term was changed to "site coverage of proposed building".
4. "Client's requirement" was changed to "required quality level of project". The item was suggested by Solomon as client's required quality could affect project overhead expenses. Although client's requirement is normally related to quality standard of the works, it will create lesser risk of misinterpretation if the "quality" aspect required by the client is specified in the factor description.

5.2.2 Design of Focus Group Discussion

With the preliminary list of factors generated from the very limited existing literature on project overheads, a focus group of senior quantity surveyors were invited to advise on the possible factors affecting project overheads expenditure. All the senior quantity surveyors had more then twenty years industrial experience. They were invited from the large contractors in Hong Kong and had more than ten years' experience in the position of a project quantity surveyor.

5.2.2.1 Analysis of Focus Group Comments

The views of the focus group were cross-examined with the revised list of factors generated from the literature review. Further addition, combination, and modification of the factors was made to develop a final list of possible factors affecting project overheads which was comprehensive and easy to understand for further analysis.

5.2.3 Opinion Survey on Impact of Factors Affecting Project Overheads

Having consolidated the comments from the focus group discussion, a complete list of possible factors affecting project overheads was prepared. However, as in Boussabaine and Elhag (1999) and others' work, only the critical factors should be included in the ANN estimation model. The list was thus used to develop the questionnaire to survey the impact of factors affecting project overhead expenses. The results from the survey would then be scrutinized by appropriate statistical method to generate the principal factors so as to use as inputs for the estimating model.

5.2.3.1 Design of Questionnaire

To facilitate future analysis, the consolidated factors were divided into external variables (variables that are uncontrollable by the contractor, e.g. project nature, contract conditions, etc.) and internal variables (variables that are related to the internal resource of the contractor and hence within its control / influence e.g. staffing, sub-contractors competence, external relations, etc.).

As a result, the questionnaire was divided into three parts containing closed type questions. Part 1 requested general information about the respondents. Part 2 contained the external variables and part 3 contained the internal variables for respondents to evaluate.

Respondents were asked to rate the impact of each variable on project overheads expenditure. Similar to typical opinion surveys, the significance of impact attributable to each variable was measured by a 3-point likert scale ranging from very significant impact =3 to slight impact = 1. Respondent could assign a zero rating to the variable if he/she thought that the impact was negligible or nil. Usually, 5-point scale was used in opinion survey to rank from for example, strongly agree to strongly disagree. In that case, the rating involves 2-sided opinions, with the central scale (point 3) as neutral. Unlike conventional 5-point scale application, the rating to be put against each variable in the current study was, in fact, single-sided in nature (i.e. from nil to significant impact). Therefore, the scale was not reduced in absolute term. Although the chosen 3-point scale was apparently narrower than the conventional type, similar application was also found in Elhag and Boussabaine's work (1998) to study variables affecting project cost and duration. The same type of 3-point scale was used and was proved to be a suitable approach.

The complete questionnaire with all the final list of factors is detailed in Appendix D.

5.2.3.2 Design of Statistical Analysis Tool - Exploratory Factor Analysis

The survey results provided the primary data for extraction of principal factors. Since the priori knowledge about project overheads is limited, exploratory factor analysis was used to scrutinize the factor list, based on the opinions of respondents in the survey. This method can summarize the relationship among various factors in a concise manner and to eliminate preconception (Gorsuch 1983). The advantage of exploratory factor analysis is to identify a smaller number of factors that can be used to represent relationships among sets of many interrelated variables (Norusis 1993). It is widely used by researchers in construction management to identify and interpret non-correlated clusters of variables (Chan et al, 2004; Fang et al, 2004; Trost and Oberlender, 2004; Ghosh and Jintanapakanont, 2004). Therefore, it was a suitable statistical tool to be used here to extract a small number of project overhead factors for the estimation model. As a result, the responses from the opinion survey on impact of factors affecting project overheads were analyzed by exploratory factor analysis to extract the principal factors for the ANN model.

As explained in earlier section, the variables in the opinion survey were group into external variables and internal variables. Therefore, the external and internal variables were analyzed separately in the factor analysis procedure so as to reflect the contractor's perspective more precisely.

5.2.3.2.1 Computer Package to Use

SPSS v. 10 software package was used to carry out the exploratory factor analysis.

5.2.3.2.2 Preliminary Analysis

Before extracting the factors, two basic assumptions of factor analysis had to be tested: multivariate normality and sampling adequacy (Lattin et al, 2003; George and Mallery, 1999; Grimm and Yarnold, 1995). With SPSS, Bartlett's test of sphericity is provided to measure multivariate normality of the variables. It also tests whether the correlation matrix is an identity matrix (i.e., a spherical set of multivariate data) (Lattin et al, 2003; George and Mallery, 1999). Besides, KMO test is available in the software to measure whether the distribution of values is adequate for conducting factor analysis (George and Mallery, 1999).

If KMO test results for a variables set is higher than 0.5 (the highest value is 1), it implies that the distribution of the values in the matrix is adequate for conducting factor analysis (George and Mallery, 1999). For the Bartlett's test of sphericity, high value indicates that the data is approximately multivariate normal and acceptable for factor analysis (George and Mallery, 1999; Lattin et al, 2003). The high value in the Bartlett's test results also confirms that the correlation matrix cannot be concluded as an identity matrix (Lattin et al, 2003).

5.2.3.2.3 Extraction of Components

If the data is proved to be multivariate normal and the correlation matrix is not an identity matrix, components (or factors) can be extracted by Principal component analysis. In the extraction process for the two sets of variables, the number of components retained was based on the eigenvalues. As suggested by Gousuch (1983), only components with eigenvalues larger than 1.00 were retained. To achieve a simple structure by minimizing the tendency towards a "general" component in the solution, Varimax rotation (of the factor loadings) was adopted (Gorsuch 1983). It is the most popular method which attempts to find new axes to represent the factors while restricting the new axes to be orthogonal to each other. In statistical terms, the orthogonality of the rotated factors is equivalent to the fact that they are uncorrelated to each other (Afifi et al 2004, p.402-403). Varimax procedure attempts to find a rotation so that the squared loadings are close to one or zero (Gorsuch 1983). If loadings are either zero or one, the factor will be most interpretable because the numerous small loadings that cause difficulties are eliminated.

After rotation, salient variables in each component were identified to explain the meaning of the factors. As recommended by Gorsuch (1983), the salient variables were selected by two criteria:

- sufficiently high absolute value of the loading (0.4 as the minimum);
- distribution of each variable's loadings across all the factors (the variable would be considered as related to that factor if it only loaded on that factor).

Besides the factor loadings, weighted average of each variable was also calculated to provide further understanding on the level of importance of these variables as perceived by the respondents. This method is recommended by Shash (1993) in analysing the ordinal data collected in likert scale; and subsequently adopted by other researchers to rank factors in similar manner (Kometa et al. 1995, Chan and Kumaraswamy 1998, Ghosh and Jintanapakanont 2004). The importance index was calculated using the following formula:

Importance index = $\sum (aX) * 100/3$ **(5-1)**

where a is the constant that expresses the weighting assigned by the respondents, ranging from 0 to 3; and $X = n/N$, where n is the frequency of the responses; and N is the total number of responses.

With the external and internal factors extracted by the factor analysis procedure, the input variables for the artificial intelligent estimating model were thus established.

5.3 Collection of Project and Cost Data for Modelling

After identifying the input variables, project data could be collected for modelling. Collection of project overheads cost data is not easy due to the confidentiality nature. However, to proof the accuracy and viability of the estimating model, the use of real project data was essential. On the other hand, since the existing literature is limited, no secondary data is found applicable to the study. Hence, project data from large contractors were collected. Two sets of information were required from each project: total project overheads cost (output data) and the project data measuring the extracted factors (input data).

Literature was reviewed to check the way to measure the input variables. However, very little literature gives such details. Only studies by Elhag and Boussabaine, 1998 & 1999 and Tam and Harris, 1996 reported the measurement of variables and are discussed below.

5.3.1 Considerations for Measurement Level

Unlike the output which was a cost item measured in the monetary value, some inputs were rather abstract in nature, e.g. complexity, external relationships. Therefore, the way to measure the inputs had to be identified first. As shown in Table 5-3, most of the attributes in Elhag and Boussabaine's studies (1998 and 1999) were measured by nominal-level. For instance, type of foundation was measured by strip; pad; raft; and piles. Some attributes were measured quantitatively by ratio-level e.g. duration of project (in weeks), size of project (in gross floor area).

Table 5-3 Selected Attributes for Projection of Project Cost (Elhag & Boussabaine, 1998) and Modelling of Project Duration (Elhag & Boussabaine, 1998)

No.	Factors	C / D	Category 1	Category 2	Category 3	Category 4
1	Type of client	C	Public	Private		
2	Type of building	C	Office	Industrial		
3	Deadline	C	Normal	Tight	Fast track	
4	No. of storey	C	Number			
5	No. of basement	C	Number			
6	Frame structure	C	Load-bearing	steel	Concrete	Average
7	Location	C	Rural	City outskirts	Inner city	
8	Site conditions	C	Green	Demolition	Average	
9	Type of foundation	C	Strip	Pad	Raft	Piles
10	Access	C	Unrestricted	Restricted	Highly restricted	Average
11	Complexity of building services	C	Low	Medium	High	
12	Procurement method	C	Traditional	D & B	Management	
13	Site topography / slope of site	C	Leveled site	Gentle sloping	Landscapes	Steep sloping site
14	Working space	C	Unrestricted	Restricted	Highly restricted	Average
15	Ground conditions	C	Non-contaminated	Contaminated	Average	
16	Type of soil	C	Good	Moderate	Poor	
17	Mark-up %	C				
18	Need for work (%)	C				
19	Duration	C	weeks			
20	Area	C, D	m^2			
21	Tender selection method	C, D	Negotiated	Competitive		
22	Type of contract	D	JCT	D&B		
23	Fluctuation in prices	D	Fixed	formula		
24	Contract sum	D				

In case of Tam and Harris's study (1996), a lot of the variables were measured by likert scale, i.e. ordinal-level. For instance, past performance of the project manager was measured by 3-point scale of 1 (slightly poorer than average) to 3 (better than average). Another example like client's strictness in supervision, 5-point likert scale of 1 (very loose) to 5 (very tight control) was used. There were attributes measured by ratio-level as well, e.g. quality of management team was measured by the quotient of number of professionally qualified staff and total number of staff.

Determining the measurement level for the variables is important as it affects the form of statistical analyses to be used. Regarding to the use of artificial neural network (the proposed modelling technique in this study), past ANN researchers like Boussabaine and Elhag (1999) used a wide range of measurement levels for the input variables. However, according to the NeuroShell 2 User's Manual, neural network expects each type of input to represent the strength of the input neuron, e.g. rainfall ranging from 0 to 25 inches. In other words, nominal-level measurement should not be used.

Having the restriction of not to use nominal-level measurement, only the other three levels of measurement were used in this study. Higher priority was given to the measurement level with higher objectivity and suitability. For instance, gross floor area of the project (ratio-level measurement) is more objective than using scale of 1=very small to 5= very large (ordinal-level measurement). However, it might not be suitable if more than one variable was grouped together after factor extraction. Consider if "past experience of project manager in similar project", "technical competence of site management staff" and "interpersonal skills of project manager" were grouped under the factor "strength of site management team" after factor analysis. Originally, the first two variables could use ratio-level measurement. E.g. number of years of experience could be used to measure the past experience of the project manager; and number of staff having professional qualification could be used to measure technical competence of site staff. However, when the three variables combined as one factor, it was impossible to measure with continuous variables. In this case, likert scale (ordinal-level measurement) would be the choice.

For the output variable (i.e. the total project overheads cost), total project overhead expenses of completed projects were collected.

5.3.2 Determination of Measurement Level

As discussed, no project information should be measured by nominal-level in order to fit in the artificial neural network modelling, as the magnitude of input represents the strength of input neuron. Therefore, when collecting the project data and information related to the input factors, ratio-level or ordinal-level measurement was used. From the results of the exploratory factor analysis, fourteen input factors were identified (which will be explained in detailed in the Data Analysis Chapter). Details of the measurement nature of the fourteen principal factors were tabulated in Table 5-4.

Table 5-4 Measurement of 14 Input Factors

	Principal factors	Abbreviations	Measurement-level	Remarks
1	Design and variations	DES_VO	ordinal-level	likert scale 1 to 5
2	Economic condition during construction stage	ECON	ratio-level	average of government indices
3	Bonds and Warranties	BOND	ordinal-level	likert scale 1 to 5
4	Project complexity	COMP	ratio-level	weighted average of scales and no. of storeys
5	Procurement nature	PROCU	ordinal-level	likert scale 1 to 5
6	Site and building shape	SHAPE	ordinal-level	likert scale 1 to 5
7	Location	LOCAT	ordinal-level	likert scale 1 to 5
8	Project size	SIZE	ratio-level	gross floor area
9	Strength of site management team	TEAM	ordinal-level	likert scale 1 to 5
10	Contractor's past relationship with project parties	RELAT	ordinal-level	likert scale 1 to 5
11	Extent of staff and plants assigned to project	STA_PLT	ordinal-level	likert scale 1 to 5
12	Financial strength of sub-contractors	SC_FIN	ordinal-level	likert scale 1 to 5
13	Sub-contractors performance	SC_PER	ordinal-level	likert scale 1 to 5
14	Contractor is client's subsidiary	SUBSID	ordinal-level	0 or 1

Most of the factors were measured in a rather straight-forward manner, using likert scale ranged from 1 (very low or very poor) to 5 (very high or very good). For *Economic condition during construction stage*, since the factor was the combination of the economic condition of Hong Kong and the economic condition of the construction market during the construction stage, published data from the government could be made use of. Firstly, the economic condition could be represented by the Consumer Price Index while the economic condition of the construction market could be measured by the Gross value of construction work performed. Both data were complied regularly by the Census and Statistics Department of the Hong Kong SAR Government and thus reliable. The combined index for economic condition would be calculated by the weighted average of the construction volume and the consumer price index (the respective weightings were based on the relative importance index found together with the exploratory factor analysis), using year 2000 as the base year. Details of the indices and calculation will be shown in the Data Analysis Chapter.

Another input factor, *project complexity* was measured by a combination of several variables: complexity of project (use of special techniques / plants / technicians), number of level of basement (of proposed structure), type of Project (e.g. commercial, residential, industrial), and number of storey of proposed structure. Since there is no established method to measure complexity of a project, the design here was based on the results of *project* *complexity* factor extraction. As commented by the professionals in the industry, if the *project complexity* factor was measured by a single likert scale, respondents might found difficulty to incorporate the impact of all the four variables. Therefore, a weighted average score was calculated from the respective characteristics of the four variables, collected as per Table 5-5 below.

Table 5-5 Measurement of Each Variable Contained in the Project Complexity Factor

Variables contained in the principal factor – Project Complexity	Measurement-level	Remarks
Complexity of project (use of special techniques / plants / technicians)	ordinal-level	likert scale 1 to 5
No. of level of basement (of proposed structure)	ratio-level	no. of basement levels
Type of Project (e.g. commercial, residential, industrial)	ordinal-level	scale 1 to 6 based on Tam & Harris (1996)'s measurement of project complexity[5]
No. of storey of proposed structure	ratio-level	no. of storeys (superstructure)

The last factor, *Contractor is client's subsidiary*, was measured by 0 (representing that the contractor was not the client's subsidiary) or 1 (representing that the contractor was the client's subsidiary).

As a result, a proforma was designed to facilitate the collection of project data which is shown in Appendix E. Contractors who provided the data were invited to fill in the proforma and returned.

5.4 Summary

In this Chapter, the development of the inputs and output for the ANN cost estimating model was examined. The input variables were determined by thorough literature review, followed by focus group review of the preliminary list of factors. The opinion survey which was designed to collect the views on impact of those factors affecting project overheads was discussed with the questionnaire presented. Details of the exploratory factor analysis technique used to extract principal factors from the preliminary list were illustrated. The factors extracted were used as input factors of the ANN model. Finally, the considerations in

[5] According to Tam & Harris (1996), 6 levels were used to measure project complexity based on the nature of work : 1 – foundation work, site formation, etc., 2 – renovation or alteration works, 3 – factory or domestic housing which required a little amount of M&E co-ordination, 4 – deluxe housing project or office building which required more subcontracting and M&E co-ordination, 5 – hotel or high-class office building, 6 – hospital or complicated project.

determining the measurement level of inputs and outputs were explained and recommended with appropriate reference to the literature.

CHAPTER 6

Design of ANN Cost Estimating Model

6 Design of ANN Cost Estimating Model

6.1 Introduction

In the Literature Review Chapter, characteristics of the common algorithms were examined. In this Chapter, the features and advantages of various estimating techniques and modelling algorithms were compared. Details of the proposed ANN model, including its architecture, were portrayed. Besides, various methods of validation were introduced to assure the reliability of the proposed model.

6.1.1 Analysis of Alternative Cost Estimating Techniques

In the Literature Review Chapter, features of different contemporary cost estimating techniques were discussed. To determine which cost estimating technique to be used in this study, a comparison was made among three alternative estimating models other than the classical approach. They are artificial neural network, fuzzy-logic and linear regression.

Table 6-1 Summary of Comparison Among Different Cost Estimating Techniques

Features	Importance	ANN	Fuzzy-logic	Linear Regression
Enable single output to be estimated	Very important	✔	✔	✔
No priori information about the relationships between inputs / output required	Very important	✔		
Satisfactory accuracy from past researchers	Very important	✔	✔	
No complicated mathematical knowledge required	Very important	✔		✔
Identification of significant factors	Very important	✔ (GMDH)		✔
Allowed non-linear relationship	Important	✔	✔	
Allowed linguistic relationship to be defined	Not important		✔	
Commercial computer software package available to execute the modelling function	Important	✔		✔

From Table 6-1 above, ANN method obviously provides more advantages compared with the other two. In fact, from the review of various contemporary techniques of cost estimation in the Literature Review Chapter, artificial neural network is widely adopted in the construction management researches. From the work of Emsley et al (2002), Boussabaine and Elhag (1999), Moselhi and Siqueira (1998) and the like, ANN cost estimation model was a reliable estimating tool with proven accuracy. In most cases, ANN out-performed other cost estimation models like regression. More important, when there is a lack of priori information about the relationships between the input and output variables, ANN can still model the input factors satisfactorily (Zhou et al, 1996; Garza and Rouhana, 1995; Boussabaine and Elhag, 1999). Although linguistic relationship is only available in fuzzy logic techniques, it is not particularly important as ANN is empowered to determine the relationships automatically.

In particular with GMDH network, a "best formula" together with the significant variables are identified when training stops. Such level of transparency in the final model is found in the GMDH network which allows greater understanding of the cost estimating function and the factors contributing to the estimate. Although the mathematical theories of GMDH algorithm are very complicated, the architecture for the study is rather simple with the help of commercial software package. Therefore, ANN was proposed in this research to build up the estimating model.

6.1.2 Design of ANN Cost Model

With the entire data ready, the neural network for estimating project overheads can be built. Several initial decisions have to be made including the computer software package and the training architecture used.

6.1.2.1 Computer Package to Use

There are a few commercial packages in the market for neural network modelling, including MatLab and NeuroShell 2. NeuroShell 2 is a popular software used by researchers in the construction industry (Moselhi and Siqueira, 1998; Hegazy and Ayed, 1998) and is much cheaper than MatLab. It is developed by the Ward Systems Group, Inc. The data in this thesis was therefore presented to NeuroShell 2 Release 4.0 package for training and testing.

6.1.2.2 Data Sets

In the last section, the measurement level of factors was identified. They were well complied with the basic requirement – magnitude of data represented the strength of the input neurons. On the other hand, according to the NeuroShell 2 User's Manual (1995), inputs should always contain the same type of data. This, however, was difficult to achieve as different input variable represented different type of data, e.g. gross floor area (in the scale of say 7,000m^2 or more), complexity of job (only from 1 to 5 measured by likert scale), duration of project (may be around 350 days or more), etc. Therefore, normalization of the data must be done before any training. This was done by applying the scaling function to the data.

The two main numeric ranges for the neural network to operate were 0 to 1 and -1 to 1. As the nature of the project data in the proposed ANN model only contained positive values (no negative values for likert scale, floor area, or no. of storeys), the numeric range adopted should be 0 to 1. Since there was the possibility for the model to encounter input data with higher values than the maximum value in the current sample, the inputs were not clipped at the top (i.e. scaled to <<0,1>> in NeuroShell 2). In other words, if data from 0 to 100 is scaled to <<0, 1>>, then a later data value of 150 will be scaled to 1.5. The same type of scaling was also applied to the output data for more efficient operation of the model.

6.1.2.3 Design of Model structure

All the literature agrees that there is no hard and fast rule for designing the neural network model structure (Boussabaine and Elhag, 1999; Li, 1995; Ward Systems Group, Inc., 1995; etc.). One of the reasons is because there is little knowledge about the relationship between the inputs and output. As priori information about the relationships between inputs and outputs, the corresponding functions, etc. were not available, Group Method of Data Handling algorithm was chosen to build the ANN architecture. Table 6-2 gives a simple comparison of supervised neural networks and the GMDH network.

Table 6-2 Comparison of GMDH Network and Supervised Neural Networks

	GMDH network	Supervised networks
Modelling principle	Inductive in nature	Deductive in nature
Ability of self-organization	No. of layers and no. of nodes are estimated by minimum of external criterion (objective choice)	No. of layers and no. of nodes are tried by users (subjective choice)
Identification of estimating function	Best formula with the minimum error is identified automatically	Not provided
Identification of significant variables	Automatic selection of essential input variables	Not available

As summarized in Table 6-2, GMDH can overcome the problem of supervised networks (in e.g. backpropagation, probabilistic and general regression networks) that require users to set a lot of parameters by trial and error. Besides, the optimal estimating function together with the significant inputs being used can be clearly identified in the GMDH model solution. Such level of transparency is only available in GMDH model. It fitted precisely the objective of this thesis (to identify the significant factors to predict project overheads). Figure 6-1 illustrates the proposed network architecture for this study.

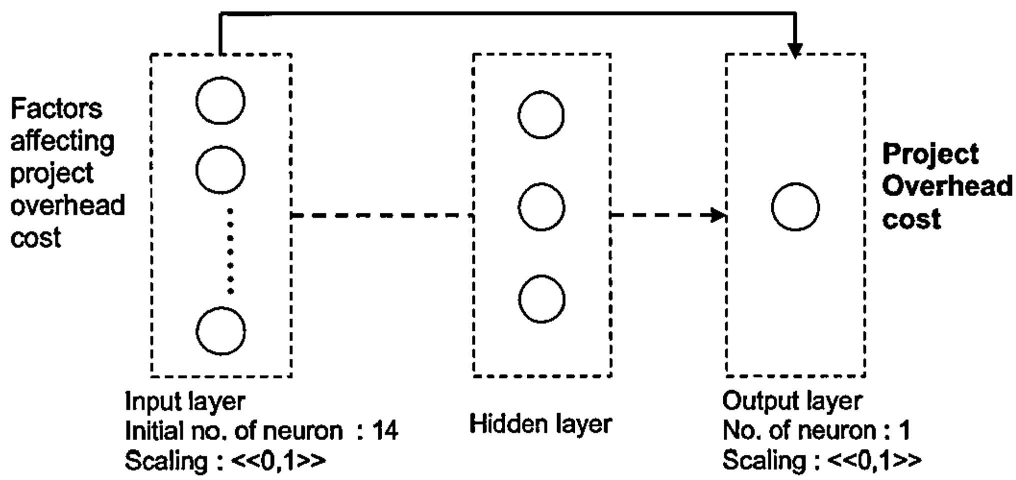

Figure 6-1 Proposed GMDH Neural Network for Project Overhead Cost Estimating Model

6.1.2.4 No. of Layers and Nodes

Unlike other supervised models, the optimal number of hidden layers is found by the GMDH algorithm automatically. Thus, this was not required to design. For the input layer, initially, all the variables extracted from the exploratory factor analysis were used as the inputs neurons. For output layer, there was only one neuron and that was the project overheads cost. The "best" model can be found from optimization and adjustment of the model parameters although there is no formal method to derive the neural network structure for a given estimation problem (Boussabaine and Elhag, 1999). As a result, when the significant variables were identified in the initial training, the non-significant variables would be removed to simplify the model and re-trained again. The process was continued until no further variables could be removed from the input neurons. In that case, the simplest model was arrived if the estimating accuracy was satisfactory.

6.1.2.5 Training Criteria

The "advanced training criteria" was chosen to allow more flexibility to set the training parameters, and to release the full power of GMDH (Ward Systems Group, Inc., 1995). Since GMDH has the power to give up unnecessary terms or variables at different stages of the algorithm to improve the model, different model optimization strategies can be opted. Smart mode of model optimization was chosen as it could, in most cases, optimized the tradeoff between calculation speed and model quality.

6.1.2.6 Choice on the Form of Polynomials

Since little priori information was known about the relationship between the inputs and output, largest freedom should be allowed for the GMDH algorithm to model for any non-linearity of the pattern. Therefore, the maximum input variables allowed to pass to the next layer for survival in each connection was set as 3. This was the maximum allowable number in NeuroShell 2 and was recommended in most cases (Ward Systems Group, Inc., 1995). Besides, the maximum product term in each connection was set as (X1·X2·X3), which meant that covariants and trivariants were allowed for non-linear modelling. This was also the maximum allowable product term in NeuroShell 2 and was also recommended (Ward Systems Group, Inc., 1995). Regarding the variable degree in connection, X^3 was chosen. The choice for these three settings concerned with the form of each polynomial could provide the maximum degree of non-linearity to be built for survival.

The maximum number of survivors affects the diversity of survivor models and the choice of significant variables. Thus, to give more flexibility to model the function, the maximum number of survivors allowed to pass on to the succeeding layer as inputs was set as the number of initial inputs.

6.1.2.7 Choice on Selection of Survivors

Selection criteria is emphasized as the most important parameter to be set in GMDH network as it is the "objective function" to determine how the survivors should be selected (Ward Systems Group, Inc., 1995). As discussed in the Literature Review Chapter, there were 6 criteria for selection, including FCPSE (Full Complexity Prediction Squared Error), PSE (Prediction Squared Error), MDL (Minimal Description Length), GCV (Generalized Cross Validation), FPE (Final Prediction Error) and Regularity. In this regard, all the six criteria were tested sequentially to find out the best model for the estimating function.

6.1.2.8 Diagnostic tool

In the search of the best neural network model, a lot of trials are involved. Hence, it is important to have an objective method to diagnose the results. Several tools in NeuroShell 2 are available to assess the training and testing performance of the ANN models:

1. Mean squared error = mean of (actual value – predicted value $)^2$.
2. Mean absolute error = mean of | actual value – predict value |.
3. R squared (the coefficient of multiple determination)

$$= 1 - \frac{\Sigma (a - \hat{a})^2}{\Sigma (a - \bar{a})^2} \qquad \textbf{(6-1)}$$

where a is the actual value, $\hat{a}$ is the predicted value, and $\bar{a}$ is the mean of the a values.

(A perfect fit will give an R squared value of 1, a very good fit near 1, and a very poor fit near 0.)

These indicators helped to identify which setting or structure of the model was the best, i.e. giving the highest accuracy in prediction.

On the other hand, when diagnosing the best GMDH model (e.g. comparing GMDH models produced by different selection criteria), besides judging from the above indicators, the "best formula" of each model was also examined. A simple formula requiring fewer inputs was considered as a better model than very complicated formula achieving similar accuracy.

6.1.3 Training the ANN Model

A major set of data (90%) can be extracted from the sample randomly for training, left behind a smaller data set (10%) for testing. Unlike backpropagation where transfer functions and training rules have to be set, the main parameter required in GMDH is the selection criteria (evaluation methods) to determine when the training should stop. As there is no priori knowledge of which selection criterion fits the proposed model, all the six methods of evaluation were tested to identify the best model for the study.

Successive training would be carried out with fewer inputs upon the identification of significant variables by the GMDH network; so as to develop the simplest model with satisfactory accuracy. Furthermore, significant factors could be identified automatically by the model.

6.1.3.1 Design of Training

With the initial parameters decided as detailed above, the model could be trained accordingly. All the six selection criteria in NeuroShell 2 (including FCPSE (Full Complexity Prediction Squared Error), PSE (Prediction Squared Error), MDL (Minimal Description Length), GCV (Generalized Cross Validation), FPE (Final Prediction Error) and Regularity) were tested sequentially to search for the best model.

6.1.3.2 Design of Retraining of the GMDH Model With Fewer Inputs

As mentioned before, six selection criteria were available and would be trained sequentially to provide six different "best" models for each set of inputs. The significant variables from each models were compared to shortlist the significant ones based on the following conditions:

1. The variable was identified as significant in the largest number of models.
2. The variable was identified as significant in the "best" model amongst the six models generated from the various selection criteria.

Based on the above conditions, if the variable(s) used in the current training was / were not significant as identified by the GMDH algorithm, such variable(s) was / were removed from the inputs and trained again by six selection criteria. The process was continued until the GMDH model identified all (or almost all) the inputs used as significant variables.

6.1.4 Design of Validation Methods

There were several validation methods applied to the model, both quantitative and qualitative in nature. Quantitative methods included cross-validation, comparison with alternative ANN architectures and comparison with conventional regression method. Qualitative method was an appraisal of the proposed model by a focus group of professional quantity surveyors. The design of the various validation approaches are explained below.

6.1.4.1 Design of Cross-Validation

The representativeness and sufficiency of the sample is often a concern in statistical analysis. Unfortunately the response from the sample is not under the control of the researchers. Therefore, a dilemma often occurs between the choice of a research topic that is not explored in the past and the ease to access the required data for the research. In case of this study, the data required for building the cost estimating model involved confidential project cost data from the contractors. It was foreseeable that great resistance from the contractors would be received to reveal their confidential cost data to external party. Therefore, reliability might be skeptical if the sample size is small. To overcome this problem, the data was cross-validated using the method suggested by Tam et al (2005). The sample of seventy-one cases was divided into nine sub-groups, each containing eight cases (and one sub-group contained seven cases). The cross-validation (for all six selection criteria) was repeated eight times for the whole sample; using eight sub-groups for training and one for production in turn. The rotation pattern amongst the nine sub-groups of data was illustrated in Table 6-3.

Table 6-3 Rotation Pattern of the 9 Sub-groups of Data

	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training sub-groups	1,2,3,4,5,6, 7,9	1,2,3,4,5,6, 8,9	1,2,3,4,5,7, 8,9	1,2,3,4,6,7, 8,9	1,2,3,5,6,7, 8,9	1,2,4,5,6,7, 8,9	1,3,4,5,6,7, 8,9	2,3,4,5,6,7, 8,9
Production sub-groups	8	7	6	5	4	3	2	1
	The set used for training / production in the original GMDH model was : Training sub-groups : 1,2,3,4,5,6,7,8 and Production sub-group : 9							

Note : 1 to 9 represent the names of the nine different sub-groups of data

Results obtained from each cross-validation round were reviewed to check if the best model gave satisfactory predictions in all rounds.

6.1.4.2 Comparison with Alternative ANN Architectures

One of the purposes of this research is to identify the most significant factors affecting project overhead cost, and thus GMDH algorithm was chosen in the ANN model. Nevertheless, whether the "significant" variables identified in GMDH network results could generalize well in other estimating models was another concern. Therefore, to prove the generalization of the significant variables to estimate project overheads, the variables were used as inputs and tested in other ANN architectures. This led to the second validation method which was to verify the reliability of the significant variables identified by GMDH algorithm. ANN architectures like backpropagation and general regression networks were experimented as they were applicable to predictive nets and popular in the research domain. While probabilistic neural network was designed for categorized outputs, it would not be used here.

6.1.4.2.1 Design of Alternative ANN Architectures for Comparison

The most popular structure of backpropagation network was the standard 3-layer backpropagation and the multi-layer feed-forward network (MLFF) with backpropagation learning algorithm. The former one is a standard and simple model but may require more hidden layers for more complex prediction. The latter one contains three hidden slabs to allow different transfer functions to be applied. It provides three different ways to view the data. As a result, different features in the data patterns can be detected during training. Thus, it was recommended as a useful predictive network (Ward Systems Group, Inc., 1995; Tam et al, 2004). In this regard, MLFF network was chosen to represent the backpropagation network for validation purpose. Besides backpropagation network, general regression network was also chosen for validation purpose. Figure 6-2 and Figure 6-3 show the MLFF network structure and the GRNN structure respectively for validation of the GMDH model.

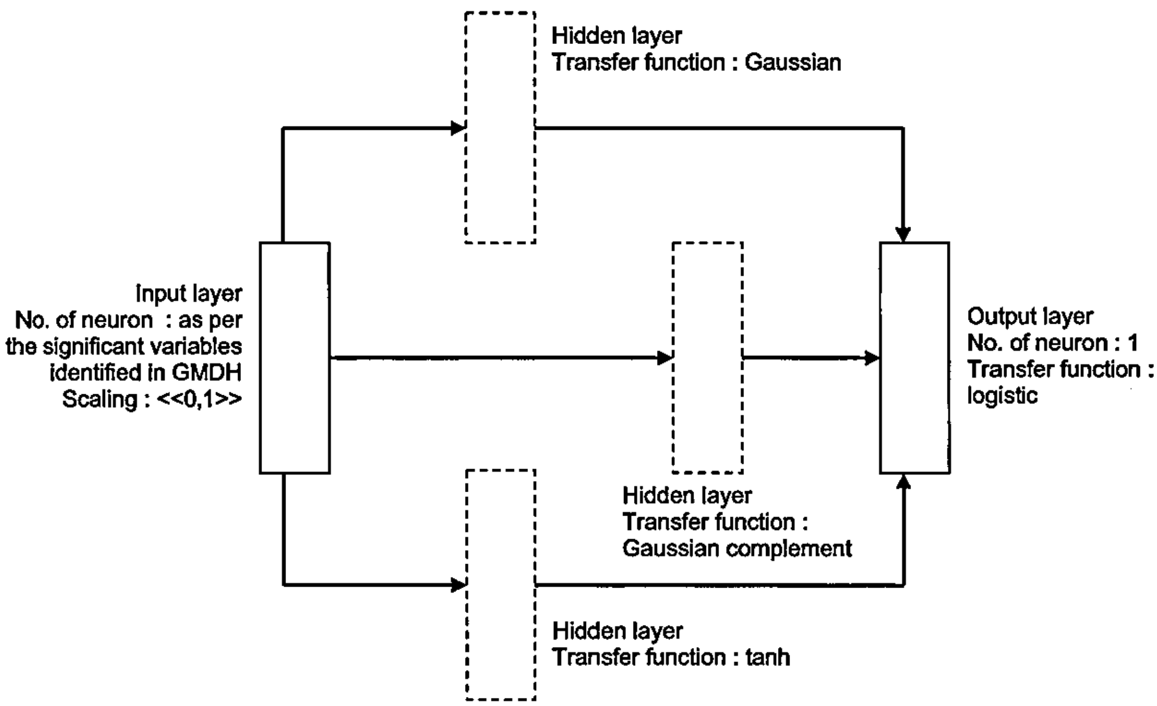

Figure 6-2 Multi-layer Feed-forward Network Structure Used in Validation

Since the purpose of this comparison was to verify the validity of the significant variables identified in the GMDH network, not to compare the predictive power of different architecture, the setting of these two models were made simple and the parameters were set to the default values. In case of the MLFF model, the transfer functions were set as default as shown in Figure 6-2. Default values that were also chosen for other parameters included :

- learning rate : 0.1;
- momentum : 0.1;
- initial weights : 0.3.

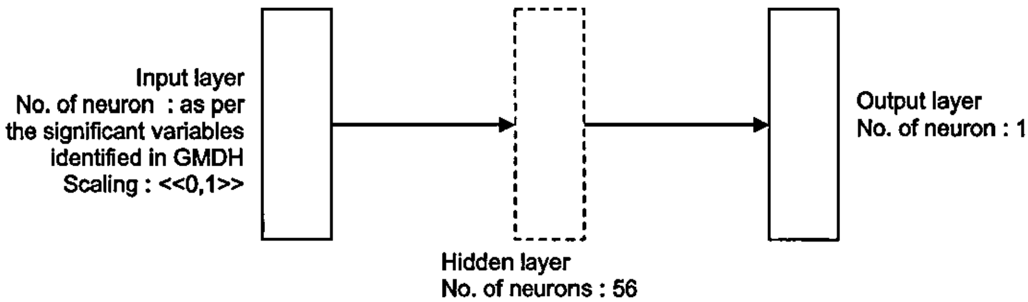

Figure 6-3 General Regression Neural Network Structure Used in Validation

For the GRNN model, no training parameter as learning rate or momentum was required. The smoothing factor (which was the only training parameter) was derived after the network was trained. The default number of neurons in the hidden layer was equal to the number of training patterns / cases. The most significant variables identified in the best GMDH network were used as the inputs to these two alternative models. The same set of training, testing and production data were presented to these two models for consistency. The results produced by these two models were then compared with those produced by the GMDH network. If the predictive results in the alternative models were also satisfactory, the variables could be regarded as reliable factors to estimate project overheads.

6.1.4.3 Comparison with Conventional Regression Model

As an alternative to neural network model, a multiple linear regression model was implemented to compare the predictive performance with each other. SPSS v.10 is a powerful and easy to use tool to carry out statistical analysis, and hence being used to work out the regression equation.

The project overhead cost estimating relationship, describing the project overhead cost (C) as a function of a number of variables as determined in the factor analysis (X_1, X_2, X_3, ...X_y). The general form of the equation was

$$C = F(X_1, X_2, X_3, \ldots X_y).$$

$$\mathbf{C = A + B_1X_1 + B_2X_2 + B_3X_3 + \ldots + B_yX_y.} \quad \textbf{(6-2)}$$

All the cases in the sample were presented to SPSS to build the regression relationship. The principal factors identified from the exploratory factor analysis were the independent variables of the regression model, and the total cost of project overheads was the dependent variable. Stepwise regression method was proposed to avoid the adverse impact to the overall predictive validity from other variables when one variable was chosen as significant (George and Mallery, 1999). The test could provide results on the error level for comparison. Besides, the regression equation produced from SPSS could also be compared against the best formula produced from GMDH network.

6.1.4.4 Design of Focus Group Appraisal

After various validation processes like cross-validation, validated with other ANN architectures and regression method, it should be confident that the reliability and accuracy of the proposed was secured. However, whether the model was acceptable to the practitioners might require further consideration. Since artificial neural network is not familiar amongst most of the practitioners, views from them had to be collected in a medium that allowed more interaction and explanation to respondents. As a result, focus group review was chosen to appraise the model. The senior quantity surveyors who had joined the earlier stage focus group discussion were invited to comment on the attractiveness of the proposed model.

It is likely that none of the focus group member has knowledge about ANN. Therefore, a brief introduction of the concepts of ANN and its application would be given. Following the explanation on the basic idea of the proposed model, members would be invited to evaluate the proposed model against five attributes, including:

1. Easy to understand
2. Simple to use
3. Time saving
4. Level of accuracy
5. Likelihood to apply.

Further comment was also invited from the group. The whole discussion was conducted in Cantonese.

6.1.4.4.1 Analysis of Focus Group Comments

The views of the focus group were compiled and average scores for the attributes were calculated. In case of individual comments, they were summarized for further analysis.

6.2 Summary

In this Chapter, the rationale of choosing artificial neural network and GMDH algorithm as the estimating model was examined and discussed. Details regarding to the design of the ANN model was developed. The model was explained including full details of architecture design and training criteria. The design of the training and re-training, including the diagnostic tool to achieve the best model was also examined.

Validation of the model was an integral part of a modelling process. Several methods were chosen to match with the nature and purpose of the study. Cross-validation was designed to reduce the reliability problem associated with small sample size. Comparing the model with alternative ANN and regression models was recommended and illustrated to validate the reliability of the significant factors. Besides, errors from different estimating models could help to verify the accuracy of the proposed model. Appraisal of the final model by industrial professionals was designed as the last test. The proposed model could be recommended for use if the testing and validation results were satisfactory.

CHAPTER 7

Data Analysis

7 Data Analysis

7.1 Introduction

In this Chapter, the data collected and the associated analytical results are presented. Since a systematic flow diagram was developed in the Methodology Chapter to explain the detailed stages of work involved in the whole research process, the data analysis and findings are explained according to the same flow. Major findings of the whole research process include:

1. Survey results and descriptive analysis on the project overhead estimation practice and accuracy.
2. Survey results and descriptive analysis on the likelihood of inaccurate estimation of each project overhead item.
3. Descriptive analysis of the project overhead cost distribution.
4. Focus group discussion results on the list of factors affecting project overheads.
5. Survey results and factor analysis on the critical factors influencing the expenditure of project overheads.
6. Training, testing, production and validation results on the estimating model of project overheads built by artificial neural network.

7.2 Exploratory Study of Project Overheads

7.2.1 Stage 1 Survey – Explore the General Practice of Project Overheads Estimation

7.2.1.1 Response to Stage 1 Survey

There were altogether one hundred and nineteen Group C contractors in the list of Approved Contractors and Suppliers for carrying out the public works as at April 2002. Forty-nine responses were received out of the sample, representing 41.1% response rate which is considered reasonable and acceptable.

7.2.1.2 Survey Results

7.2.1.2.1 Methods of estimating project overheads

As shown in Figure 7-1, forty-six respondents (94%) of the sample responded that the project overheads were estimated in detail with reference to the contract conditions. The other three contractors (around 6%) estimated the project overheads as a percentage of the total value of measured works. This illustrates that a very high proportion of estimators will spend their time and effort in preparing estimates for the project overheads. The percentage of detailed estimation was higher than the situation in other countries like the U.S.A. (83% in Hegazy and Moselhi's study, 1995) or Saudi Arabia (71% in Assef et al's study, 1999).

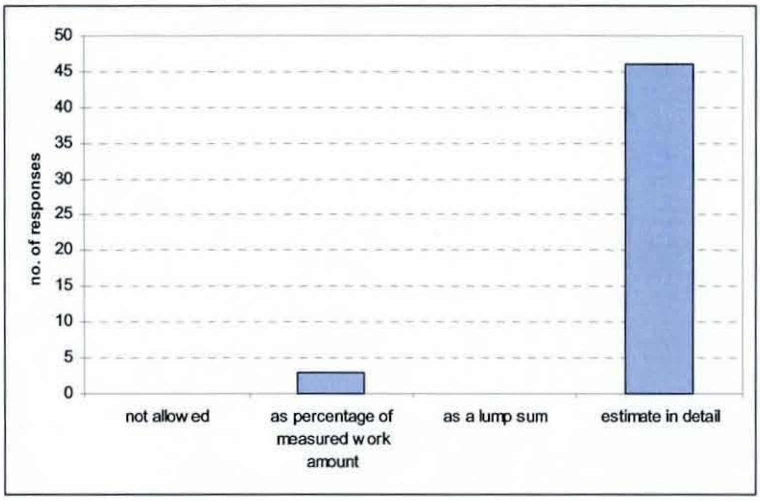

Figure 7-1 Methods of Estimating Project Overheads (from Stage 1 Survey)

7.2.1.2.2 Methods to determine the allowance for individual project overhead items

From the forty-six respondents who estimated the project overheads in detail, majority of them (65%) used their experience and professional judgment to arrive at the individual item allowance. Around 35% estimated the item amount based on cost data. None used sophisticated estimation models. The results match well with the practice in western countries. Estimators' judgment was the dominating method in estimating functions (Akintoye and Fitzgerald, 2000).

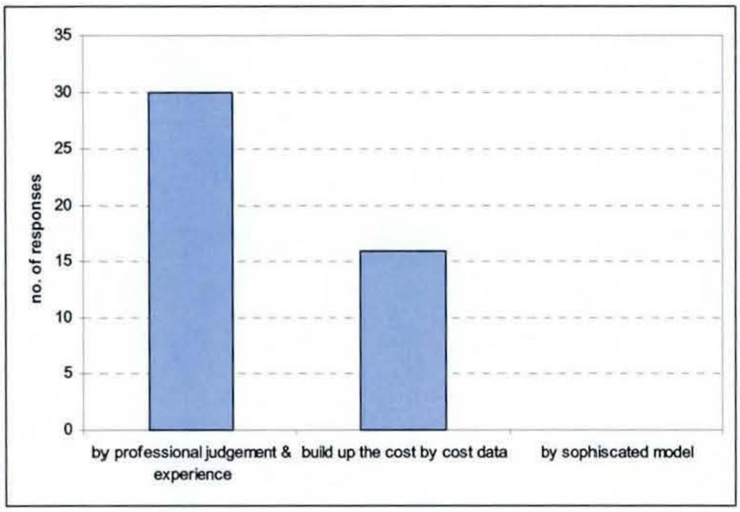

Figure 7-2 Methods to Determine Project Overhead Allowance (from Stage 1 Survey)

7.2.1.2.3 Source of cost data

The sixteen estimators used cost data either referred to the company in-house database (19%) or to the quotations from sub-contractors / suppliers (31%) or referred to both (50%). No estimators used the schedule of rates published by the Hong Kong government or PQS firms. This suggests that estimates based on cost data contained a good mix of solid knowledge and latest cost figures to provide a realistic bid. Details are shown in Figure 7-3.

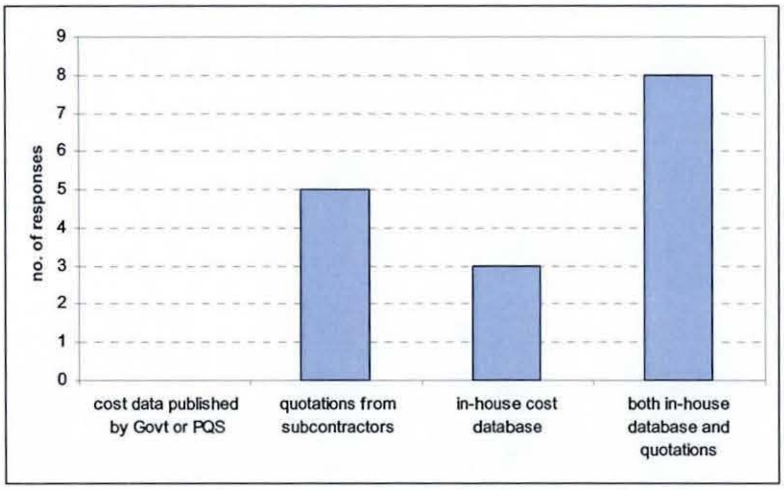

Figure 7-3 Source of Cost Data for Estimating Project Overheads (from Stage 1 Survey)

7.2.1.2.4 Availability of feedback or review of estimation accuracy

For those sixteen companies that project overhead items were estimated by detailed building up of individual items as per the CIOB Code of Estimating Practice, around 90%

of them had proper feedback / review mechanism in place to review their estimation accuracy. This reflects a good awareness of the estimation accuracy among the contractors. Figure 7-4 summarizes the responses of the contractors graphically.

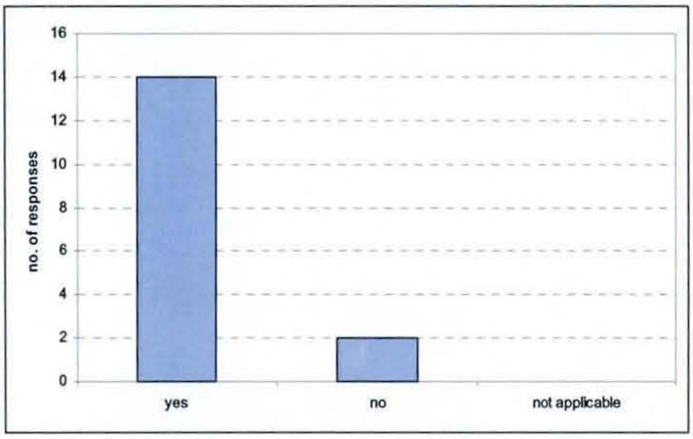

Figure 7-4 Availability of Feedback or Review System Regarding Estimation Accuracy (from Stage 1 Survey)

7.2.2 Follow-up Interview After Stage 1 Survey

7.2.2.1 Response to Follow-up Interview of Stage 1 Survey

Following the survey, twenty-two senior estimators of the construction companies who had responded to the survey were willing to join individual interviews. Although it only represented 18.5% of the sample, their views were useful to establish more in-depth knowledge regarding the current practice and problems in project overheads estimation, and the identification of important overhead items based on their respective likelihood of inaccurate estimation.

7.2.2.2 Follow-up Interview Results in Stage 1 Survey

7.2.2.2.1 Most important component in estimating

Among the various parts of the tender document : project overheads, measured work, profit margin and attendance, majority of the interviewees thought that setting profit margin was the most important (around 64%) because this was the prime interest of the company shareholders. Around 27% said that pricing measured work was the most important. Most of them who chose measured work as being the most important claimed that this was the "main body" of the tender document, and contributed the largest amount to the tender sum. Only 9% of them thought that estimating project overheads was the most important. This minority thought that project overheads were often the last resort that the contractors could squeeze for savings. Details of the views of interviewees are depicted in Figure 7-5.

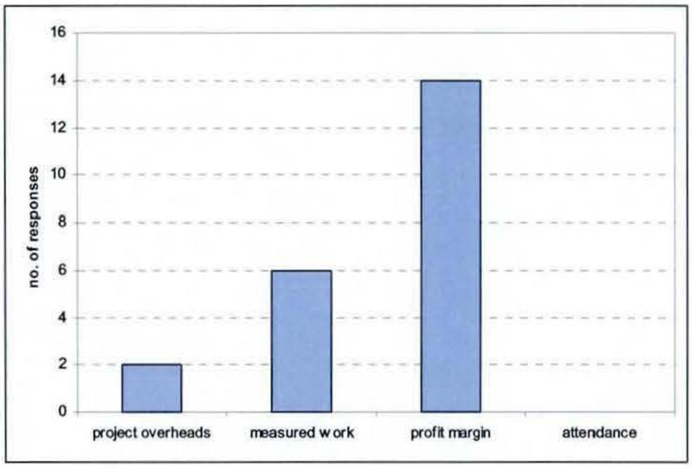

Figure 7-5 Most Important Component in Estimating

7.2.2.2.2 Staffing for estimating

All of the twenty-two interviewees claimed that both junior and senior estimators were involved in estimating. All of them agreed that junior estimators were involved in preparing estimates for the measured work, because almost all of the measured works would be sublet. Juniors were mainly involved in sending out the tender invitations, preparing tender analysis for the quotations received, and billing the measured items. Around 23% of them would ask their juniors to consolidate cost data for pricing project overheads. For the project overheads estimating and profit setting, all of the interviewees said that the former was done by the seniors whereas the latter was done by the directors of the company. This indicated that companies intuitively believed estimation of project overheads demanded a higher level of professional knowledge. This also matched with the earlier findings from this survey that estimation of project overheads relied heavily on the professional judgment of the estimators.

7.2.2.2.3 Risk transferability

This was to identify the perceived level of risk transferability of different estimates: project overheads, measured work, profit margin and attendance in case of errors or inaccuracies during estimation. Around 82% of the interviewees agreed that the risk of wrong estimation in project overheads was the most difficult to transfer. Contrasted with project overheads, no interviewees chose measured work because it would be sublet to sub-contractors, and most of the inaccuracies in estimation were transferable. A small percentage (18%) of interviewees said that the error in profit setting could not be transferred elsewhere.

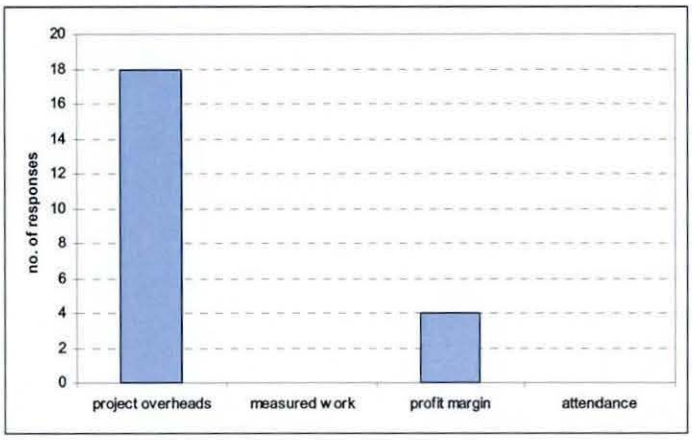

Figure 7-6 Item with Lowest Level of Risk Transferability if Inaccurately Estimated

7.2.2.2.4 General accuracy of project overheads estimation

Around 45% of the interviewees commented that over- or under-estimation of project overheads was common in their estimates. On top of that, 32% of the interviewees felt that the inaccuracies in project overheads estimation were acceptable because there were a lot of uncertainties in the project which were difficult to forecast. Only 23% of interviewees felt that the project overheads estimate was accurate.

7.2.2.2.5 Project overhead items most likely to be over- or under-estimated

There were diversified views collected regarding which project overhead items were the most likely to be over- or under-estimated. The interviewees were invited to quote at most 3 project overhead items which they thought were more likely to be inaccurately estimated. The findings are summarised in Table 7-1:

Table 7-1 Views on the Project Overhead Items Most Likely to be Estimated Inaccurately

Items	Site management salaries and expenses	Protection to works	Cleaning & removal of rubbish	Site offices, temporary stores Contract conditions	Testing	Others@
No. of responses	16	14	13	7	7	6
No. of responses ÷ total no. of items*	25.4%	22.2%	20.6%	11.1%	11.1%	9.5%
No. of responses ÷ no. of interviewees#	72.7%	63.6%	59.1%	31.8%	31.8%	27.3%

Note : *N= 63
#n=22
@"Others" covers miscellaneous items including drawings prepared by contractors, insurance, noise and dust control.

As shown in the above table, majority of the interviewees said that site management salaries and expenses were likely to be inaccurately estimated. The other two items: protection to works and general cleaning were also said to have a higher likelihood of inaccurate estimation.

7.2.2.2.6 List of Significant Project Overhead Items

Around 86% of the interviewees agreed that the Preliminaries section of the Hong Kong Standard Method of Measurement of Building Works (1979) was a useful basis from which to build the framework of project overhead items (full list of the HKSMM Preliminaries items were shown in Table 2-1 of the Literature Review Chapter).

Table 7-2 Consolidated Comments from the Interviewees on the List of Project Overheads for Pricing / Study

	Consolidated comments from interviewees	% of interviewees proposed
1.	Combined items 13 (Foreman), 22 (Watching) and 30 (Attendance) of the HKSMM2 for their similarity in nature	91%
2.	Combined items 5 (Form of Contract), 6 (Particulars to be inserted in Appendix of Conditions) and 20 (Conditions of payment) of the HKSMM2	77%
3.	Combined items 9 (Notices and fees) and 11 (Industrial training levy) of HKSMM2 since they were of similar nature and accounted a fixed percentage of the contract value	86%
4.	Combined items 17 (Injury to persons and damage to property), 18 (Insurances, etc.) and 21 (Surety or bond) of HKSMM2 for their similarity in nature	100%
5.	Combined items 28 (Water) and 29 (Lighting and power) of HKSMM2	77%
6.	Combined items 31 (Hoardings, etc.) and 33 (Temporary roads) of HKSMM2	68%
7.	Excluded item 7 of HKSMM2 (Working hours, rates of wages, etc.) since normal working hours were applied in large projects, except for some fitting out works (which is outside the scope of this study)	45%
8.	Excluded item 15 of HKSMM2 (Sub-letting) because this was already a general practice in HK and any cost incurred could be covered by site management and attendance	14%
9.	Excluded item 16 of HKSMM2 (Artists or tradesmen not sub-contractors) for attendance to artists would be allowed in the site management item	68%
10.	Excluded item 19 of HKSMM2 (Provisional and prime cost sums) for no direct cost implication to main contractor	32%
11.	Excluded item 24 of HKSMM2 (Treasure trove, coins, etc.) for no direct cost incurred to main contractor	23%
12.	Excluded item 32 of HKSMM2 (Works by public authorities) since attendance to the workers of the public authorities would be allowed in the site management item	68%
13.	Excluded item 34 of HKSMM2 (Drying the building) since it was not significant in H.K.	23%
14.	Excluded item 36 of HKSMM2 (Defects after completion) since rectification of defects would be covered by subcontractors on a 'back to back' basis, and the main contractor's attendance during Defects Liability Period would be allowed in site management item	36%
15.	Included "Principles of measurement of the Bills of Quantities" as it was a common item in Preliminaries Section of the Bills of quantities in H.K.	59%
16.	Included "Drawings to be prepared by the Main Contractor" as this item incurred cost to the contractor and it was a common obligation of the contractor	82%
17.	Included "Restrictions to Dust and Noise Nuisance" as this item incurred cost to the contractor and it was a common obligation of the contractor	95%

There are altogether thirty-six items in the HKSMM, but some of them were suggested to be streamlined or combined by the interviewees. These comments regarding the inclusion/exclusion of the project overhead items for pricing / study are listed out in Table 7-2 above. In general, the comments helped to streamline the list of project overheads and to reflect the likely cost commitment of the contractor more realistically. A final framework that consisted of the twenty-three critical project overhead items (which reflected a majority view of more than 50%) was developed and detailed in Table 7-3.

Table 7-3 List of Representative Project Overheads Items that Should Be Included for Pricing / Study

Item	Significant project overheads items to be included for pricing / study		
1	Definitions of various contractual parties	13	Protection of finished works
2	Description of works	14	Protection of adjacent / existing works
3	Nature of site and site inspection	15	Safety precautions
4	Site possession and completion	16	Facilities to be provided by Main Contractor : plants, scaffolding
5	Site management, watchman and attendance to NSCs	17	Site offices, stores, latrines
6	Principles of measurement of the Bills of Quantities	18	Setting out
7	Drawings to be prepared by the Main Contractor	19	Samples, mock ups
8	Contract conditions and amendments	20	Testing
9	Methods of measuring and valuing variations	21	Power and water supply
10	Fees and Levies	22	Temporary works e.g. hoardings, temporary roads, signboard
11	Restrictions to noise and dust nuisance	23	Cleaning and removal of rubbish
12	Insurances		

The established list of project overhead items can be used as a basis to study the extent of estimation inaccuracy of each project overhead item.

7.2.3 Stage 2 Survey – Explore the Likelihood of Inaccurate Estimation

In Stage 1 survey, the general practice of estimating project overheads was identified – majority estimated project overheads in detailed based on their professional judgment. More importantly, almost half of the respondents felt that the project overhead estimates were inaccurate although the risk of inaccurate estimation was the most difficult to transfer.

With the critical project overheads items identified at the end of the Stage 1 survey, more in-depth investigation of the likelihood of inaccurate estimation of project overheads could be conducted in the Stage 2 survey.

7.2.3.1 Response to Stage 2 Survey

According to the List of *Approved Contractors and Suppliers for Public Works* maintained by the Works Bureau of the Hong Kong Special Administrative Region Government for Public Works Approved List as at March 2002, there were one hundred and nineteen Group C contractors. They are the large, local and international contractors in Hong Kong. There were forty responses received, representing a response rate of 33.6%, which was considered reasonable and acceptable (a minimum of thirty-two usable responses, and response rate of 30% from a minimum sample size of one hundred and seven is recommended by Fellows and Liu, 2003).

7.2.3.2 Response Demographics

Among the forty respondents, twenty of them were building contractors, five were civil engineering contractors and fifteen of them engaged in both types of construction work. The majority of them (around 85%) were engaged in the construction of new works.

7.2.3.3 Survey Results

7.2.3.3.1 Method of Estimating Project Overheads

As shown in Figure 7-7, thirty-seven respondents (92.5%) of the sample stated that the project overheads were estimated in detail with close reference to the contract conditions and requirements. In other words, these contractors estimated the likely resources required for each project overhead item based on the tender drawings, specifications and contract conditions, and then built up the likely cost to be incurred. Only 3 contractors (equivalent to 7.5%) estimated the project overheads as a percentage of the total value of measured work. The results in Stage 2 survey was very similar to those in Stage 1 survey, indicating that large contractors in Hong Kong are cautious and detail in working out estimates for project overheads.

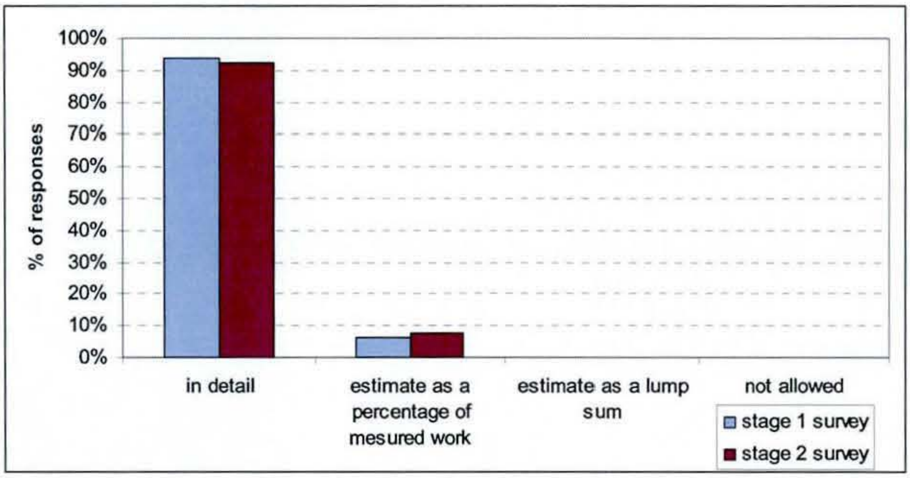

Figure 7-7 Comparison of Stage 1 and Stage 2 Survey Results on Methods of Estimating Project Overheads

This illustrates that a very high proportion of contractors spent their time and effort preparing the estimates for the project overheads. The result varies from previous studies where a much larger percentage of contractors (29%) estimated the project overheads as a percentage of the measured works or even a lump sum (Assaf et al, 1999).

7.2.3.3.2 Accuracy of Project Overheads Estimation

The next question concerned the general accuracy of project indirect costs estimation. As depicted in Figure 7-8, 60% of respondents said that the estimation was about the same as the actual expenditure. 30% claimed that the project overheads were generally under-estimated and only 10% felt that the items were generally over-estimated.

Compared with the results in Stage 1 survey, more responses felt that their project overheads were accurate in Stage 2 survey (60%) than in Stage 1 survey (23% only). Nevertheless, the possibility of bias (to understate the level of inaccurate estimation) cannot be eliminated entirely in both surveys and the actual extent of inaccurate estimation may be even higher.

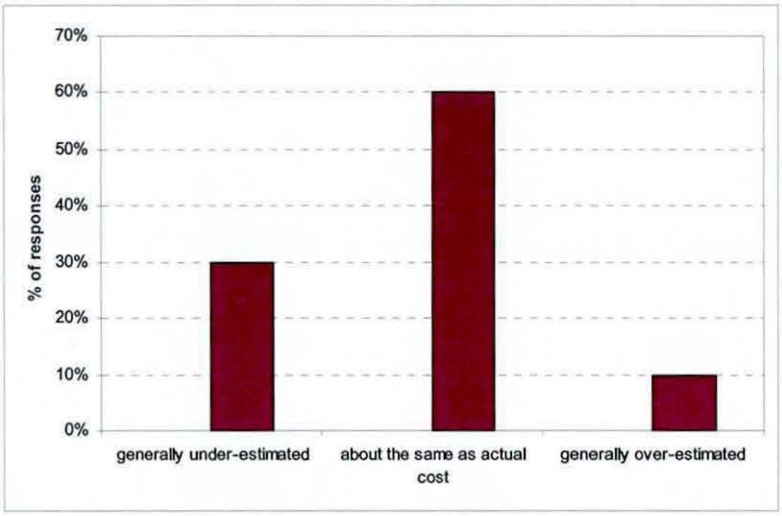

Figure 7-8 Accuracy of Project Overheads Estimation (from Stage 2 Survey)

For those who commented that there was under-estimation in project overheads, only 16.7% said that the shortfall could be recovered by prolongation claims. 50% said that such shortfall could be partly recovered by the prolongation claims and the remaining 33.3% commented that the shortfall could not be recovered.

As perceived by the respondents, under-estimation was more likely than over-estimation when estimating project overheads due to competitive tendering. Whilst the majority of these respondents thought that the shortfall (as a result of under-estimation) could not be recovered fully, the extent of potential loss to the company could be considerable.

7.2.3.3.3 Overall Estimation Accuracy of Project Overhead Items

The next part of the questionnaire collected views from estimators on the extent of under / over-estimation of the project overhead items according to their experience with their companies. They were asked to assign a weight to each project overhead item from a score range of 1 (=over-estimated) to 5 (=under-estimate). A score of "3" would mean that the item was accurately estimated during the tender stage. The mean scores, standard deviations and the percentages of response of over- / under-estimation for all the project overhead items are tabulated in Table 7-4.

Table 7-4 Summary of Descriptive Statistics on Estimation Accuracies of the Project Overhead Items

	Project overhead items	Mean	S.D.	% of respondents		
				Over-estimated (score ≤ 2)	Under-estimated (score ≥ 4)	Inaccurately estimated (score ≠ 3)
1	Definitions of various contractual parties	3.025	0.58	7.5%	10%	17.5%
2	Description of works	3.075	0.42	5%	12.5%	17.5%
3	Nature of site and site inspection	3.175	0.68	7.5%	25%	32.5%
4	Site possession and completion	3.125	0.61	10%	20%	30%
5	Site management, watchman and attendance to NSCs	3.400	0.78	7.5%	45%	52.5%
6	Principles of measurement of the Bills of Quantities	3.125	0.56	10%	22.5%	32.5%
7	Drawings to be prepared by the Main Contractor	3.325	0.80	12.5%	37.5%	50%
8	Contract conditions and amendments	3.250	0.63	2.5%	27.5%	30%
9	Methods of measuring and valuing variations	3.050	0.60	10%	17.5%	27.5%
10	Fees and levies	3.100	0.50	8%	18%	25%
11	Restrictions to noise and dust nuisance	3.525	0.75	5%	47.5%	52.5%
12	Insurances	3.025	0.66	17.5%	17.5%	35%
13	Protection of finished works	3.525	0.85	12.5%	55%	67.5%
14	Protection of adjacent / existing works	3.450	0.75	7.5%	45%	52.5%
15	Safety precautions	3.375	0.63	5%	40%	45%
16	Facilities to be provided by Main Contractor : plants, scaffolding	3.225	0.77	15%	32.5%	47.5%
17	Site offices, stores, latrines	3.025	0.70	17.5%	22.5%	40%
18	Setting out	3.125	0.46	5%	17.5%	22.5%
19	Samples, mock ups	3.300	0.82	10%	35%	45%
20	Testing	3.000	0.64	20%	20%	40%
21	Power and water supply	3.150	0.66	12.5%	25%	37.5%
22	Temporary works e.g. hoardings, temporary roads, signboard	3.325	0.66	7.5%	37.5%	45%
23	Cleaning and removal of rubbish	3.425	0.75	12.5%	52.5%	65%

Note : N = 40 ; score range from 1 to 5

It can be seen that no item was found to be accurately estimated by all respondents, despite the fact that detailed estimation was the most common method of estimating undertaken. In general, the mean scores of <u>all</u> the items were above 3 (except for testing), indicating a general tendency of under-estimation.

7.2.3.3.4 Over-estimated Project Overhead Items

As discussed in the Design of Exploratory Study of Project Overheads Chapter, Relative Accuracy Index of over-estimation (RAI_o) was adapted from the formula used by Kumaraswamy and Chan (1998) to calculate relative importance index.

The RAI_o and the relative rankings of over-estimation likelihood for the project overhead items are listed in Table 7-5.

Table 7-5 Summary of Rankings of Project Overhead Items (Ranked by the Likelihood of Over-estimation)

Rank	Project overhead items	RAI$_o$	% of respondents scoring ≤ 2
1	*Protection of finished works*	*0.9074*	*12.5%*
2	*Cleaning and removal of rubbish*	*0.9123*	*12.5%*
3	Site offices, stores, latrines	0.9140	17.5%
4	Testing	0.9167	20%
5	*Facilities to be provided by Main Contractor: plant, scaffolding*	*0.9259*	*15%*
6	*Drawings to be prepared by the Main Contractor*	*0.9275*	*12.5%*
7	Insurances	0.9293	17.5%
8	*Samples, mock ups*	*0.9359*	*10%*
9	*Site management, watchman and attendance to NSCs*	*0.9394*	*7.5%*
10	Power and water supply	0.9444	12.5%
11	Methods of measuring and valuing variations	0.9444	10%
12	Principles of measurement of Bills of Quantities	0.9506	10%
13	Nature of site and site inspection	0.9506	7.5%
14	Site possession and completion	0.9540	10%
15	Protection of adjacent / existing works	0.9545	7.5%
16	Definitions of various contractual parties	0.9596	7.5%
17	Temporary works e.g. hoardings, temporary roads, signboard	0.9600	7.5%
18	Restrictions to noise and dust nuisance	0.9667	5%
19	Fees and levies	0.9700	7.5%
20	Safety precautions	0.9722	5%
21	Contract conditions and amendments	0.9733	2.5%
22	Description of works	0.9785	5%
23	Setting out	0.9798	5%

Note : the items that are shown in italics are found likewise as the top 10's of the RAI$_o$, RAI$_u$ and RAI$_l$ lists

Although the RAI$_o$ listed in Table 7-5 represented the likelihood of over-estimation, the percentages of response were low (the highest percentage was only 20%). This indicates that over-estimation is not a common problem among the estimation of these project overhead cost items. Among the twenty-three items, "protection of finished works" was the most likely to be over-estimated. Although the general requirement of protection is typically laid down in the contract documents, the actual expenditure on protection to finished works depends on the co-operation of the sub-contractors' workers, the housekeeping of the site, the level of congestion of the site area, the time constraints of the activities, etc. Hence, it is easily over-estimated as the future project conditions are not predictable at the tender stage. The second item that was most likely to be over-estimated was "cleaning and removal of rubbish". Although cleaning and removal of rubbish is a common obligation of the main contractor, the amount of rubbish produced often depends on the wastage level on site, the method of work (e.g. whether prefabrication or in-situ methods are used), and the method of packaging the materials. In addition, the co-operation of sub-contractors will affect the tidiness of the site, which in turn has a direct impact on the amount and location of spoil/rubbish for disposal. Therefore, it is easy to over-estimate this item if the estimator takes a prudent view. Site accommodation was the third item that was most likely to be over-estimated. This is also understandable because even though the standard of site accommodation is prescribed by the contract, the final standard is still subject to the discretion of the client's representatives. Contractors are usually reluctant to argue with the client's representatives on the acceptable level of site accommodation in order to maintain a good working relationship. Therefore, it is sensible to allow more contingency for this item, leading to a higher likelihood of over-estimation.

7.2.3.3.5 Under-Estimated Project Overhead Items

Similar to the level of over-estimation, Relative Accuracy Index for under-estimation (RAI_u) was used to evaluate the relative level of under-estimation amongst different project overhead items.

The respective RAI_u of each project overhead items are listed in Table 7-6.

Table 7-6 Summary of Rankings of Project Overhead Items (Ranked by the Likelihood of Under-estimation)

Rank	Project overhead items	RAI_u	% of respondents scoring ≥ 4
1	*Protection of finished works*	*0.7524*	*55%*
2	*Cleaning and removal of rubbish*	*0.7905*	*52.5%*
3	Restrictions to noise and dust nuisance	0.7928	47.5%
4	Protection of adjacent / existing works	0.8108	45%
5	*Drawings to be prepared by the Main Contractor*	*0.8182*	*37.5%*
6	*Site management, watchman and attendance to NSCs*	*0.8198*	*45%*
7	*Samples, mock ups*	*0.8426*	*35%*
8	Safety precautions	0.8509	40%
9	*Facilities to be provided by Main Contractor: plant, scaffolding*	*0.8529*	*32.5%*
10	Temporary works e.g. hoardings, temporary roads, signboard	0.8559	37.5%
11	Contract conditions and amendments	0.8857	27.5%
12	Nature of Site and site inspection	0.8922	25%
13	Power and water supply	0.8952	25%
14	Principles of measurement of the Bills of Quantities	0.9063	22.5%
15	Site offices, stores, latrines	0.9091	22.5%
16	Site possession and completion	0.9091	20%
17	Testing	0.9167	20%
18	Insurances	0.9192	17.5%
19	Methods of measuring and valuing variations	0.9293	17.5%
20	Fees and levies	0.9369	17.5%
21	Setting out	0.9386	17.5%
22	Description of works	0.9510	12.5%
23	Definitions of various contractual parties	0.9510	10%

Note : the items that are shown in italics are found likewise as the top 10's of the RAI_o, RAI_u and RAI_t lists

From Table 7-5 and Table 7-6, a number of items were identified as possessing a higher risk of both over- and under-estimation. For example, the top two items in both tables are the same. This reflects the difficulty and uncertainty encountered when estimating these items, giving rise to diverse decisions when determining the estimates. From the list of RAI_u in Table 7-6, most of the estimators thought that they often under-estimated protection of finished works followed by cleaning and removal of rubbish. Both were claimed to be under-estimated by more than half of the respondents. The main reasons are that the expenditure on these two items is reliant on the future site conditions, planning of work and sub-contractors' performance; but these are difficult to estimate. As a result, estimators would tend to assume the sub-contractors to share the liability of protection to finished works and cleaning. However, the extent of protection and cleaning provided by the sub-contractors is often minimal and insufficient to meet the requirements of the complicated site environment. In addition, damage caused by inclement weather (e.g. flooding, collapse of scaffolding) is common during the summer time, but the financial impact of these accidents may not be estimated during the tender stage. Therefore, these items are often under-estimated, as confirmed by the survey results.

It was noted that the absolute values of RAI_u in Table 7-6 were much lower than that of the RAI_o in Table 7-5, indicating a greater degree of under-estimation than over-estimation. This matches well with the findings from earlier questions which asked the respondents about the general accuracy of their project overhead estimates. As shown in Table 7-5 and Table 7-6, the percentages of respondents scoring greater than 4 (i.e. under-estimation) was much larger than the percentages for over-estimation. This reflects the intuitive tendering decisions of estimators. If the contractors want to win the bid, they will tend to allow very little contingency in their estimates. Hence, under-estimation is more likely than over-estimation when the items involve uncertainties under a competitive bidding environment.

7.2.3.3.6 Inaccurately Estimated Project Overhead Items

Here, the RAI_i was used to illustrate the relative inaccuracy of project overhead estimation. The RAI_i and the relative rankings were calculated and tabulated in Table 7-7.

Table 7-7 Summary of Rankings of Project Overhead Items (Ranked by the Likelihood of Inaccurate Estimation)

Rank	Project overhead items	RAI_l	% of respondents scoring ≠ 3
1	*Protection of finished works*	*0.7417*	*67.5%*
2	*Cleaning and removal of rubbish*	*0.7750*	*65%*
3	Restrictions to noise and dust nuisance	0.7863	52.5%
4	*Drawings to be prepared by the Main Contractor*	*0.7982*	*50%*
5	Protection of adjacent / existing works	0.8000	52.5%
6	*Site management, watchman and attendance to NSCs*	*0.8000*	*52.5%*
7	*Samples, mock ups*	*0.8167*	*45%*
8	*Facilities to be provided by Main Contractor : plants, scaffolding*	*0.8250*	*47.5%*
9	Safety precautions	0.8417	45%
10	Temporary works e.g. hoardings, temporary roads, signboard	0.8417	45%
11	Site offices, stores, latrines	0.8583	40%
12	Nature of Site and site inspection	0.8649	32.5%
13	Testing	0.8667	40%
14	Power and water supply	0.8667	37.5%
15	Contract conditions and amendments	0.8704	30%
16	Insurances	0.8750	35%
17	Principles of measurement of the Bills of Quantities	0.8796	32.5%
18	Site possession and completion	0.8829	30%
19	Methods of measuring and valuing variations	0.8919	27.5%
20	Fees and levies	0.9167	25%
21	Definitions of various contractual parties	0.9189	17.5%
22	Setting out	0.9250	22.5%
23	Description of works	0.9352	17.5%

Note : the items that are shown in italics are found likewise as the top 10's of the RAI_o, RAI_u and RAI_l lists

In fact, the RAI_l can be regarded as the combined index for both RAI $_o$ and RAI $_u$. As seen from the formula for RAI_l, all responses were included in the calculation. Therefore, it was logical that Protection of finished works; and Cleaning and removal of rubbish were the most likely to be estimated inaccurately. Besides, it was justified that the majority of the top 10's in the RAI_l list appeared as the top 10's in Table 7-5 and Table 7-6 as well (as highlighted in italics). All the top ten items in Table 7-7 were likely to be inaccurately estimated by more than 45% of respondents. In other words, these items were very likely to be over-estimated by some estimators and under-estimated by some others. To explain, it simply reflects the dilemma faced by the estimators. These items possess entangled uncertainties that are difficult to forecast at the outset of the tender stage. As a result, the estimators may either over-estimate them (if they do not want to lose money in case of contract award) or under-estimate them (if they do not want to lose the opportunity to win the contract). Depending on the tendering strategy of the company and the pricing attitude of the estimator, project overhead items that possess so much uncertainty cannot be estimated accurately.

7.2.3.3.7 Cross-comparison of Relative Rankings Among Different Nature of Companies

This part cross-compares the relative inaccuracies of project overhead costs estimation among different types of companies. As mentioned before, the responses of this survey came from three types of companies: building contractors, civil engineering contractors, and contractors that engaged in both building and civil engineering works. Table 7-8 shows the RAI_o of the twenty-three project overhead items from the three types of companies.

Table 7-8 Summary of RAI_o and Relative Rankings of Project Overhead Items Across 3 Types of Companies

	Project overhead items	RAI_o (Relative Rank in bracket)					
		Building contractors		Civil Engineering Contractors		Contractors engaged in both types of works	
1	Definitions of various contractual parties	0.9215	(8)	1.0000	(22.5)	1.0000	(20)
2	Description of works	0.9555	(19)	1.0000	(18.5)	1.0000	(22.5)
3	Nature of site and site inspection	0.9285	(10)	1.0000	(9)	0.9697	(15)
4	Site possession and completion	0.9048	(4)	1.0000	(18.5)	1.0000	(20)
5	Site management, watchman and attendance to NSCs	0.9286	(9)	1.0000	(12.5)	0.9333	(8)
6	Principles of measurement of the Bills of Quantities	0.9487	(17)	0.9167	(5.5)	0.9667	(13.5)
7	Drawings to be prepared by the Main Contractor	0.9394	(14)	0.8889	(2.5)	0.9259	(7)
8	Contract conditions and amendments	0.9394	(13)	1.0000	(12.5)	1.0000	(20)
9	Methods of measuring and valuing variations	0.9048	(6)	1.0000	(18.5)	0.9722	(16)
10	Fees and levies	0.9375	(12)	1.0000	(22.5)	1.0000	(22.5)
11	Restrictions to noise and dust nuisance	0.9667	(21)	1.0000	(12.5)	0.9524	(11)
12	Insurances	0.9412	(15)	0.9167	(5.5)	0.9167	(2.5)
13	Protection of finished works	0.8889	(3)	1.0000	(12.5)	0.8889	(1)
14	Protection of adjacent / existing works	0.9444	(16)	1.0000	(12.5)	0.9524	(11)
15	Safety precautions	0.9762	(22)	0.8889	(2.5)	1.0000	(18)
16	Facilities to be provided by Main Contractor : plants, scaffolding	0.9333	(11)	0.9167	(5.5)	0.9167	(4)
17	*Site offices, stores, latrines*	*0.9048*	*(5)*	*0.9333*	*(8)*	*0.9167*	*(2.5)*
18	Setting out	1.0000	(23)	0.9167	(5.5)	0.9762	(17)
19	Samples, mock ups	0.9167	(7)	1.0000	(18.5)	0.9444	(9)
20	Testing	0.8889	(2)	1.0000	(18.5)	0.9231	(6)
21	Power and water supply	0.9608	(20)	0.7778	(1)	0.9667	(13.5)
22	Temporary works e.g. hoardings, temporary roads, signboard	0.9556	(18)	1.0000	(12.5)	0.9524	(11)
23	Cleaning and removal of rubbish	0.8788	(1)	1.0000	(18.5)	0.9167	(5)

Note : the item that is shown in italics is found likewise as the top 10's of the RAI_o lists from the three types of companies

Table 7-9 and Table 7-10 depict the RAI_u and RAI_l for the three types of companies respectively.

Table 7-9 Summary of RAI_u and relative Rankings of Project Overhead Items Across 3 Types of Companies

	Project overhead items	RAI*u* (Relative Ranking in bracket)		
		Building contractors	Civil Engineering Contractors	Contractors engaged in both types of works
1	Definitions of various contractual parties	0.9778 (23)	1.0000 (22.5)	0.9048 (13.5)
2	Description of works	0.9375 (21)	0.9333 (17.5)	0.9744 (22)
3	Nature of site and site inspection	0.9167 (17)	0.8000 (1)	0.8974 (12)
4	Site possession and completion	0.9048 (14)	0.9333 (17.5)	0.9048 (13.5)
5	*Site management, watchman and attendance to NSCs*	*0.8704 (7)*	*0.8667 (5.5)*	*0.7381 (4)*
6	Principles of measurement of the Bills of Quantities	0.8958 (11)	0.9167 (10.5)	0.9167 (16)
7	Drawings to be prepared by the Main Contractor	0.8125 (3)	0.8333 (2.5)	0.8205 (10)
8	Contract conditions and amendments	0.8431 (6)	0.9167 (10.5)	0.9286 (18)
9	Methods of measuring and valuing variations	0.9111 (16)	0.9333 (17.5)	0.9487 (21)
10	Fees and levies	0.9216 (19)	1.0000 (22.5)	0.9333 (19)
11	Restrictions to noise and dust nuisance	0.7719 (2)	0.9167 (10.5)	0.7857 (6.5)
12	Insurances	0.9216 (20)	0.9167 (10.5)	0.9167 (16)
13	*Protection of finished works*	*0.7451 (1)*	*0.8667 (5.5)*	*0.7179 (2)*
14	Protection of adjacent / existing works	0.8333 (5)	0.9167 (10.5)	0.7857 (6.5)
15	Safety precautions	0.8947 (10)	0.8333 (2.5)	0.8000 (8)
16	Facilities to be provided by Main Contractor : plants, scaffolding	0.9020 (13)	0.9167 (10.5)	0.7692 (5)
17	Site offices, stores, latrines	0.8824 (8)	1.0000 (21)	0.9167 (16)
18	Setting out	0.9167 (18)	0.9167 (10.5)	0.9762 (23)
19	Samples, mock ups	0.9020 (12)	0.9333 (17.5)	0.7381 (3)
20	Testing	0.8889 (9)	0.9333 (17.5)	0.9444 (20)
21	Power and water supply	0.9444 (22)	0.8333 (4)	0.8810 (11)
22	Temporary works e.g. hoardings, temporary roads, signboard	0.9074 (15)	0.9167 (10.5)	0.8095 (9)
23	Cleaning and removal of rubbish	0.8125 (4)	0.9333 (17.5)	0.7143 (1)

Note : the items that are shown in italics are found likewise as the top 10's of the RAI_u lists from the three types of companies

From Table 7-8, Table 7-9 and Table 7-10, the relative rankings of items in terms of their likelihood of over-; under- and inaccurate estimation are very different among the three types of companies. If focused on the top ten project overhead items in each table, less than four items are commonly found in all three types of companies. It is apparent that the rank agreement of the relative rankings of over-estimation, under-estimation and inaccurate estimation is fairly weak across the different types of companies.

Table 7-10 Summary of RAI_i and Relative Rankings of Project Overhead Items Across 3 Types of Companies

	Project overhead items	RAI_i (Relative Ranking in bracket)		
		Building contractors	**Civil Engineering Contractors**	**Contractors engaged in both types of works**
1	Definitions of various contractual parties	0.9074 (20)	1.0000 (22.5)	0.9048 (17.5)
2	Description of works	0.9074 (21)	0.9333 (18)	0.9744 (23)
3	Nature of site and site inspection	0.8704 (14)	0.8000 (3.5)	0.8810 (14)
4	Site possession and completion	0.8519 (12)	0.9333 (18)	0.9048 (17.5)
5	*Site management, watchman and attendance to NSCs*	*0.8333 (7)*	*0.8667 (7.5)*	*0.7333 (4)*
6	Principles of measurement of the Bills of Quantities	0.8704 (15)	0.8667 (7.5)	0.8974 (16)
7	Drawings to be prepared by the Main Contractor	0.7963 (4)	0.8000 (3.5)	0.8000 (9.5)
8	Contract conditions and amendments	0.8148 (5)	0.9167 (12.5)	0.9286 (19.5)
9	Methods of measuring and valuing variations	0.8519 (11)	0.9333 (18)	0.9286 (19.5)
10	Fees and levies	0.8833 (18)	1.0000 (22.5)	0.9333 (21)
11	Restrictions to noise and dust nuisance	0.7667 (2)	0.9167 (12.5)	0.7778 (6.5)
12	Insurances	0.8833 (16)	0.8667 (7.5)	0.8667 (12.5)
13	*Protection of finished works*	*0.7333 (1)*	*0.8667 (7.5)*	*0.7111 (1)*
14	Protection of adjacent / existing works	0.8167 (6)	0.9167 (12.5)	0.7778 (6.5)
15	Safety precautions	0.8833 (18)	0.8000 (2)	0.8000 (9.5)
16	Facilities to be provided by Main Contractor : plants, scaffolding	0.8667 (13)	0.8667 (7.5)	0.7556 (5)
17	Site offices, stores, latrines	0.8333 (8)	0.9333 (18)	0.8667 (12.5)
18	Setting out	0.9167 (22.5)	0.8667 (7.5)	0.9556 (22)
19	Samples, mock ups	0.8500 (10)	0.9333 (18)	0.7333 (3)
20	Testing	0.8333 (9)	0.9333 (18)	0.8889 (15)
21	Power and water supply	0.9167 (22.5)	0.7500 (1)	0.8667 (11)
22	Temporary works e.g. hoardings, temporary roads, signboard	0.8833 (18)	0.9167 (12.5)	0.8000 (8)
23	Cleaning and removal of rubbish	0.7833 (3)	0.9333 (18)	0.7111 (2)

Note : the items that are shown in italics are found likewise as the top 10's of the RAI_i lists from the three types of companies

To affirm the weak rank agreement of the relative inaccuracy rankings among the three types of companies, a Rank Agreement Factor (which is explained in the Design of Exploratory Study of Project Overheads Chapter) was computed.

In this survey, the RA_{max} is 11.4783 (taking into account the total number of project overhead items being twenty-three). Details of the determination of RA_{max} are shown in Appendix F. As a result, the percentage agreement / disagreement amongst the three types of companies were calculated as shown in Table 7-11.

Table 7-11 Cross Comparison of Relative Inaccuracies as Perceived by Different Types of Companies

Groups Under Comparison for the Relative Inaccuracies	Rank Agreement Factor (RAF)	Percentage Disagreement (PD)	Percentage Agreement (PA)
Over-estimation			
Building vs Civil	227	85.98%	14.02%
Building vs Both	148	56.06%	43.94%
Civil vs Both	142	53.79%	46.21%
Under-estimation			
Building vs Civil	144	54.55%	45.45%
Building vs Both	122	46.21%	53.79%
Civil vs Both	145	54.92%	45.08%
Overall inaccurate estimation			
Building vs Civil	187	70.83%	29.17%
Building vs Both	111	42.05%	57.95%
Civil vs Both	151	57.20%	42.80%

Table 7-11 summarizes the cross-comparison of the perceived relative inaccuracy rankings of the project overhead costs estimation among the three types of companies. Most of the PAs were only around 50%, with two of them even below 30%. The percentage agreements were fairly low and it was therefore unable to conclude that the relative ranks of estimation inaccuracies for the project overhead items (which indicated the perceived level of estimation inaccuracies) were agreed among the three types of companies. This is due to the differing requirements of building and civil engineering contracts. For building contracts, the number and variety of trades is much greater than civil engineering contracts. Site restrictions and requirements in building contracts are also very different from civil engineering contracts. As a result, less samples and cleaning / removal of rubbish are required in civil engineering works. It can be seen from Table 7-8, Table 7-9 and Table 7-10 that although these items were ranked relatively high by the building contractors, the opposite was the case for civil engineering contractors. Likewise the Power and water supply item that was ranked at the bottom of the list by the building contractors was ranked as the first among the civil engineering contractors (in the RAI_i). This is because the major trades in the civil engineering projects - excavation and concreting; consume substantial amount of power and water. However, due to the large site area involved in typical civil engineering projects, consumption of water and power is difficult to control and hence more difficult to estimate. This brings out an important consideration when collecting project data for the further analysis of project overhead cost. The same type of project (building or civil engineering) should be collected to eliminate differences in relative accuracy and financial impact.

Despite the low rank agreement among different types of companies, some project overhead cost items were identified as having high likelihood of under-, over- and inaccurate estimation by all three types of companies (as marked in italics in Table 7-8, Table 7-9 and Table 7-10). For example, Site offices, stores and latrines was claimed to have high a likelihood of over-estimation by all types of companies. The reason is that the requirements of site accommodation are irrespective of the nature of the project. As mentioned earlier, the final acceptance of site accommodation is based on the discretion of the client's representatives. Hence, this item is often over-estimated by the contractors to ensure a good working relationship with the consultants is maintained. On the other hand, Site management and Protection to finished works were perceived by all types of contractors as more likely to be under-estimated. As commented by the respondents, site management staff sundry expenses (e.g. entertainment fee and traveling allowance) were often included in the Site management item. Although most of the companies had set down limits on these expenses for staff of different grades, it was common to have some staff overspent beyond the limit while none would spend less than allowed. Hence, the site management item exhibits a high likelihood of under-estimation among the three types of companies. Besides this item, Protection of finished works was also likely to be under-estimated by all contractors because irrespective of contract type, most sub-contractors will not provide sufficient protection to their finished works. The damage caused by inclement weather is also difficult to estimate and is unlikely to be counted in the tender estimate resulting in under-estimation by all contractors. Inaccurate estimation reflects the combined effect of over- and under-estimation. Thus, Site management and Protection of finished works are found to have a high likelihood of inaccurate estimation among the three types of contractors.

7.2.4 Project Overhead Cost Data Collection

Following the Stage 1 and Stage 2 surveys, actual project overhead cost data were collected from the large contractors in Hong Kong in order to have a comprehensive understanding of the nature of project overhead costs.

7.2.4.1 Project Overhead Cost Data Collected

Invitations were sent to the Contracts Managers of these companies, requesting to have access to one or more of their projects cost data. However, due to the sensitivity nature of cost data, most of the construction companies refused to disclose their confidential project information. After follow-up phone calls and referral contacts, breakdown of project overheads expenditure and the respective project information of twenty completed building projects were successfully collected.

The project overheads cost data was studied together with the contract conditions and bills of quantities. This can enable a better interpretation of the project overhead cost by referring to the special requirement of the client or site condition.

The contract sum of the twenty projects ranged from HK$120 to 550 million. All of them were new building projects including four public housing, nine private housing and seven commercial building projects. They were all constructed during 1997 to 2001.

7.2.4.2 Analysis of Project Overhead Cost Data

7.2.4.2.1 Project overheads distribution

In general, the project overheads accounted for 11% - 19% of the total project cost, which is in line with the literature (not more than 20% of total project costs (Assaf et al. 1999, Solomon 1993)). Regarding the cost breakdown structure of the project overheads, different sampled construction companies adopted slightly different strategies for cost control purpose. For example, some contractors recorded the costs on main contractor site office and consultant site office together while some kept them as separate items. Therefore, some cost items are combined as a matter of comparison. Altogether there were twenty-one items consolidated, with seven items (site management; mechanical plants; site cleaning; insurances and surety bond; temporary works; and scaffolding) accounting for more than 80% of the total project overheads cost. On the other hand, half of the items in the list accounted for a total of 4% of the project overhead costs only. Since there were some sundry items spent by different companies for varying purposes e.g. work commencement ceremony[6], photos processing, etc., an item "miscellaneous" was created to include all these sundry costs. The details of itemized cost distribution are shown in Table 7-12.

Table 7-12 Analysis of Major Project Overheads Cost

Item	Project overhead items	Average % of project overheads
1	Site management	35.9%
2	Mechanical plants	12.4%
3	Cleaning and removal of rubbish	10.0%
4	Insurances and surety bond	7.4%
5	Setting out	6.2%
6	Temporary works e.g. hoardings, temporary roads, signboard	5.9%
7	Scaffolding	5.9%
8	Power	4.5%
9	Site offices, stores, latrines	3.2%
10	Fees and levies	2.5%
11	Watchman	2.5%
12	Water	1.0%
13	Testing	0.7%
14	Safety precautions	0.6%
15	Contract conditions and amendments	0.4%
16	Drawings to be prepared by the Main contractor	0.3%
17	Miscellaneous e.g. entertainment fees, petty cash	0.3%
18	Telephone & fax	0.2%
19	Protection	0.06%
20	Samples, mock ups	0.06%
21	Restrictions to noise and dust nuisance	negligible, less than 0.05%
		100.0%

7.2.4.2.2 Major project overheads cost items

Similar to the project overhead cost breakdown prepared by Solomon (1993), staffing cost remained as the largest cost centre among all the project overheads cost items. However, the relative percentage of this item was much greater in the current study, up to an average of around 36% (in Solomon's study only around 26%). The expenditure on site staff was in the range of 24% - 46%. This reflects the increasing level of site management expertise and skills required by the clients nowadays. The second highest cost centre was mechanical plants. For the majority of the projects analysed, the cost of mechanical plants was in the range of 9% - 13%. Although this item remained as the second costly item in the project overheads, its relative percentage (12.4%) was much smaller than that in Solomon's study (22.3%). As a result of technological advancement, cost of producing plants and the like is much reduced in the last decade. Besides, the overall inflation in staff salary is much higher

[6] It is customary in the Chinese culture to prepare ceremony to pray for important events like construction work commencement to ensure smooth operation throughout the whole process. Roasted pig, fruits, Chinese wine, etc. will be presented as offerings.

than the cost of other factors of production (e.g. materials and plants). Therefore, it is reasonable to see a drop in plant cost / percentage.

The relative rankings and percentages for the rest of the project overheads items were quite different from the findings in Solomon's study (1993). One of the reasons may be due to the difference between tender estimates (collected in Solomon's study) and actual expenditure (collected in this study) where the former figures are influenced by the company's tendering strategy. The third contributory item is cleaning and removal of rubbish. It accounted for 10% of the total project overheads, almost doubled the percentage suggested by Solomon. Nowadays, construction companies are more aware of safety and environmental issues. Good housekeeping and site cleanliness is not only a concern from the clients, but also one of the selling points of the construction companies. Therefore, it is reasonable that more money is spent in keeping the site clean and tidy.

The forth item, insurances and surety bond, contributed more than 7% of the total project overheads, a lot more than the figure shown in Solomon's study (1.7%). According to the project managers of the construction companies, the insurance expenditure will continue to climb as a result of the global increase in premiums after the series of terrorist attacks in the last few years.

Another interesting item is site offices, stores and latrines. It ranked the ninth in this study, contributed an average of 3.2% to the overall project overheads. However, in Solomon's study, this item ranked the forth as one of the major items in the list. In the tender, there is a tendency to put a higher price in site accommodation item strategically in order to have more money paid at the start of the project. Therefore, the tender costs used by Solomon for site accommodation exhibited a much higher percentage to the total project overheads.

7.2.4.2.3 Relative Percentage vs Estimation Accuracy of Project Overhead Items

Based on the Relative Accuracy Indices calculated earlier in the Stage 2 Survey results, the average percentages of project overheads were further cross-compared with the relative estimation inaccuracies. Since the cost breakdown structures of the construction companies do not match entirely with the significant project overhead items as listed in Table 7-7, the percentages of some items in Table 7-12 were combined together (e.g. site management and watchman; water and power). As a result, a total of sixteen items were consolidated for comparison and listed in Table 7-13.

Table 7-13 Summary of Average Percentage and Relative Accuracy Index of Each Project Overhead Item

Item	Project overhead items	Average % of project overheads	Relative Accuracy Index RAI_I
1	Protection of adjacent / existing / finished works	0.1%	0.7709
2	Cleaning and removal of rubbish	10.0%	0.7750
3	Restrictions to noise and dust nuisance	0.0%	0.7863
4	Drawings to be prepared by the Main Contractor	0.2%	0.7982
5	Site management, watchman and attendance to nominated sub-contractors	38.4%	0.7999
6	Samples, mock ups	0.1%	0.8167
7	Facilities to be provided by Main Contractor : plants, scaffolding	18.2%	0.8250
8	Safety precautions	0.6%	0.8417
9	Temporary works e.g. hoardings, temporary roads, signboard	5.9%	0.8417
10	Site offices, stores, latrines	3.4%	0.8583
11	Testing	0.7%	0.8667
12	Power and water supply	5.6%	0.8667
13	Contract conditions and amendments	0.4%	0.8704
14	Insurances	7.4%	0.8750
15	Fees and levies	2.5%	0.9167
16	Setting out	6.2%	0.9250

Some interesting relations were found in the above table. First, some items that the estimators described as difficult to estimate were in fact, accounting only a small amount in the project overhead costs e.g. Protection; Restrictions to noise and dust nuisance; and samples / mock ups. However, there were costly items like setting out; insurances that encountered a relatively low likelihood of inaccurate estimation. As found in earlier surveys, estimators spend the most expertise in estimating project overheads, particularly in those items with a higher likelihood of inaccuracy. From the above data, it is evidenced that they are not utilizing their capacity to the optimum level as a lot of their time is spent to minimize risk that contributes to relatively trivial costs.

A two-dimensional matrix in Figure 7-9 helps to present the relationship between relative cost amount and estimation accuracy of the sixteen project overhead items in a clearer manner. However, a lot of points were concentrated along the x-axis. This is due to the fact that a substantial amount of project overhead items were contributing to a very small percentage of the total project overhead cost. In this regard, a logarithm function was applied to the percentage values of each project overhead item. The adjusted matrix in

Figure 7-10 provides a well-scattered diagram for better differentiation and decision-making.

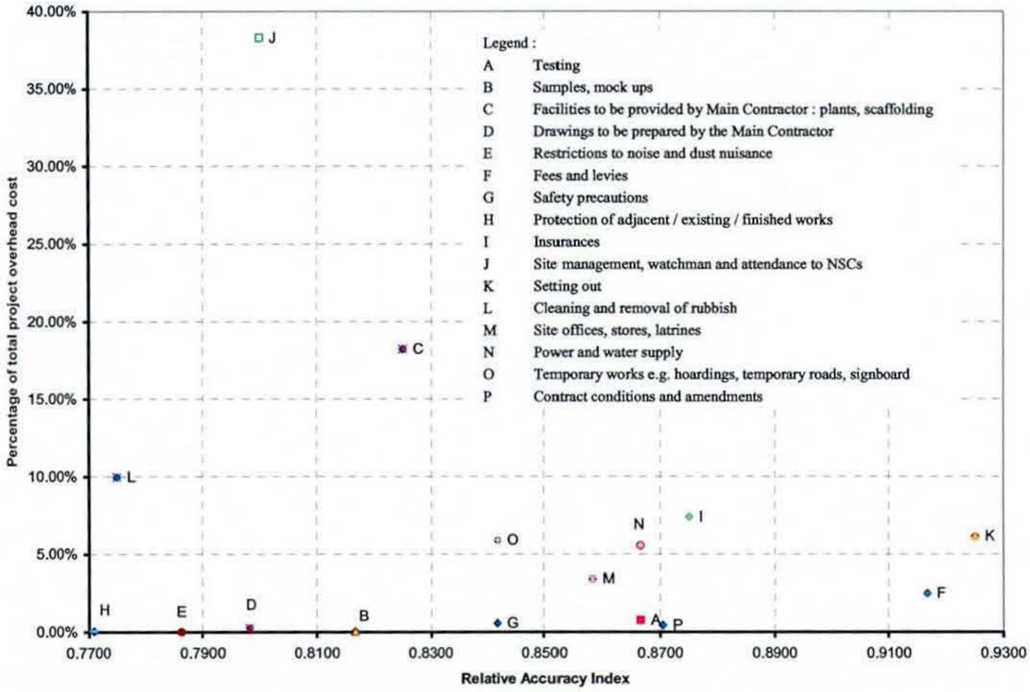

Figure 7-9 Matrix of Relative Project Overhead Percentage Against Relative Estimation Accuracy Index

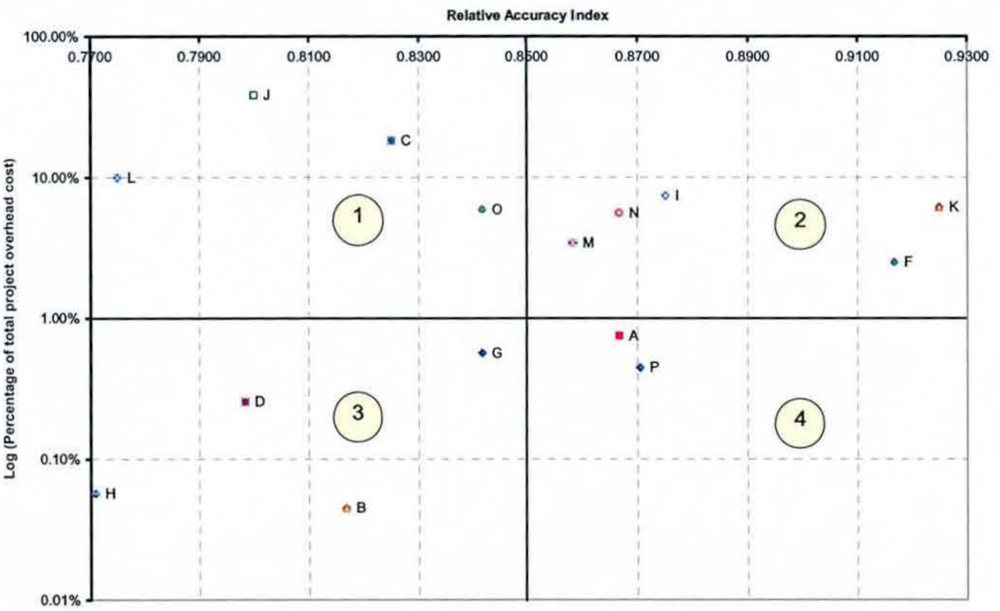

Figure 7-10 Matrix of log(percentage of total project overhead cost) Against Relative Accuracy Index

From Figure 7-10, four groups of project overhead items could be classified to reflect their potential cost and risk in estimation.

1. 1st Quarter: project overhead items that contributed relatively high to total project overhead and their estimation inaccuracy was also high. Items included :
   - facilities provided by main contractor e.g. plants, scaffolding;
   - site management, watchman and attendance to NSCs;
   - clearing, removal of rubbish; and
   - temporary works, e.g. hoarding, temporary roads, signboards.
2. 2nd Quarter: project overheads that contributed relatively high to total project overhead cost but their estimation inaccuracy was low. Items included :
   - fees and levies;
   - insurances;
   - setting out;
   - site offices, stores, latrines; and
   - power and water supply.
3. 3rd Quarter: project overheads that contributed relatively low to total project overhead cost but their estimation inaccuracy was high. Items included :
   - samples, mock ups;
   - drawings to be prepared by the main contractor;
   - safety precautions;
   - protection of adjacent / existing / finished works; and
   - restrictions to noise and dust nuisance.
4. 4th Quarter: project overheads that contributed low to total project overhead cost and their estimation inaccuracy was low.
   - testing; and
   - contract conditions amendments.

The matrix helps estimators to make appropriate decisions to allow their estimating resources amongst different category of items. Further elaboration will be made in the Analysis and Validation of ANN Model Section.

7.3 Development of Estimation Model for Project Overheads

The exploratory study successfully revealed a lot of useful information that was not found in the existing literature, e.g. the estimating practice of project overheads, potential inaccuracy of each project overhead item, and the distribution pattern of project overhead costs. The next part of this research is to develop the estimating model for project overheads. First, the results leading to the development of final list of input variables through focus group discussion and opinion survey are elaborated. Then, results of the exploratory factor analysis are examined in order to establish the final list of input variables to the estimation model. Lastly, the training, testing and validation results of the GMDH model are examined, together with the results of alternative regression equation. The respective estimating accuracies of these models are compared before concluding the best model and significant factors of measuring project overheads.

7.3.1 Develop the Comprehensive Factor List

As discussed in the Development of Inputs and Outputs to Cost Model Chapter, the preliminary list of factors affecting project overhead expenses was generated from the literature review. Since the exact literature on factors affecting project overheads was very limited, other literature related to factors affecting project cost and duration, project time and cost overruns and project risks were also examined. As a result, a preliminary list containing twenty-eight factors was generated.

7.3.1.1 Results of Focus Group Discussion on Factors Affecting Project Overhead Cost

In order to make sure that the preliminary list is comprehensive to include all possible factors, the list was given to the eight senior quantity surveyors for focus group discussion. All of the focus group members possessed more than twenty years of industrial experience, and were currently working in a Group C contractor in Hong Kong. The discussion lasted for two and a half hour, and the comments of the group were summarized in Table 7-14.

Table 7-14 Summary of Focus Group Comments Regarding Factors Affecting Project Overhead Cost

Item	Preliminary Factor Description	Recommendations from Focus Group
1	Duration of project	Agreed
2	Type of project	Type of Project (e.g. commercial, residential, industrial)
3	Scale of project	Size of Project (gross floor area of proposed building)
4	Building height	Expanded to : No. of storey of proposed structure No. of level of basement (of proposed structure)
5	Complexity of project (use of special techniques / plants / tradesman)	Complexity of project (use of special techniques / plants / technicians)
6	Nature of structure frame	Deleted, better to integrate with "Complexity of project"
7	Shape of site	Agreed
8	Shape of structure to be constructed	Shape of building
9	Accessibility of site	Agreed
10	Site coverage of proposed building	Agreed
11	Required quality level of project	Agreed
12	Type of contract	Type of contract (e.g. standard form, government form)
13	Payment schedule	Expanded to : Payment terms Punctuality of payment by client
14	Quality of design information	Completeness of project information and drawings from architect / engineer (drawings, design, specifications)
15	Change order/variations	Expanded to : Extent of variations Promptness of response from designers
16	The client's strictness in supervision	Strictness in supervision by client / client's representatives
17	Contract terms (e.g. EOT clause, amount of LD, time limit to submit claims)	Extent of bond / warranty requirement from the project
18	Ground conditions	Soil conditions
19	Extent of sub-letting	Extent of sub-letting of project works
20	Risk related to sub-contractors' performance and strength	Expanded to : Technical competence of sub-contractors Financial stability and cash flow of sub-contractors in construction stage Workload of subcontractors
21	Size of mechanical plant fleet	Extent of mechanical plants used for the project
22	Equipment available to contractor	Extent of mechanical plants owned by the contractor
23	Project leader's experience	Expanded to : Past experience of project manager in similar project Claims consciousness of project manager who will be assigned to the project Interpersonal skills of project manager who will be assigned to the project
24	Contractor's past performance or image	Deleted, "Project leader's experience" is more representative
25	Staff and labour size	No. of management staff assigned on site for the project
26	Percentage of professional qualified staff	Technical competence of site management staff
27	Safety and social risk	Deleted, not critical in Hong Kong
28	Project schedule (time constraint)	Deleted, not critical in Hong Kong

Besides the items that were agreed to keep in the list of factors affecting project overhead cost, some of the items were recommended to expand. Firstly, *building height* was suggested to split into two items, one to measure the *no. of storey of the building* and one to measure *no. of basement level.* As a lot of buildings in Hong Kong are designed with basement, building height may be misleading as it refers to the height of superstructure only. The original factor: *payment schedule* failed to reflect entirely the difficulty to receive payment from the client. Therefore, it was suggested to expand into *payment terms* and *punctuality of payment by client.* Regarding to variations, most of them incurred financial implications to the contractor. Therefore, the focus group suggested to add one more factor : *promptness of response from designers* which closely related to the follow-up action of variations in a project. If a designer can promptly respond to contractor's request / query, a lot of abortive work can be avoided even there has to be design change. Otherwise, abortive work/materials and even co-ordination problems may occur due to late confirmation of changes. For *risk related to sub-contractors' performance and strength,* it was suggested to expand into three factors to make the definition clearer: *technical competence of sub-contractors, financial stability and cash flow of sub-contractors in construction stage* and *workload of sub-contractors.* The new items represent the likely risks related to sub-contractors employment, i.e. if they are technically or financially incapable. The focus group also commented that project manager of the project was the director of the whole construction process. The original factor, *Project leader's experience,* was not enough to illustrate the full impact of a project manager's characters and abilities. Therefore, it was expanded into three factors: *past experience of project manager in similar project, claims consciousness of project manager who will be assigned to the project,* and *interpersonal skills of project manager who will be assigned to the project.*

Nevertheless, four items were recommended to be deleted. These included *nature of structure frame, contractor's past performance or image, safety and social risk* and *project schedule (time constraint).* The *nature of structure frame* mainly affects the complexity level of the project, e.g. to use certain kind of heavy plant for construction, special construction sequence, etc. Hence, it was better to incorporate this into *complexity of project* to avoid duplication. In terms of the financial and progress control of the project, the most crucial people are the project manager and the project team. Even if the company is very experienced in constructing the project, the project will not be successful if the project team is not competent. Therefore, *contractor's past performance or image* was deleted as the skills and experienced of project manager is covered in other items. *Safety and social risk* was suggested in the literature as one of the factors affecting construction risk. Safety provision is a mandatory requirement which is well understood and practiced by the contractors. On the other hand, we never have social pressures affecting the progress of construction works in Hong Kong. Even if there is project arising public's interest e.g. environmental issue concerned by green bodies, the concerned parties' interest is usually settled before commencement of work. Hence, the *safety and social risk* was deleted according to the majority of the focus group's view. For the *schedule (time constraint)* item, every project in Hong Kong is facing this problem as all the clients, including the government bodies, customarily allow the shortest possible duration for construction. Hence, the group believed that when all projects had to cope with this issue on an equal basis, this item would not particularly affect any one project. Thus, it was finally deleted from the list.

Whilst there were items deleted from the list, there were new items suggested to be added. The new items are summarized in Table 7-15. The new items were related to four main areas: project, economics, internal resources, and relationships.

Table 7-15 Summary of Additional Factors Affecting Project Overhead Expenses as Suggested by Focus Group

Group	Suggested new factors
Project	Estimated value of project works
	Selection method by client (i.e. negotiation or competitive tendering)
	Procurement method of project (e.g. traditional, design and build, management contract)
	Extent of off-site fabrication or prefabrication in the project works
Economics	Economic conditions of Hong Kong in construction stage
	Economic conditions of the construction market in construction stage
	Interest rate during the construction stage
	Inflation rate during the construction stage
contractor	Extent of standardization in site management documentation (e.g. standard policy and forms used for inspection, query, etc.)
	Degree of computerization in site office
	General salary level of project management staff
	Size of the company (the contractor)
Relationships	Contractor is client's subsidiary firm
	Relationship with client in past projects
	Relationship with designer in past projects
	Relationship with PQS in past projects

The focus group finally agreed on the 47-item list for factors affecting project overheads. Furthermore, the group also suggested to divide the factors into external variables (factors that were beyond the control of the contractor, e.g. project nature, contract conditions, etc.) and internal variables (factors within the control of contractor, e.g. staffing, sub-contractors competence, external relations, etc.) to reflect the contractor's perspective better. The list of factors is shown in Table 7-16.

Table 7-16 Factors Affecting Project Overhead Expenses for Opinion Survey

Item ref.	Variables
	External variables
E.1	Estimated value of project works
E.2	Duration of Project
E.3	Type of Project (e.g. commercial, residential, industrial)
E.4	Size of Project (gross floor area of proposed building)
E.5	No. of storey of proposed structure
E.6	No. of level of basement (of proposed structure)
E.7	Complexity of project (use of special techniques / plants / technicians)
E.8	Shape of site
E.9	Shape of building
E.10	Site coverage of proposed building
E.11	Accessibility of site
E.12	Required quality level of project
E.13	Selection method by client (i.e. negotiation or competitive tendering)
E.14	Procurement method of project (e.g. traditional, design and build, management contract)
E.15	Type of contract (e.g. standard form, government form)
E.16	Payment terms
E.17	Completeness of project information and drawings from architect / engineer (drawings, design, specifications)
E.18	Extent of bond / warranty requirement from the project
E.19	Soil conditions
E.20	Promptness of response from designers
E.21	Extent of variations
E.22	Strictness in supervision by client / client's representatives
E.23	Punctuality of payment by client
E.24	Economic conditions of Hong Kong in construction stage
E.25	Economic conditions of the construction market in construction stage
E.26	Interest rate during the construction stage
E.27	Inflation rate during the construction stage
	Internal Variables
I.1	Extent of off-site fabrication or prefabrication in the project works
I.2	General salary level of project management staff
I.3	Extent of sub-letting of project works
I.4	Technical competence of sub-contractors
I.5	Extent of standardization in site management documentation (e.g. standard policy and forms used for inspection, query, etc.)
I.6	Degree of computerization in site office
I.7	Extent of mechanical plants owned by the contractor
I.8	Size of the company (the contractor)
I.9	Contractor is client's subsidiary firm
I.10	Relationship with client in past projects
I.11	Relationship with designer in past projects
I.12	Relationship with PQS in past projects
I.13	Past experience of project manager in similar project
I.14	Claims consciousness of project manager who will be assigned to the project
I.15	Interpersonal skills of project manager who will be assigned to the project
I.16	Technical competence of site management staff
I.17	No. of management staff assigned on site for the project
I.18	Extent of mechanical plants used for the project
I.19	Financial stability and cash flow of sub-contractors in construction stage
I.20	Workload of subcontractors

7.3.2 Survey Results on Impact of Factors Affecting Project Overheads

As explained in the Methodology Chapter, the list of factors developed in the focus group discussion (as shown in Table 7-16) was used to design the questionnaire to collect contractors' views on the impact of these factors to project overhead expenses. This was to make sure that only the significant factors were extracted for use as the input variables in the estimating model to predict the project overhead cost.

7.3.2.1 Response to Survey on Impact of Factors Affecting Project Overheads

According to the list as at 31 July 2005, there were one hundred and nine Group C contractors who were allowed to tender for public works of HK$50 million or more. In the survey, there were seventy-nine responses received, representing a response rate of 72.48% which was considered satisfactory (a minimum response rate of 30% from a minimum sample size of one hundred and seven is recommended by Fellows and Liu, 2003).

7.3.2.2 Response Demographics

63% of the respondents were general contractors engaged in civil engineering, building and retrofitting works. The rest of the respondents were civil engineering contractors specialised in civil engineering, foundation and site formation works. All of the respondents had more than fifteen years experience in the quantity surveying profession.

7.3.2.3 Preliminary Analysis of Survey Results

As highlighted in the Development of Inputs and Outputs to Cost Model Chapter, two basic assumptions of factor analysis have to be tested: multivariate normality and sampling adequacy; before carry out factor analysis. Results of the two tests for the external and internal variables are shown in Table 7-17.

Table 7-17 Results of KMO Test and Bartlett's Test of Sphericity

Test	External variables	Internal variables
KMO test	0.659	0.731
Bartlett's test of sphericity	909.435 at 0.000 significance	630.774 at 0.000 significance

The KMO test results for both sets of variables were higher than 0.5 (0.659 and 0.731 respectively) which indicated that the distribution of the values in the matrix was adequate for conducting factor analysis (George and Mallery, 1999). The Bartlett's test of sphericity results were high enough (909.435 and 630.774 respectively) with significance <0.05, indicating that the data was approximately multivariate normal and acceptable for factor analysis (George and Mallery 1999, Lattin et al 2003). The high values in the Bartlett's test results also confirmed that the correlation matrix could not be concluded as an identity matrix (Lattin et al 2003).

7.3.2.4 No. of Components Extracted

With the positive results in multivariate normality and sampling adequacy, components could be extracted by exploratory factor analysis to identify the principal factors affecting project overheads for the ANN estimating model. In the extraction process of the two sets of variables, the number of components retained was based on the eigenvalues. As explained in the Development of Inputs and Outputs to Cost Model Chapter, only components with eigenvalues larger than 1.00 were retained. There were eight external factors and six internal factors extracted with eigenvalues > 1, as indicated in Table 7-18.

Table 7-18 Eigenvalues of External Factors and Internal Factors Extracted

External Factor	Eigenvalue	Internal Factor	Eigenvalue
1	5.9428	1	5.4396
2	2.8817	2	2.9085
3	2.0119	3	1.6721
4	1.7301	4	1.3894
5	1.6358	5	1.1486
6	1.5217	6	1.0309
7	1.2997		
8	1.1647		

Figure 7-11 and Figure 7-12 portray the scree plots of the eigenvalues of the external factors and internal factors respectively. From the scree plot of external factors (Figure 7-11), it was after factor 8 where the slope of the line changed, which confirmed the number of factors extracted to be eight. Similar 'elbow' was found after factor 6 in Figure 7-12 for the scree plot of internal factors.

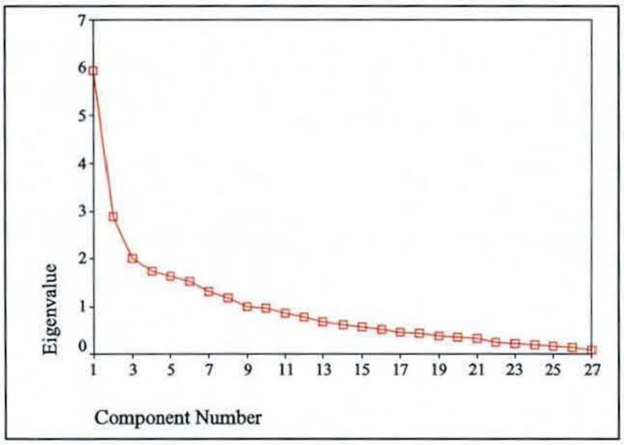

Figure 7-11 Scree Plot of Eigenvalues of External Factors

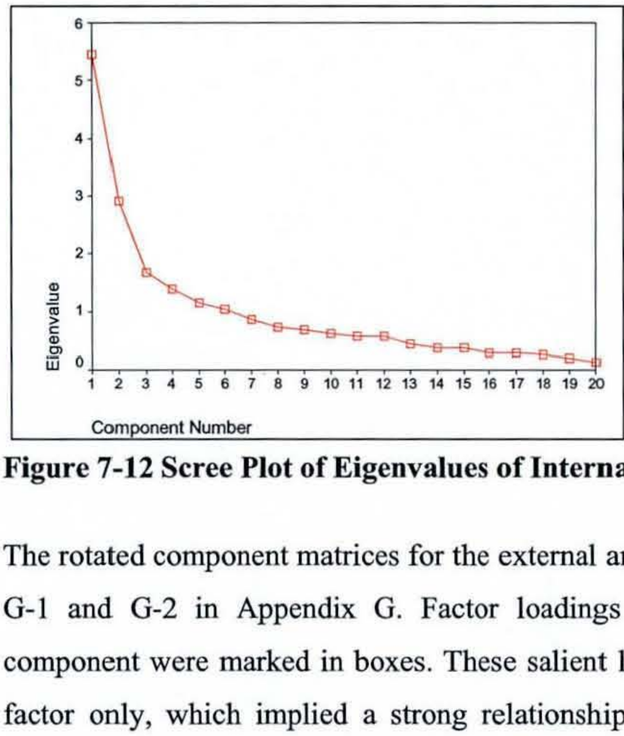

Figure 7-12 Scree Plot of Eigenvalues of Internal Factors

The rotated component matrices for the external and internal factors are tabulated in Tables G-1 and G-2 in Appendix G. Factor loadings higher than 0.6 under each extracted component were marked in boxes. These salient loadings were found to be loaded on one factor only, which implied a strong relationship existed between the variables and the factor. For instance, variables E.17, E.20 and E.21 exhibited high loadings in component 1 (from 0.60 to 0.80) but not in any other components. They were therefore grouped under component 1. Those variables not included in the extracted factors (both external and internal) were relatively low in their factor loadings, ranging from 0.03 to 0.4. On the other hand, from the correlation matrices in Tables G-3 and G-4 (in Appendix G), higher correlations were found among the variables that were put under the same component. For example, the correlation of E.17 and E.20 to E.21 was 0.565 and 0.586 respectively.

Therefore, factor labels were designed according to the nature of variables under the factor concerned. Tables 7-19 and 7-20 detail the variables identified for each factor. As shown in the tables, the extracted external factors and internal factors accounted for 67.364% and 67.945% of the total variance respectively. Since a minimum of 60% total variance was met, the extraction was considered acceptable (Ghosh and Jintanapakanont, 2004).

Table 7-19 Principle External Factors Affecting Project Overheads Expenditure

Factor	Ref.	Variables included in the factor	Factor loading	Variance explained %	Cumulative variance %
Design and variations	E.17	Completeness of project information and drawings from architect / engineer (drawings, design, specifications)	0.800	9.386	9.386
	E.20	Promptness of response from designers	0.714		
	E.21	Extent of variations	0.600		
Economic condition during construction stage	E.24	Economic conditions of Hong Kong in construction stage	0.901	9.276	18.662
	E.25	Economic conditions of the construction market in construction stage	0.888		
Bonds and Warranties	E.18	Extent of bond / warranty requirement from the project	0.732	8.960	27.622
	E.27	Inflation rate during the construction stage	0.713		
Project complexity	E.3	Type of Project (e.g. commercial, residential, industrial)	0.656	8.793	36.415
	E.5	No. of storey of proposed structure	0.615		
	E.6	No. of level of basement (of proposed structure)	0.672		
	E.7	Complexity of project (use of special techniques / plants / technicians)	0.715		
Procurement nature	E.13	Selection method by client (i.e. negotiation or competitive tendering)	0.731	8.476	44.891
	E.14	Procurement method of project (e.g. traditional, design and build, management contract)	0.629		
	E.15	Type of contract (e.g. standard form, government form)	0.741		
Site and building shape	E.8	Shape of site	0.752	8.208	53.099
	E.9	Shape of building	0.721		
Location	E.11	Accessibility of site	0.708	8.179	61.278
	E.22	Strictness in supervision by client / client's representatives	0.721		
Project size	E.4	Size of Project (gross floor area of proposed building)	0.746	6.086	67.364

Table 7-20 Principal Internal Factors Affecting Project Overheads Expenditure

Factor	Ref.	Variables included in the factor	Factor loading	Variance explained %	Cumulative variance %
Strength of site management team	I.13	Past experience of project manager in similar project	0.680	14.854	14.854
	I.15	Interpersonal skills of project manager who will be assigned to the project	0.776		
	I.16	Technical competence of site management staff	0.773		
Contractor's past relationship with project parties	I.10	Relationship with client in past projects	0.728	13.429	28.283
	I.11	Relationship with designer in past projects	0.852		
	I.12	Relationship with PQS in past projects	0.834		
Extent of staff and plants assigned to project	I.17	No. of management staff assigned on site for the project	0.677	12.583	40.866
	I.18	Extent of mechanical plants used for the project	0.799		
Financial strength of sub-contractors	I.3	Extent of sub-letting of project works	0.670	10.608	51.474
	I.19	Financial stability and cash flow of sub-contractors in construction stage	0.730		
Sub-contractors performance	I.4	Technical competence of sub-contractors	0.682	8.938	60.412
	I.5	Extent of standardization in site management documentation (e.g. standard policy and forms used for inspection, query, etc.)	0.667		
Contractor is client's subsidiary	I.9	Contractor is client's subsidiary firm	0.682	7.534	67.945

To enhance the understanding of these principal factors, the relative importance index and relative rank of each factor was calculated (details of the calculation method is shown in equation 5-1 in Chapter 5). Table 7-21 and Table 7-22 present the relative rank and importance index of each factor and the relative importance of each variable within a factor.

Table 7-21 Importance Index of External Factors

Rank	Variables included in the factor	Importance Index
1st	**Project size**	**73.418**
1	Size of Project (gross floor area of proposed building)	73.418
2nd	**Project complexity**	**68.987**
1	Complexity of project (use of special techniques / plants / technicians)	80.591
2	No. of level of basement (of proposed structure)	68.776
3	Type of Project (e.g. commercial, residential, industrial)	64.557
4	No. of storey of proposed structure	62.025
3rd	**Location**	**65.401**
1	Accessibility of site	71.308
2	Strictness in supervision by client / client's representatives	59.494
4th	**Economic condition during construction stage**	**63.080**
1	Economic conditions of the construction market in construction stage	65.401
2	Economic conditions of Hong Kong in construction stage	60.759
5th	**Bonds and Warranties**	**58.017**
1	Extent of bond / warranty requirement from the project	60.338
2	Inflation rate during the construction stage	55.696
6th	**Procurement nature**	**54.852**
1	Procurement method of project (e.g. traditional, design and build, management contract)	67.089
2	Type of contract (e.g. standard form, government form)	49.789
3	Selection method by client (i.e. negotiation or competitive tendering)	47.679
7th	**Design and variations**	**54.430**
1	Extent of variations	59.072
2	Completeness of project information and drawings from architect / engineer (drawings, design, specifications)	57.806
3	Promptness of response from designers	46.414
8th	**Site and building shape**	**45.357**
1	Shape of site	47.257
2	Shape of building	43.460

Table 7-22 Importance Index of Internal Factors

Rank	Variables included in the factor	Importance Index
1st	**Extent of staff and plants assigned to project**	**74.895**
1	No. of management staff assigned on site for the project	75.949
2	Extent of mechanical plants used for the project	73.840
2nd	**Financial strength of sub-contractors**	**60.759**
1	Extent of sub-letting of project works	62.869
2	Financial stability and cash flow of sub-contractors in construction stage	58.650
3rd	**Strength of site management team**	**56.399**
1	Technical competence of site management staff	64.135
2	Past experience of project manager in similar project	56.540
3	Interpersonal skills of project manager who will be assigned to the project	48.523
4th	**Contractor is client's subsidiary**	**53.586**
1	Contractor is client's subsidiary firm	53.586
5th	**Sub-contractors performance**	**53.165**
1	Technical competence of sub-contractors	54.430
2	Extent of standardization in site management documentation (e.g. standard policy and forms used for inspection, query, etc.)	51.899
6th	**Contractor's past relationship with project parties**	**39.100**
1	Relationship with client in past projects	45.148
2	Relationship with designer in past projects	39.241
3	Relationship with PQS in past projects	32.911

7.3.2.4.1 External variables

As mentioned before, the external variables refer to the variables that cannot be controlled by the contractor. The meaning of each factor is interpreted in the order of their importance index.

7.3.2.4.1.1 1st Factor – Project size

Although this factor accounted for the least total variance (6.086%), it ranked the highest amongst the eight factors. It contained only one variable, size (gross floor area) of the project. Gross floor area is a useful indicator of the "workload" involved in a project, and thus often used as a multiplier when estimating construction cost from unit rates. Therefore, it is reasonable to find project size as one of the major factors affecting project overhead expenses.

7.3.2.4.1.2 2nd Factor – Project complexity

This factor ranked the second, accounting 8.793% of the total variance. Four items: complexity of project (use of special techniques / plants / technicians), number of basement levels, type of project, and number of storeys of proposed structure were included in this factor. Among the four variables, complexity of project was the most important. The factor represented the level of difficulty and complexity to construct the facility. Therefore, it directly affects the extent of site facilities and management skills required for the satisfactory completion of the works.

7.3.2.4.1.3 3rd Factor – Location

This factor accounted for 8.179% of the total variance. Two variables were contained in this factor: accessibility of site and strictness in supervision by client / client's representatives. Both of these variables are closely related to the site location. As discussed in the literature review, a lot of authors advocated the importance of location (i.e. accessibility) to project overhead expenses. If the site is remotely located, the contractor will have to construct temporary roads/bridges and provide alternative transport facilities. These provisions are for the purpose to maintain the flow of workers and materials; as well as the inspections by the client's representatives. Very often, additional trips have to be provided for the client's visits as they are not regularized. It is certain that more overhead allowance has to be made if the client closely supervises the job.

7.3.2.4.1.4 4th Factor – Economic condition during construction stage

Factor 4 accounted for 9.276% of the total variance. It composed of two items: economic conditions of Hong Kong in construction stage and the economic conditions of the construction market in the construction stage. In the last decade, Hong Kong has experienced a drastic recession after the financial crash in 1997. The Composite Consumer Price Index dropped from 106.9 to 92 (-13.94%), while the gross value of construction work performed was declining significantly from HK$ 242,843 million to HK$ 163,883 million (-32.51%) from 1997 to 2003 (Census and Statistics Department, Hong Kong SAR Government, 2005). Obviously, the economic conditions will affect the overall price level as well as the level of construction activities. As reflected by the importance index in Table 7-21, the economic condition of the industry casts a more direct impact on the project overhead costs than the economic condition of Hong Kong.

7.3.2.4.1.5 5th Factor– Bonds and Warranties

Bonds and warranties factor accounted for 8.96% of the total variance. Two variables had high loading in this factor: extent of bond / warranty requirement from the project and inflation rate during the construction stage. In recent years, public and private clients in Hong Kong have escalated their requirements of bonds / warranty significantly to reduce their own risk of latent defects after occupation, like water leakage from waterproofing treatment / structural glazing. The indemnity from the bonds is often valid for ten years or even fifteen years. From the point of view of a contractor, providing these surety bonds to the client means a loss of opportunity to gain interest. Since the interest rate and the inflation rate are closely linked together, an upward trend of inflation implies a greater loss of interest.

7.3.2.4.1.6 6th Factor– Procurement nature

Factor 6 reflected the cost impact arising from the procurement arrangements of the contractor. It included procurement methods, nature of contract and selection method. This factor accounted for 8.476% of the total variance. As indicated from the importance indices of the three variables, the impact of procurement method (of contractor) to the project overhead expenses ranked the highest. The result is reasonable as different procurement methods are associated with different roles and obligations, and certainly different liabilities of the contractor. For example, the design-and-build contractor will have more flexibility to control the standard of site accommodation than the contractor employed under traditional designer-led procurement system.

7.3.2.4.1.7 7th Factor– Design and variations

This factor accounted for the largest amount of the total variance (9.386%). It encompassed three variables : completeness of project information and drawings from architect / engineer, promptness of response from designers, and extent of variations. They represented the impact of design variations to the expenditure of project overheads. It is clear that design changes in the construction stage will have a direct impact on some project overhead provisions like re-submission of samples, revision of shop drawings, closer supervision, etc. Whilst the extra expenditure on the additional supportive work is difficult to isolate and prove, it is often non-reimbursable in the valuation of variations.

7.3.2.4.1.8 8th Factor– Site and building shape

The 8th factor related to site and building shape accounted for 8.208% of the total variance. The factor contained two variables: shape of site and shape of building. Both of these factors influenced the provision / design of site facilities. For example, a site with an irregular shape may lead to higher cost of hoarding, protective fencing, temporary lighting, etc. On the other hand, irregular building shape increases scaffolding area, and requires more hoisting equipment to provide sufficient coverage for vertical transportation of materials. Nevertheless, compared with the other seven factors, impact arising from this factor is the easiest to estimate as the site and building shape is almost fixed throughout the construction stage.

7.3.2.4.2 Internal variables

There were altogether twenty internal variables analyzed, covering various aspects of a contractor's internal resources. Six factors were extracted and explained according to their relative importance as indicated in Table 7-22.

7.3.2.4.2.1 1st Factor – Extent of Staff and Plants Assigned to Project

Extent of staff and plant assigned to project ranked the first, contributing 12.583% of the total variance. It consisted of two variables: number of management staff assigned on site for the project, and the extent of mechanical plant used for the project. As revealed from past literature and earlier project data analysis in Section 7.2.4.2.1, staffing and plant accounted for almost 50% of the project overhead costs. It is therefore understandable for this factor to be the most significant one affecting project overheads as perceived by the contractors.

7.3.2.4.2.2 2nd Factor – Financial Strength of Sub-contractors

Financial strength of sub-contractors, contributed 10.608% of the total variance, ranked the second among the internal factors. Two variables represented the significant loadings of this factor: extent of sub-letting of project works and financial stability / cash flow of sub-contractors in the construction stage. In Hong Kong, construction projects are highly fragmented with most of the trade works being sublet to sub-contractors. If a sub-contractor has financial difficulty, he will normally cut his overhead expenses in e.g. drawing submission, cleaning, safety provisions, protection to finished works, etc. to maintain sufficient cash to complete the subcontract works. Unless the main contractor changes to another sub-contractor (which may cost even more), he has to pay extra manpower and money to support the sub-contractor and provide the outstanding project overheads.

7.3.2.4.2.3 3rd Factor – Strength of Site Management Team

The 3rd factor was related to the strength of site management team. It contributed to the greatest amount of variance (14.854%). This factor contained a combination of three variables: technical competence of site management staff, past experience of project manager in similar project and interpersonal skills of the project manager. Clearly, these three variables represented the strength of the whole management team under the leadership of the project manager. Strong management team helps to make savings in project overheads, for instance, less construction waste, shorter materials transit distance, better utilisation of mechanical plants, etc.

7.3.2.4.2.4 4th Factor – Contractor is Client's Subsidiary

This factor contained only one variable: contractor is client's subsidiary firm; and accounted for 7.534% of the total variance. This factor explained the convenience and advantage in vertical business integration. In Hong Kong, a few contractors are subsidiaries of large property developers and are greatly benefited from such a relationship.

7.3.2.4.2.5 5th Factor – Sub-contractors Performance

This factor accounted for 8.938% of the total variance. It contained two variables : technical competence of sub-contractors and the extent of standardization in site management documentation. It was the second factor attributable to sub-contractors' background. In general, if the contractor employs a competent crew of sub-contractors and maintains well-organised standard forms to manage these sub-contractors, the contract works should progress smoothly and efficiently. This will bring savings in project overheads to the main contractor, e.g. less waste, better housekeeping, optimized use of facilities, etc.

7.3.2.4.2.6 6th Factor – Contractor's Past Relationship with Project Parties

The 6th factor, Contractor's past relationship with project parties, accounted 13.429% of the total variance. This factor contained a combination of 'relationship' variables: relationship with client in past projects, relationship with designer in past projects, and relationship with Professional Quantity Surveyor (i.e. consultant quantity surveyor) in past projects. Good external relationship is an important factor to the success of a business. In the construction process, a lot of overhead provision requires approval by the client or his representative (e.g. standard of site accommodation and office equipment, quality of submissions, level of housekeeping, etc.). If relationship with these parties is good, the level / standard of these overhead provisions can be negotiable.

In short, the eight principal external factors extracted were : (1)*Design and variations*, (2)*Economic condition during construction stage*, (3)*Bonds and Warranties*, (4)*Project complexity*, (5)*Procurement nature*, (6)*Site and building shape*, (7)*Location* and (8)*Project size*. The six principal internal factors extracted were : (1)*Strength of site management team*, (2)*Contractor's past relationship with project parties*, (3)*Extent of staff and plants assigned to project*, (4)*Financial strength of sub-contractors*, (5)*Sub-contractors performance* and (6)*Contractor is client's subsidiary*. Regarding the various factors suggested by the past theorists as significant attributes to project overheads, this study re-affirms some of them as critical, including : *project complexity*, *accessibility*, *client's strictness in supervision*, *extent of sublet work*, *cash flow of sub-contractors*, *contract type*, and *extent of staffing and mechanical plants employed*.

7.3.3 Collection of Project Overheads Data for ANN Model

Before building the artificial neural network estimating model, project overhead cost data together with the project details related to the principal factors (as extracted by the exploratory factor analysis) were collected from the large contractors in Hong Kong.

7.3.3.1 Project Data Collected

The data collection process took almost two years from 2004 to 2006. Although invitations were sent to the Quantity Surveying Manger / Chief Quantity Surveyor of the Group C contractors at the outset, nil reply was received. Therefore, telephone calls were made to those who had joined the opinion survey before, and only two companies were willing to provide project data required. Although the data should be collected by random sampling, it was impossible to collect sufficient sample due to the sensitivity nature of the cost data. Therefore, past colleagues and friends who were project quantity surveyors working in large contractors were contacted and finally data from seventy-one construction projects were collected. Summary of the project data collected is shown in Appendix H.

7.3.3.2 Project Profile

portrays the project profile of the data collected. All the projects were building projects constructed during 1997 to 2004 (excluding the defects liability period). The gross floor area of the project ranged from 3,500 m^2 in low-rise detached houses project to 183,000 m^2 in high-class commercial building. The sample covers different variety of building projects with different degree of quality standard and complexity.

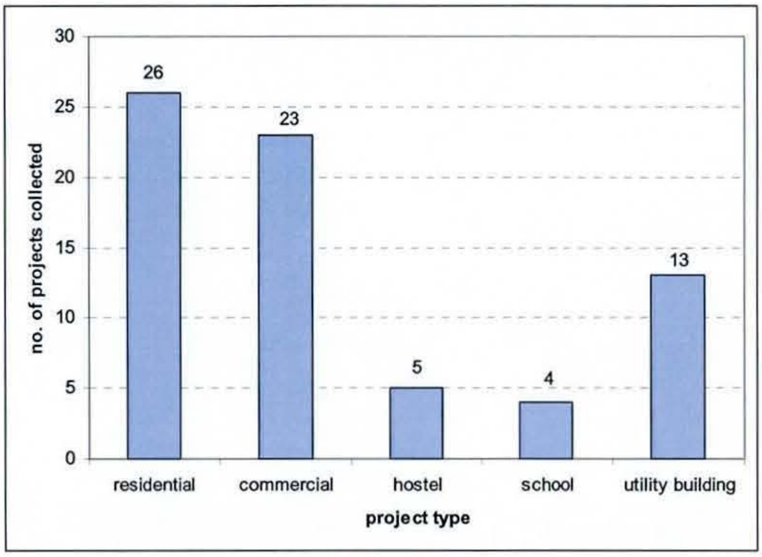

Figure 7-13 Profile of the Projects Classified by Project Nature

7.3.3.3 Project Data Transformation

As explained in the Development of Inputs and Outputs to Cost Model Chapter, 2 of the input factors were measured by an aggregation of several variables and required transformation before input to the ANN model. They are the *economic condition* and *project complexity*. For *Economic condition during construction stage*, it was measured by Consumer Price Index to represent the economic condition of Hong Kong and the Gross value of construction work performed to represent the economic condition of the construction market. The combined index was measured by weighted average of the construction volume and the consumer price index, using year 2000 as the base year. The relative weighting of construction volume and overall economy was based on the relative importance of the two variables found in the factor analysis (in Table 7-21), which was 65.401 and 60.759 respectively. Details of the combined index used to measure the *Economic condition during construction stage* across different years are shown in Table 7-23.

Table 7-23 Combined Index of "Economic Condition" Across Time

Year	Volume of construction work (in HK$ million)	Transformed construction volume index (base year 2000 = 100)	Consumer Price Index	Combined Index
	a	b = (a - 208,026)/208,026 x 100 +100	c	$d = \frac{b \times 65.401 + c \times 60.759}{(65.401+60.759)}$
1981	33,968	16.329	30.017	22.921
1982	38,082	18.306	33.300	25.527
1983	38,853	18.677	36.600	27.309
1984	40,307	19.376	39.758	29.192
1985	39,333	18.908	41.158	29.624
1986	44,860	21.565	42.617	31.703
1987	55,837	26.841	45.050	35.611
1988	69,267	33.297	48.583	40.659
1989	86,007	41.344	53.550	47.223
1990	99,842	47.995	59.042	53.315
1991	109,879	52.820	65.667	59.007
1992	120,529	57.939	71.967	64.695
1993	148,449	71.361	78.317	74.711
1994	166,923	80.241	85.192	82.625
1995	191,537	92.074	92.883	92.464
1996	216,787	104.211	98.758	101.585
1997	242,843	116.737	104.508	110.848
1998	240,687	115.700	107.492	111.747
1999	226,702	108.978	103.233	106.211
2000	208,026	**100.000**	**100.000**	**100.000**
2001	196,564	94.490	97.775	96.072
2002	184,801	88.836	94.800	91.708
2003	163,883	78.780	92.350	85.315

Another principal factor, *project complexity* was measured by a combination of several variables : complexity of project (use of special techniques / plants / technicians), number of level of basement (of proposed structure), type of Project (e.g. commercial, residential, industrial), and number of storey of proposed structure. Similarly, the relative weightings of the four variables was based on the relative importance index found in the factor analysis. The weighted average score was calculated as per Table 7-24 below.

Table 7-24 Relative Weightings of the 4 Variables Used to Measure Project Complexity Factor

Variables contained in the principal factor – Project Complexity	Relative importance index from factor analysis
Complexity of project (use of special techniques / plants / technicians)	80.59
No. of level of basement (of proposed structure)	68.78
Type of Project (e.g. commercial, residential, industrial)	64.56
No. of storey of proposed structure	62.03

The weighted average score for *Project complexity* = $\frac{\Sigma x_i \cdot w_i}{\Sigma w_i}$ **(7-1)**

where x_i is the score of variable i and w_i is the relative importance index of variable i.

7.3.4 Artificial Neural Network Cost Estimating Model

As explained in the Design of ANN Cost Estimating Model Chapter, ANN model using Group Data Handling Method (GMDH) algorithm was used as the principal method to develop the required estimating model as it was the most transparent with a polynomial function provided after training. Since NeuroShell 2 provides six predetermined criteria to select the best model under the GMDH algorithm, six trials were conducted sequentially with the same set of inputs and outputs, each using one selection criterion. In other words, the six polynomial equations produced represent the results of using six different methods to pick out the best model under the same algorithm (i.e. GMDH). When a different selection method is used in the iteration process to find the best model, it is likely that the results produced in each trial are quite different. Therefore, all the results regarding the accuracy level, correlation between predicted and actual outputs, the best polynomial equation, etc. were tabulated in Appendix I for comparative purpose. The training and production results generated from the six selection criteria were then analyzed accordingly. The results were compared with those generated from the feedforward backpropagation and general regression models. The outcomes of cross-validation of the best model were also examined. Finally, the comparison of the performance of GMDH model and prediction of multiple regression equation was presented.

7.3.4.1 Training and Production Results of ANN Model Using GMDH Algorithm With All Inputs

To maintain a reasonable data set for testing and production, around 10% of the collected data was set aside for testing and production respectively. The extraction of testing and production data was done randomly by NeuroShell 2. Eight patterns were extracted for the production set. This set of data was set aside and used in all production trials (in all types of ANN architectures). In case of the GMDH algorithm using regularity as the selection criterion (and backpropagation and GRNN models in validation), another testing set of seven patterns were extracted from the training.

Thus, there were sixty-three training cases for models not requiring testing set and fifty-six training cases for models requiring testing (including GMDH network using regularity selection criterion, backpropagation and GRNN models). All the data were scaled to the range of <<0,1>> without any clipping at the top and bottom. Table I-1 in Appendix I shows the training, testing and production results of all the six GMDH models using all fourteen variables as inputs. The results can be summarized as:

- R squared: (training) 0.7951 – 0.9558
  (production) 0.7597 – 0.9574
- Correlation coefficient: (training) 0.8917 – 0.9777
  (production) 0.8718 – 0.9648
- no. of significant variables in the best models: 5 – 11

All the six GMDH models provided good prediction, based on the R squared values obtained for the testing and production set (all around 0.8 or higher). Amongst the six models, the highest R squared value in the training and production sets was 0.9558 and 0.9574 respectively, achieved by the 4th model using GCV as the selection criteria. Nevertheless, the best model should be the first one using FCPSE as the selection criteria. Although it gave slightly lower R squared values in both the training and production set patterns when compared with the 4th model (training: 0.9192; production: 0.9202), the best formula was the simplest and lesser layers were required. On the other hand, the worst model was Regularity network, producing very complex polynomials though the R squared values were high.

Best formula (by GMDH algorithm using 14 input variables and FCPSE selection criterion):

$Y = -1.3*X_8 + 0.15 - 0.5*X_2 + 0.19*X_{10} + 0.16*X_{11} + 3.4*X_8*X_{11} + 0.38*X_2\text{^}2 + 0.57*X_5\text{^}2 - 0.65*X_2*X_5 + 1.2*X_6*X_8$

Where Y = (OUT_COST-427,000)/129,373,000; X_2 = (ECON-85.32)/26.43; X_5 = (PROCU-1)/4; X_6 = (SHAPE-1)/4; X_8 = (SIZE-3,500)/179,500; X_{10} = (RELAT-1)/4 and X_{11} = (STA_PLT-1)/4

The correlation coefficients of the six models were very high, all around 0.9. This indicates a very strong relationship between the actual and predicted values.

The six models generated very different formula for prediction, and all of them used only some but not of the fourteen input variables. Since only half of the input variables were significant in the six models, it was decided to remove the non-significant ones to see if any improved model (with good predictive accuracy and simpler formula) could be produced.

7.3.4.2 Training and Production Results of ANN Models Using 12 Variables as Inputs

Table 7-25 shows the list of most significant variables identified by NeuroShell 2 in the GMDH network. This served as a useful guideline to decide which variables to be selected. SIZE and STA_PLT variables were commonly found as significant in all models. For ECON, PROCU, SHAPE and RELAT, they were identified as significant in five models. The rest of the variables were identified as significant by at least two models. Therefore, the two variables: LOCA and SUBSID which did not appear in the list of any model were removed first.

Table 7-25 Significant Variables Identified in the 6 GMDH Models (Using 14 Variables)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	SIZE	SIZE	SIZE	SIZE	SIZE	SIZE
	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT
	ECON	ECON	ECON	ECON		ECON
	PROCU	PROCU	PROCU	PROCU	PROCU	
	SHAPE	SHAPE	SHAPE	SHAPE		SHAPE
	RELAT	RELAT	RELAT	RELAT		RELAT
		TEAM		TEAM	TEAM	TEAM
			COMP	COMP		
			DES_VO			DES_VO
				SC_FIN		SC_FIN
				SC_PER		SC_PER
				BOND	BOND	

The data was put in the GMDH network to train again, using the twelve selected variables as inputs. The results of training, testing and production were detailed in Table I-2 of Appendix I. Summary of the results are:

- R squared: (training) 0.8239 – 0.9613
  (production) 0.7056 – 0.9538
- Correlation coefficient: (training) 0.9077 – 0.9805
  (production) 0.8443 – 0.9774
- no. of significant variables in the best models: 3 – 9

The accuracy of the six models was not reduced significantly as a result of removing two input variables (LOCA and SUBSID). In fact, the R squared values of some models were improved, including the 1st (from 0.9202 to 0.9211), 2nd (from 0.7998 to 0.8684) and 5th model (0.7597 to 0.9538). In general, the prediction of the six models were good, all with the R squared values for the testing set above 0.8 and the R squared values of production set above 0.7. The correlation coefficients of all models in training and production sets were above 0.8, which indicated a strong positive relationship between the actual and predicted values. Amongst the six models, the highest R squared value in training was 0.9613 produced by the model using GCV selection criteria, and the highest R value in production was 0.9538 achieved by the model using FPE as the selection criteria.

However, the best formula produced by these two models was quite complicated, especially the one produced by the model using GCV as selection criterion. The simplest formula was produced by PSE selection criterion using two layers only. However, its R squared values were slightly lower than that of the second simplest model which used FCPSE as the selection criterion. FCPSE selection criterion produced very simple formula using three layers, with R squared of 0.9122 (in training) and 0.9211 (in testing). Therefore, having considered the best formulae and R squared values of the six models, the 1st model (FCPSE selection criterion) provided the best prediction. In other words, FCPSE selection criterion produced the best model when using fourteen or twelve variables compared with other selection criteria.

Best formula (by GMDH algorithm using 12 input variables and FCPSE selection criterion):

$Y = 0.15 - 0.5*X_2 - 0.98*X_7 + 0.19*X_9 + 0.16*X_{10} + 3.4*X_7*X_{10} + 0.38*X_2\text{^}2 + 0.57*X_5\text{^}2 - 0.66*X_2*X_5$

Where Y = (OUT_COST-427,000)/129,373,000; X_2 = (ECON-85.32)/26.43; X_5 = (PROCU-1)/4; X_7 = (SIZE-3,500)/179,500; X_9 = (RELAT-1)/4 and X_{10} = (STA_PLT-1)/4

It was also noted that the FCPSE selection criterion produced even simpler formula in the model using twelve inputs than the one using fourteen inputs. Furthermore, one less variable (SHAPE) was required to predict the cost with similar accuracy.

7.3.4.3 Training and Production Results of ANN Models Using 8 Variables as Inputs

Table 7-26 lists out the most significant variables identified by NeuroShell 2 in the 12-variable models.

Table 7-26 Significant Variables Identified in the 6 GMDH Models (Using 12 Variables)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th – FPE	6th - Regularity
Most significant variables	SIZE	SIZE	SIZE	SIZE	SIZE	SIZE
	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT
	PROCU		PROCU	PROCU	PROCU	PROCU
	RELAT		RELAT	RELAT	RELAT	RELAT
	ECON		ECON	ECON	ECON	
		TEAM	TEAM	TEAM		TEAM
			COMP	COMP		
			SHAPE	SHAPE		
				BOND		
						SC_FIN
						SC_PER
						DES_VO

All the twelve variables appeared in the models, although none of the model used up all the twelve variables as inputs. SIZE and STA_PLT were identified as significant variables in all the six models. PROCU, RELAT, ECON and TEAM were identified as significant in four models. Since BOND, SC_FIN, SC_PER, DES_VO were only identified as significant variables in one model out of six, they were removed in the next round.

Thus, SIZE, STA_PLT, PROCU, RELAT, ECON, TEAM, COMP, and SHAPE were selected as inputs to train the GMDH models again to see if there was any improved prediction. Table I-3 in Appendix I presents the training and production results of the various models using GMDH algorithm. Major results are summarized as follows:

- R squared: (training) 0.800 – 0.9523
  (production) 0.826 – 0.9211
- Correlation coefficient: (training) 0.8944 – 0.9759
  (production) 0.9133 – 0.9723
- no. of significant variables in the best models: 4 – 8

In general, the prediction of the models using eight variables was more satisfactory than those using fourteen or twelve variables, as the R squared values of training and production sets were all above 0.8. Besides, all the best formulae generated were simpler than before (except the one produced by Regularity still remained very complex). The correlation coefficients also indicated very strong positive relationship between the predicted values and the actual values in both training set and production set (all at or above 0.9). The highest R squared value in training was 0.9523 achieved by the model using MDL as the selection criterion, whereas in production set was 0.9211 achieved by the 1st model using FCPSE as the selection criterion. In this trial, the best formula was undoubtedly produced by the 1st model using FCPSE as selection criterion. It produced the simplest formula, the least number of layers (only three), the highest R squared in production (0.9211) and very high R squared in training (0.9122).

Best formula (by GMDH algorithm using 8 input variables and FCPSE selection criterion):

$Y = 0.15 - 0.5*X_1 - 0.98*X_5 + 0.19*X_7 + 0.16*X_8 + 3.4*X_5*X_8 + 0.38*X_1\text{^}2 + 0.57*X_3\text{^}2 -0.66*X_1*X_3$

Where Y = (OUT_COST-427,000)/129,373,000; X_1 = (ECON-85.32)/26.43; X_3 = (PROCU-1)/4; X_5 = (SIZE-3,500)/179,500; X_7 = (RELAT-1)/4 and X_8 = (STA_PLT-1)/4

7.3.4.4 Training and Production Results of ANN Models Using 5 Variables as Inputs

After reducing the number of inputs to about half of the original number, the ANN models still capable to produce satisfactory results on the prediction. As shown in Table 7-27, not all the eight variables were utilized in the GMDH models. For instance, in the 1st, 2nd and 5th model, only five or fewer variables were used. Therefore, the training was tried again with the use of fewer inputs.

Table 7-27 Significant Variables Identified in the 6 GMDH Models (Using 8 Variables)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	PROCU	PROCU	PROCU	PROCU	PROCU	PROCU
	SIZE	SIZE	SIZE	SIZE	SIZE	SIZE
	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT
	RELAT		RELAT	RELAT	RELAT	RELAT
		TEAM		TEAM	TEAM	TEAM
	ECON		ECON	ECON		ECON
			SHAPE	SHAPE		SHAPE
			COMP	COMP		COMP

First of all, all the models identified PROCU, SIZE, STA_PLT as significant variables. On the other hand, RELAT was identified by five models as significant. Therefore, these four variables were selected for further training. Since ECON was identified as significant by four models and being one of them as the best model producing the simplest formula, it was also selected as input for further training. As a result, the inputs to be trained again were : PROCU, SIZE, STA_PLT, RELAT and ECON.

Table I-4 in Appendix I lists out the results of GMDH modelling with the use of five input variables. Major results can be summarized as;

- R squared: (training) 0.8221 – 0.9329
  (production) 0.8182 – 0.9919
- Correlation coefficient: (training) 0.9067 – 0.9659
  (production) 0.9062 – 0.996
- no. of significant variables in the best models: 4 – 5

There was slight improvement in some models as well as the contrary in terms of the R squared values in training and production. On the whole, the performance of the new models with only five inputs remained satisfactory, with the R squared values of all the models higher than 0.8 indicating the error between predicted values and actual values was small. The correlation coefficients were high above 0.9 implying strong relationship between the predicted values and the actual values. Besides, all the models identified ECON, PROCU, SIZE, RELAT, STA_PLT as the significant variables (except for the 1st model which only identified the first four). This indicated that the five variables shortlist as the network inputs were optimal and sufficient to predict the project overhead cost.

The highest R squared in training was 0.9323 produced by the GCV selection criterion whereas in production was 0.9919 produced by the Regularity selection criterion. Nevertheless, the formula produced by Regularity was obviously over-complex, whilst the one produced by GCV was quite lengthy as well. Thus, both of them could not be considered as the best model.

Although the 1st model using FCPSE selection criterion produced slightly lower R squared value in the training set (0.8818) than before, the performance in production was the best in all the trials (R squared = 0.9754). Besides, the formula was also the simplest amongst the various models and trials. Therefore, the best ANN model using GMDH algorithm was to use 5 inputs with FCPSE as the selection criterion.

Best formula to predict project overhead cost :
by GMDH algorithm using 5 input variables and FCPSE as selection criterion

$Y = -0.92*X_3 + 0.21*X_4 + 0.17*X_5 + 3.7*X_3*X_5 - 0.44*X_1*X_3$

Where Y = (OUT_COST-427,000)/129,373,000; X_1 = (ECON-85.32)/26.43; X_3 = (SIZE-3,500)/179,500; X_4 = (RELAT-1)/4 and X_5 = (STA_PLT-1)/4

7.3.4.5 Validation of the GMDH Model

7.3.4.5.1 Cross-validation

As described in the Design of ANN Cost Estimating Model Chapter, the sample was divided into nine sub-groups for eight rounds of cross-validation. The five significant variables (ECON, SIZE, RELAT, STA_PLT, PROCU) were used as inputs. The results of cross-validation for each selection criterion are tabulated in Table I-5 to Table I-10 of Appendix I.

The overall results showed that the eight cross-validation trials gave similar outcomes as the original training and production set (as in Table I-4 in Appendix I). Among these, the maximum absolute error was 52,331,179 and the minimum absolute error was 0, both from the Regularity selection criterion model. Most of the models produced satisfactory R-squared values, above 0.7. The highest R-squared value was 0.9883 with the set H testing data produced by the model using Regularity selection criterion. However, there was one exceptionally low R-squared value (0.493) also produced by the Regularity selection criterion with the set D production data. The correlation coefficients of all models were above 0.8, indicating strong positive relationship between the actual and predicted values.

In particular with the model using FCPSE selection criterion which produced the best model, all the trials using this criterion indicated good prediction. Amongst the eight trials, the highest R-squared value was 0.9305 and the lowest was 0.889 in training. In the production data, the highest R-squared value was 0.9405 and the lowest was 0.7046. The cross-validation results gave greater reliability to the use of GMDH model to predict the project overhead cost. Besides, the satisfactory cross-validation results in the 8 trials using FCPSE selection criterion also gave confidence to the generalization of the best formula created by this model.

7.3.4.5.2 Comparison With Alternative ANN Models

The second part of the validation was to compare the estimating accuracy with other ANN models. The ANN architectures chosen included Multi-layer feed-forward with backpropagation algorithm (MLFF) and General regression neural network (GRNN). Both models used the default settings as described in the Design of ANN Cost Estimating Model Chapter. Similar to the cross-validation models, the five significant variables identified by

GMDH algorithm were used as inputs here. Results of the two models were shown in Table 7-28.

Table 7-28 Results of MLFF and GRNN Models (Using 5 Variables As Inputs)

	MLFF	GRNN
Training		
No. of patterns	56	56
R squared	0.8137	0.9398
Mean squared error	127,291,213,733,618	41,112,722,958,738
Mean absolute error	9,135,952	3,317,321
Min. absolute error	330,495	0
Max. absolute error	30,649,928	22,076,200
Correlation coefficient r	0.9105	0.9697
Testing		
No. of patterns	7	7
R squared	0.6538	0.8888
Mean squared error	619,348,255,292,047	198,929,737,980,156
Mean absolute error	16,693,218	9,086,365
Min. absolute error	2,010,680	0
Max. absolute error	59,790,768	33,025,988
Correlation coefficient r	0.8649	0.9507
Production		
No. of patterns	8	8
R squared	0.7488	0.5183
Mean squared error	461,040,440,877,639	1,434,394,687,224,360
Mean absolute error	13,879,969	19,161,918
Min. absolute error	934,010	3,056,388
Max. absolute error	54,655,488	105,009,272
Correlation coefficient r	0.8902	0.5896

The results produced in the MLFF model was satisfactory, although less accurate when compared with the results from the GMDH model. The R squared values from the MLFF model for the training, testing and production sets were all above 0.65. The correlation coefficients were also above 0.8, indicating strong positive relationship between the predicted values and the actual values. For the GRNN model, although very high accuracy produced in the training and testing sets, considerably low R-squared value was found in the production set (only 0.5183). The smoothing factor automatically calculated was 0.0643529. Nevertheless, on the whole, the two models were capable to estimate the project overheads with only the five variables producing acceptable to satisfactory level of accuracy.

The comparison results indicated that the five significant variables identified in the GMDH model, namely *Economic condition during construction stage* (ECON), *Procurement nature* (PROCU), *Project size* (SIZE), *Contractor's past relationship with project parties* (RELAT), and *Extent of staff and plants assigned to project* (STA_PLT) were reliable factors to predict project overhead cost of a project. They could be used to predict the project overheads satisfactorily not only by the GMDH network, but also other ANN architectures like MLFF and GRNN. Besides, if compared the prediction of GMDH model with the other two ANN models, GMDH provided better results. R squared values of both training and production were high in the GMDH network (0.8818 in training and 0.9754 in production).

7.3.4.5.3 Comparison With Regression Analysis

Conventional regression analysis was used to check whether the ANN model proposed in this study was a better estimating tool. As there was no need to reserve a pool of data for testing, all the seventy-one cases were presented to the stepwise regression analysis. The results were tabulated in Table 7-29.

Table 7-29 Model Summary of Stepwise Regression Analysis

Model		1	2
R		0.8048	0.8922
R squared		0.6477	0.7961
Adjusted R squared		0.6426	0.7901
Standard error of estimate		19,119,239	14,652,364
Degree of freedom	regression	1	2
	residual	69	68
Overall F		126.8375	132.7215
Significance of F		0.0000	0.0000

Model 1 : OUT_COST = 18,105,412.5978 + 640.3737 (SIZE)

Model 2 : OUT_COST = -23,337,799.14 +532.4383 (SIZE) - 15,488,602.1262 (STA_PLT)

As model 2 produced higher R squared and with at least 95% confidence for the B values (i.e. constants and coefficients of variables) in the regression equation, model 2 was a better predictive model than model 1.

As a result, the regression equation generated was :

OUT_COST = -23,337,799.14 +532.4383 (SIZE) + 15,488,602.1262 (STA_PLT)

The analysis identified two significant variables, SIZE and STA_PLT (as shown in Model 2 of Table 7-29). The multiple R value (R=0.8922) showed a substantial correlation between the two predictors and the output (project overhead cost). The R-squared value was 0.7961, much lower than the one produced by the GMDH model (R squared = 0.9014 for all seventy-one cases; 0.8818 for training set and 0.9754 for production set). The β values of

SIZE and STA_PLT were 0.6540 and 0.4407 respectively, indicating a relative higher influence of SIZE on the outcome than STA_PLT. On the other hand, although only two significant variables (SIZE and STA_PLT) were identified by the regression analysis, they were identified as significant in the GMDH model as well. The results demonstrated clearly that the GMDH model outperformed the regression analysis to predict the project overhead cost.

7.3.4.5.4 Appraisal by Practitioners

To make sure that the proposed ANN model is user-friendly and applicable to the industrial environment, eight senior quantity surveyors who had participated in the focus group discussion before in scrutinizing the factor list were invited. However, due to tight schedules of some focus group members, only seven of them could participated in the appraisal of the ANN model. All of them possessed more than twenty years industrial experience, and were currently working in a Group C contractor in Hong Kong.

The overall discussion lasted for one hour and fifteen minutes. The basic idea of the ANN model, the significant factors affecting project overheads, the best model (equation) found by GMDH network and the predictions of the model were presented and explained to the members. Members were asked to rate the proposed model in five attributes. Results of their ratings were tabulated in 7-30 below.

7-30 Summary of Ratings Given by Focus Group Members

Attributes	Members' Ratings*							Average Rating
	1	2	3	4	5	6	7	
Easy to understand	4	3	3	4	5	4	3	3.71
Simple to use	5	4	3	4	4	3	3	3.71
Time saving	5	4	4	4	5	5	4	4.43
Accuracy	5	4	5	4	5	5	4	4.57
Likelihood to apply	5	2	5	4	4	5	4	4.14

* Note : Rating was given in the scale of 1 to 5. Higher rating indicates better performance or higher likelihood level.

The average rating of each attribute is an average of the ratings given by the seven members. From the overall ratings, *time saving* and *accuracy* scored close to the highest score, indicating an unanimous agreement in the satisfactory efficiency and accuracy of the proposed model. However, the average scores of *easy to understand* and *simple to use* were mediocre. Members who gave a lower score in these two attributes expressed concern in ANN modelling. Although the arrival of the predicted value using the best model equation was very simple, the concepts of ANN were far too complicated. If the company wanted to apply the technique to build a tailored ANN model, users must acquired some background knowledge of ANN. This may be too difficult to most estimators in the industry. Members also supplemented that the robustness of the ANN package cast difficulty for the practitioners to fully utilize and understand it.

Six out of seven members indicated positive view towards the likelihood of applying the proposed ANN model to the contractors. They commented that the model was an effective tool to estimate project overheads. On the other hand, such model could also be applicable in consultant quantity surveying firms to build up better estimates for project overheads. However, one member held an opposite thought that application of the model in Hong Kong contractors was quite unlikely. He explained that artificial intelligent model like ANN was not known by most practitioners. Due to their lack of knowledge, estimators would not renounce their traditional practice. Furthermore, top management of some contractors was conservative and preferred to have input from professionals, rather than artificial intelligence, in the estimating process. Although there are different views regarding the likelihood to apply the ANN model in the contractors, majority of the members held a positive view and thought that the proposed model was viable in terms of the accuracy and relatively low investment in the IT infrastructure.

All of the members conceded that the model or equation did provide a satisfactory estimate with minimal estimating effort. The technique should worth development into a tailored model in contractors or project consultants in order to enhance the estimating accuracy further.

7.4 Summary

This Chapter explained the results collected and demonstrated the analysis for the whole study. The exploratory study confirmed that estimators in Hong Kong spent a lot of effort in preparing estimates of project overheads. Although it was noted that the project overhead estimates were important and the associated errors were the most difficult to transfer, the accuracy was not satisfactory. From the analysis of twenty projects data, site management fee contributed the largest amount of project overhead cost (35.9%), followed by mechanical plants (12.4%), cleaning and removal of rubbish (10%) and insurances and bonds (7.4%).

From the analysis of the opinion survey on the impact of factors affecting project overhead cost, fourteen principal factors were identified by exploratory factor analysis. These fourteen factors were then used as inputs to build the GMDH estimating model to estimate project overheads. Project cost data from seventy-one projects were collected from the large contractors for analysis. The GMDH model outperformed other alternative ANN models (MLFF and GRNN) and the conventional regression analysis, giving satisfactory results of R squared > 0.9. In the GMDH modelling process, five significant variables were identified as reliable factors to predict project overheads, including ECON, PROCU, SIZE, RELAT, and STA_PLT. Multiple linear regression analysis also identified two of these five variables as significant (SIZE and STA_PLT).

In the eight cross-validation trials of GMDH network, all of them produced satisfactory estimation (all R squared values > 0.7 except the trial in set D production data produced by Regularity selection criterion which only gave 0.593). The training and validation results proved that GMDH network was a suitable and reliable tool to predict project overhead cost. Besides, the five variables identified by GMDH network were tried as inputs in multi-layered feedforward network and general regression neural network. Both of them produced satisfactory results as well. Therefore, these variables could be used to predict project overheads cost with satisfactory accuracy. They included *Economic condition during construction stage* (ECON), *Procurement nature* (PROCU), *Project size* (SIZE), *Contractor's past relationship with project parties* (RELAT), and *Extent of staff and plants assigned to project* (STA_PLT).

Finally, the proposed ANN model was appraised by a focus group comprising seven senior quantity surveyors working in the large contractors. They unanimously agreed that the proposed model was efficient and accurate. In terms of application amongst the contractors, almost the whole group thought that application to the industry was likely. Only one member concerned about the resistance from some conservative contractors / practitioners due to their lack of knowledge in ANN modelling. Nonetheless, in view of the attractiveness of the model, all members agreed that it was worthwhile to develop the model as a tailored system in the contractors / consultant firms.

CHAPTER 8

Analysis and Validation of ANN Model

8 Analysis and Validation of ANN Model

8.1 Introduction

After examining the results of the various stages of study in the Data Analysis Chapter, the nature of construction project overheads and estimation was ascertained. The established facts about project overhead cost and the estimating practice are very useful as there is no data of this kind published in the past. Estimating model for project overheads was also successfully developed in the Data Analysis Chapter. In this Chapter, some areas of concern are highlighted which may be interested to academics who want to pursue further study in the subject of project overheads and for professionals who perform related duties in the industry. Limitations of this research are also explained, with identification of areas for future study.

8.2 General Nature of Project Overheads

In the first part of the exploratory study, the overall nature and characteristics of project overheads expenditure was determined. The findings affirm the complex nature of project overheads: diversified estimating method, difficult to estimate, difficult to transfer the estimating risk to other parties, reliant on uncontrollable factors, etc. The percentage of project overheads to the total project cost, as found in the study, ranged from 11% to 19% (quite in line with the literature of not more than 20% by Assaf et al, 1999 and Solomon, 1993). This is a substantial percentage when compared with the low profit margin in the competitive market nowadays. On the other hand, the percentage range of project overhead contribution (to total cost) is wide. This further confirms that the practice of applying a percentage to the cost of measured work in order to estimate the project overheads allowance is inappropriate. The minority of estimators who are using this technique to estimate project overheads should look for a more accurate approach.

8.3 Misalignment between Estimating Practice and Estimator's Perceptions

From the exploratory study, it is found that the estimation of project overheads is more of an intuitive operation than a scientific process. Estimators believed that project overheads accounted for a small proportion of total project cost, and were comparatively less important than other estimates. Nevertheless, all of them intuitively allocated the most expertise from senior estimators to prepare detailed estimates of project overheads. On the other hand, majority of them agreed that the risk of estimation errors in project overheads was the most difficult to transferred, but most of the senior estimators still relied on their subjective knowledge and experience to work out the estimates. These results identify that there is misalignment between the perceived importance of the project overheads estimation and the level of resources allocated to the estimating task. It is also evidenced that the estimating process for project overheads is heavily relied on "professional judgment" which is subjective in nature. This is a major concern as raised by a lot of theorists in the past regarding estimating practice (Lowe and Skitmore, 1994; Oteifa and Baldwin, 1991). Unfortunately, no significant improvement is made after a decade. Cost estimating is a factual process (Akintoye and Fitzgerald, 2000). To improve the accuracy of estimation, good knowledge management by the contractor is indispensable, and the estimating process should make reference to objective information to improve the accuracy of estimates. The use of estimating model to estimate project overheads avoids the said errors and biases.

8.4 Estimating Strategy for Itemized Project Overhead Cost

Although an ANN estimating model is devised in this study as a fast and reliable method to estimate project overhead cost, contractors may still estimate individual project overhead item cost for cost control purpose or for client's payment. From the analysis, it is noticed that estimators spent a lot of expertise in estimating project overheads in a detailed manner. With the reveal of project overheads cost distribution in this study, estimation effort and resource can be more thoroughly planned to economize the use of estimating resource.

Based on Figure 7-10, estimating resource can be better allocated according to the likely inaccuracy and potential cost of the item. As shown in Figure 8-1, four basic approaches to tackle the estimation of a single project overhead item are thus derived in case if itemized project overhead cost estimates are required. They are listed below in the order of priority.

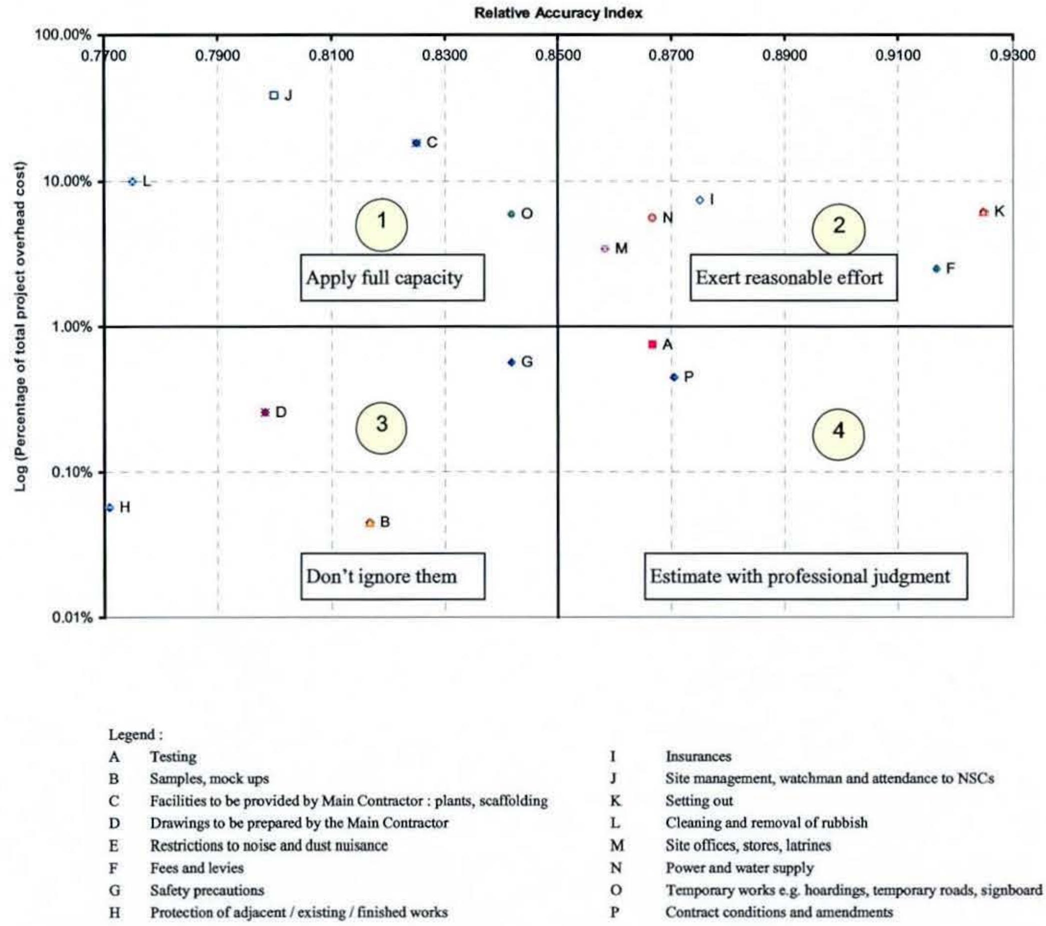

Figure 8-1 Decision Matrix for Estimating Resource Allocation

1. Apply full capacity - for the project overhead items that lied in the 1st quarter: high in percentage contribution to total project overhead cost and high in estimation inaccuracy.

   Items lied in this quarter included site management; watchman and attendance to nominated sub-contractors; facilities to be provided by Main Contractor: plants, scaffolding; cleaning and removal of rubbish; and temporary works e.g. hoardings, temporary roads, signboard. These items accounted a large proportion of project overhead cost and were the most difficult to estimate. Therefore, full estimating capacity should be allowed to these items.

2. Exert reasonable effort – for the items lied in the 2nd quarter: high in percentage contribution to total project overhead cost and low in estimation inaccuracy.

Items lied in this quarter included insurances; power and water supply; site offices, stores, latrines; setting out; fees and levies. These items were costly but their estimates exhibited a low likelihood of inaccuracy. Since expenditure on these items is generally high, an accurate estimation is necessary. Nevertheless, the uncertainty involved in these items is less and a reasonable level of estimation effort to be paid on them is justifiable.

3. Don't ignore them – for the 3rd quarter items: low in percentage contribution to total project overhead cost and high in estimation inaccuracy.

   Items in this category included safety precautions; drawings to be prepared by the Main Contractor; protection of adjacent / existing / finished works; and samples / mock ups. They represented the trivial, uncertain project overhead items. Although their expenses are considerably low, their actual cost impact can be greater than expected. Hence, they were put in the 3rd priority and should not be ignored, especially when estimating resource is available.

4. Estimate with professional judgment – for the 4th quarter items: low in percentage contribution to total project overhead cost and estimation inaccuracy.

   Items lied in this quarter included testing; and contract conditions. These items were neither significant nor uncertain in nature. To save estimating resources, estimating to these items can be simply relied on the estimators' past experience.

If a total project overhead cost is to be calculated rather than itemized cost, contractors can consider adopting the ANN estimating model as proposed in the Data Analysis Chapter to improve their estimating efficiency.

8.5 GMDH algorithm for Cost Estimation

ANN is advocated as a useful predictive tool and it proves its power again in this study. Although GMDH algorithm is not popular in the construction cost estimation domain, this study evidenced that it is a reliable and accurate tool.

Table 8-1 compares the actual costs against the values predicted by the GMDH model. Most of the predictions fall in line with the actual figures. As indicated in Figure 8-2, the predicted values of the best model found by GMDH network follow closely the actual values. This pattern can be confirmed by the high correlation coefficient of the predicted and the actual values (as explained in the Data Analysis Chapter).

Table 8-1 Summary of the Actual and Predicted Project Overhead Cost Estimated by the GMDH Model

Ref	Actual cost		Predicted cost		Difference
	original	scaled	original	scaled	(absolute value)
1	16,543,000	0.1246	18,002,921	0.1359	1,459,921
2	427,000	0.0000	(357,138)	(0.0061)	784,138
3	25,128,000	0.1909	18,292,674	0.1381	6,835,326
4	16,432,000	0.1237	23,203,464	0.1761	6,771,464
5	9,970,000	0.0738	13,894,445	0.1041	3,924,445
6	13,996,000	0.1049	17,886,188	0.1350	3,890,188
7	35,022,000	0.2674	33,610,109	0.2565	1,411,891
8	50,236,000	0.3850	46,722,293	0.3578	3,513,707
9	11,913,000	0.0888	20,439,453	0.1547	8,526,453
10	40,076,000	0.3065	33,338,865	0.2544	6,737,135
11	82,450,000	0.6340	61,879,354	0.4750	20,570,646
12	42,721,000	0.3269	60,912,153	0.4675	18,191,153
13	65,403,000	0.5022	65,033,886	0.4994	369,114
14	31,260,000	0.2383	30,899,748	0.2355	360,252
15	70,380,000	0.5407	77,402,004	0.5950	7,022,004
16	55,301,000	0.4242	63,778,765	0.4897	8,477,765
17	68,072,000	0.5229	70,092,452	0.5385	2,020,452
18	48,461,000	0.3713	48,411,848	0.3709	49,152
19	48,140,000	0.3688	59,831,176	0.4592	11,691,176
20	78,390,000	0.6026	75,784,024	0.5825	2,605,976
21	45,320,000	0.3470	52,411,252	0.4018	7,091,252
22	4,982,000	0.0352	12,642,949	0.0944	7,660,949
23	47,216,000	0.3617	34,844,411	0.2660	12,371,589
24	128,044,000	0.9864	110,721,329	0.8525	17,322,671
25	83,510,000	0.6422	65,543,494	0.5033	17,966,506
26	9,870,000	0.0730	11,912,983	0.0888	2,042,983
27	33,050,000	0.2522	21,663,565	0.1641	11,386,435
28	38,730,000	0.2961	29,511,519	0.2248	9,218,481
29	16,450,000	0.1239	18,376,594	0.1387	1,926,594
30	23,316,000	0.1769	21,858,636	0.1657	1,457,364
31	44,280,000	0.3390	41,315,820	0.3161	2,964,180
32	20,555,000	0.1556	17,870,412	0.1348	2,684,588
33	5,527,000	0.0394	5,748,824	0.0411	221,824
34	15,549,000	0.1169	17,628,594	0.1330	2,079,594

35	3,580,000	0.0244	12,479,079	0.0932	8,899,079
36	129,800,000	1.0000	119,311,779	0.9189	10,488,221
37	42,790,000	0.3274	27,296,418	0.2077	15,493,582
38	89,020,000	0.6848	64,425,863	0.4947	24,594,137
39	44,631,000	0.3417	63,395,186	0.4867	18,764,186
40	14,977,000	0.1125	11,773,332	0.0877	3,203,668
41	21,520,000	0.1630	13,501,448	0.1011	8,018,552
42	68,000,000	0.5223	49,057,962	0.3759	18,942,038
43	26,520,000	0.2017	33,917,121	0.2589	7,397,121
44	52,460,000	0.4022	73,270,393	0.5630	20,810,393
45	120,273,000	0.9264	115,375,086	0.8885	4,897,914
46	46,000,000	0.3523	55,397,969	0.4249	9,397,969
47	26,093,000	0.1984	24,840,616	0.1887	1,252,384
48	31,180,000	0.2377	31,770,042	0.2423	590,042
49	83,975,000	0.6458	68,971,222	0.5298	15,003,778
50	63,836,000	0.4901	41,561,827	0.3180	22,274,173
51	15,386,000	0.1156	18,215,487	0.1375	2,829,487
52	3,275,000	0.0220	136,186	(0.0022)	3,138,814
53	12,840,000	0.0959	11,203,964	0.0833	1,636,036
54	37,490,000	0.2865	29,576,717	0.2253	7,913,283
55	115,390,000	0.8886	115,276,070	0.8877	113,930
56	10,862,000	0.0807	12,079,047	0.0901	1,217,047
57	22,830,000	0.1732	25,309,027	0.1923	2,479,027
58	51,003,000	0.3909	40,079,383	0.3065	10,923,617
59	35,865,000	0.2739	52,891,464	0.4055	17,026,464
60	56,490,000	0.4333	63,638,580	0.4886	7,148,580
61	62,093,000	0.4767	70,533,963	0.5419	8,440,963
62	41,200,000	0.3152	45,455,590	0.3481	4,255,590
63	85,486,000	0.6575	71,442,854	0.5489	14,043,146
64	118,494,000	0.9126	118,372,328	0.9117	121,672
65	37,317,000	0.2851	50,456,490	0.3867	13,139,490
66	4,337,000	0.0302	12,717,435	0.0950	8,380,435
67	40,196,000	0.3074	27,850,526	0.2120	12,345,474
68	3,541,000	0.0241	13,913,687	0.1042	10,372,687
69	5,742,000	0.0411	4,957,030	0.0350	784,970
70	48,570,000	0.3721	35,137,903	0.2683	13,432,097
71	33,840,000	0.2583	30,587,303	0.2331	3,252,697

As mentioned before, the mathematical principles of GMDH are very complicated. However, the application of the algorithm in ANN modelling with the help of commercial software package is rather straight-forward. Besides, the solution of GMDH is entirely transparent (with the best formula and significant variables identified) to allow the users to conduct further analysis or comparative studies in an easy manner. Also, the self-organizing nature of the network requires very few parameters to be set, thus alleviating the experimentation with different settings. Thus, the GMDH algorithm should gain more popularity among the researchers in the future.

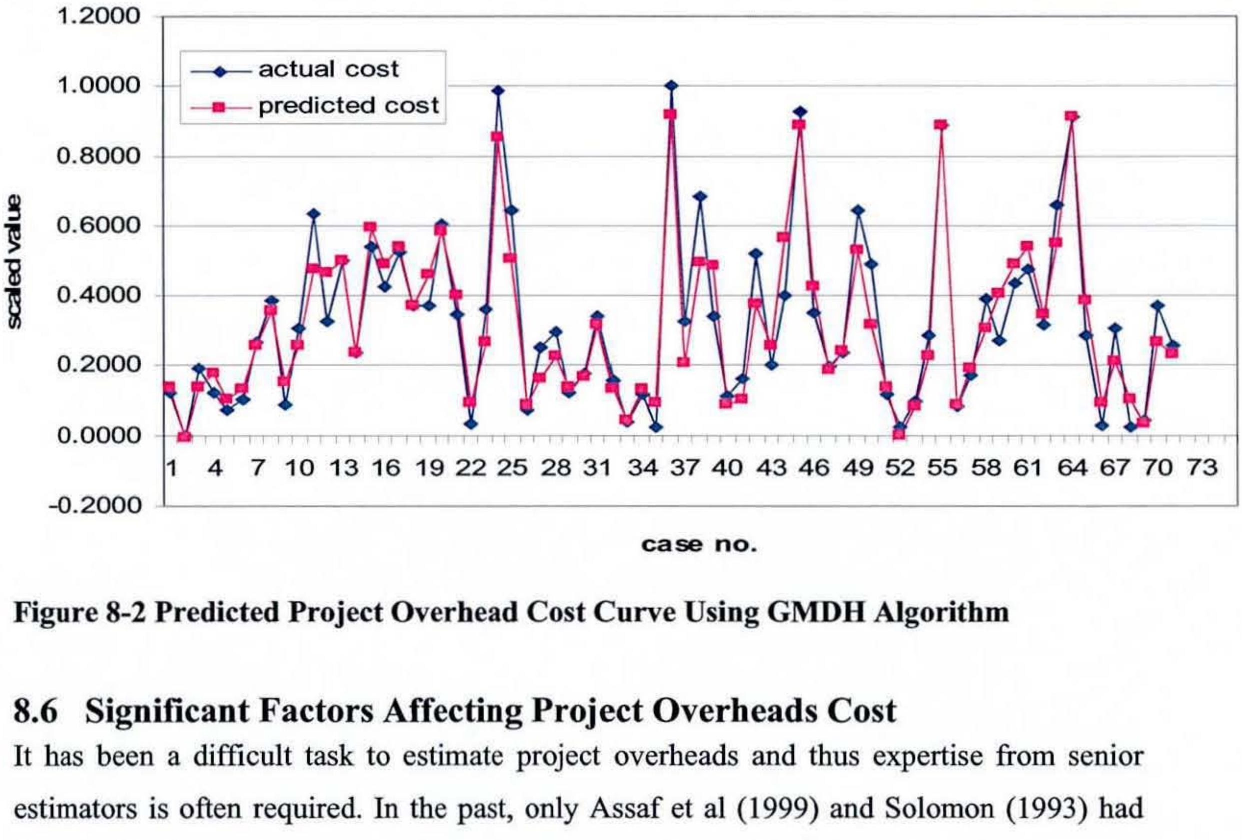

Figure 8-2 Predicted Project Overhead Cost Curve Using GMDH Algorithm

8.6 Significant Factors Affecting Project Overheads Cost

It has been a difficult task to estimate project overheads and thus expertise from senior estimators is often required. In the past, only Assaf et al (1999) and Solomon (1993) had done analysis on the factors affecting project overheads. Details of their findings were examined in the Literature Review Chapter. Solomon proposed the factors based on his observations on consultants' estimates, and Assaf et al suggested the factors by descriptive analysis of the opinions of contractors. However, both of their studies failed to test or validate the proposed factors with real project data. Unlike this study, the proposed factors were tested and validated (using project data) with other ANN models and proved satisfactory predictions.

Through the identification of significant factors from the fourteen input variables, several observations can be highlighted:

- *Project size* and *staff and plants assigned to the project* are two of the five factors that were identified by the regression analysis as significant factors. It can be easily explained as larger project should consume more expense in all project overhead items. Gross floor area of the project is the most direct way to ascertain the "size" of project. On the other hand, as identified in the exploratory study (and past study by Solomon, 1993), around 50% of the project overhead cost goes to the staffing and plant cost, which is the largest attribute compared with other project overhead items. Therefore, if the project assigned with more staff and / plants, the likely expenditure in this area will increase.
- Regarding the procurement nature factor, different procurement scenarios can lead to different impact to project overheads costs. Adopting a simpler procurement system not only saves the management effort or liability from the project owner, but also provides cost saving in project overheads expenditure. Therefore, during the pre-contract stage, project owners should carry out thorough consideration in procurement system selection, taking into account the likely cost savings.
- Another significant factor, *contractor's past relationship with project parties*, reflects the importance of "B2B" business nowadays. In the past, construction work is more of a "production" process. Contractors can survive with the provision of economic and quality works. Under globalization, buyer's market dominates almost every market segment. Project owners are no longer satisfied with quality and cost only. Construction becomes a "service" to the client. Therefore, good relationship with external counterparts is the ultimate asset of a contractor to compete with its competitors for new projects; to grow and survive. As a contractor, management should pay more attention in extending or maintaining good relationship with project owners, designers and consultants. In view of the large contractors in Hong Kong which rely on their project managers to take care of the customer relations, this kind of management strategy should be reviewed. External relations of contractors should be established on a more corporate level and in a more proactive manner.
- The last significant factor, *economic condition during construction stage*, affects the cost of operations of all businesses. Nevertheless, the impact of this factor is often overlooked by researchers. No researchers proposed this attribute as a significant factor affecting project overheads or project cost in the existing literature. Although

the economic environment is an external and uncontrollable factor, the impact from such should be properly considered before developing any business cost model or strategy. As mentioned in the Data Analysis Chapter, this factor is suggested by the focus group. It evidenced the success of incorporating practitioners from the industry to "collaborate" in researches.

- In the past literature, a lot of textbook authors suggest that "complexity" is a significant factor affecting project overheads. However, the results proved that this is not true. Similarly, "location" is another attribute advocated by a lot of them but it hardly exhibits any significance in the ANN modelling.

8.7 Limitations

8.7.1 Source of Data

Although the study involved in this thesis is quite extensive, from exploratory examination of the project overheads nature to the development of cost estimation model, all the views and project data were collected from the contractors in Hong Kong. In the Hong Kong construction industry, there are a lot of customs and traditions related to the Chinese culture, e.g. to send gifts to working partners in Chinese festivals like Mid-Autumn Festive, Chinese New Year, and Tuen Ng Festival (Dragon Boat Festival). Invitations to lunch / dinner are also made by the contractors to the consultants and close partners as a celebration. Other project overhead cost like religious ceremony for the work commencement is also an important event in all projects. All these outgoings are part of the project overhead expenses but do not appear in the cost of other overseas projects. Likewise, project overheads like pension is not required in Hong Kong but common in western countries. Therefore, project overheads analysis based on different countries' data may give slightly different distribution.

Besides, as discussed in the Data Analysis Chapter when finalizing the factor list for the model, the focus group suggested to delete the "social and safety risk" as this is not a concern in Hong Kong. However, as commented in the literature, corruption it is an important concern affecting project performance in developing countries (Rosnaum, 1997). Different cultures in different countries may have some impact to the factors affecting project overheads. However, due to the scope of this project, the cultural impact to project overheads was not assessed.

8.7.2 Sample Size

In this study, seventy-one project cases were collected from the large contractors. In view of the vast amount of projects carried out in Hong Kong, the cases used here only represent a very small sample in the population. Besides, ANN trains better if more data is available. The prediction can be further improved if more data can be collected for training and production. However, due to the sensitivity nature of the cost data, only a small portion of contractors are willing to disclose their internal data for research executed by an external party.

8.7.3 Understate Estimating Inaccuracy

In the Stage 1 survey, it was commented by majority of the respondents (around 77%) that inaccuracies did exist in project overheads estimation, particularly for items like site management salaries, protection to works, site cleaning and site offices. In Stage 2 survey, inaccurate level of each project overhead item was surveyed. Although the surveys were conducted with experienced senior estimators, inaccurate estimation is not an area that the estimators will like to admit to. Hence, the possibility of bias (to understate the level of inaccurate estimation) in the responses cannot be eliminated entirely. In other words, the actual extent of inaccurate estimation may be greater than was found in this research. However, the extent of inaccurate estimation (in real terms) was unable to ascertain here.

8.7.4 Nature of Project

Different project nature associates with different project overhead requirements (as discussed in the analysis of estimating inaccuracy amongst different company nature). For instance, protection to finished works is very important in building projects where lots of finishing trades are found. However, in civil engineering projects, this item is not significant at all. Therefore, the cost data used in the ANN model were all collected from building projects. In other words, the distribution and estimating model for project overheads related to civil engineering projects were not investigated which may produce different results in the significant factors affecting project overhead.

CHAPTER 9

Conclusion and Recommendations

9 Conclusion and Recommendations

Estimating project overheads is one of the estimating tasks that must be undertaken by contractors or consultants. Different past literature suggested different indicators for project overheads percentage, ranging from 6% to 30% of the total project cost. The very few past studies regarding project overheads revealed that most of the contractors estimated project overheads in a detailed manner (i.e. estimated each project overhead item cost, not estimated the total project overhead cost as a percentage of total cost). However, most of them estimated the item cost based on their professional judgment. The studies also indicated that conventional method of building up rates for the estimates is the still most common method in estimating. Despite the advancement of information technology in the last decade, the adoption of sophisticated estimating tool or artificial intelligence in estimating tasks is minimal.

Nevertheless, there are advocates to the importance of project overheads estimation especially when most of the works are sublet to subcontractors. The main area in a tender where the contractor can seek to gain competitive edge is the adjustment in project overheads. Whilst there are enormous amount of researches in the cost estimating and construction management domains, the studies related to project overheads in construction projects is negligible. This put forth the objectives of this research to explore more knowledge about the estimating of project overheads, as well as to establish an efficient, reliable estimating model to estimate project overheads.

Since the existing literature related to project overheads estimation is so limited, this research started from exploring the practices and nature of project overheads estimation before developing the estimating model. Therefore, the methodology was divided into two main parts (exploratory study of project overheads and development of project overheads estimation model) which were further expanded into six phases. Opinion surveys and project data were collected from the Group C contractors in Hong Kong (under the "List of Approved Contractors and Suppliers for carrying out the Public Works maintained by the Works Bureau of the Hong Kong SAR Government) which were very useful for this thesis. All the objectives of this thesis were satisfactorily achieved, and the hypotheses proved. The major findings and areas of concern are summarized in the following sections.

9.1 Summary of Major Findings

9.1.1 Exploratory Study of Project Overheads

The exploratory study of project overheads covered three main areas: project overheads estimating practice of contractors, project overheads estimating accuracy among contractors and distribution pattern of project overhead cost. As said, data collection and surveys were made with large contractors in Hong Kong to ascertain the information about the three areas.

9.1.1.1 Estimating Practice of Project Overheads

Out of one hundred and nineteen number of Group C contractors, forty-nine responses were received in the questionnaire survey on estimating practice of project overheads.

1. More than 90% of the responded contractors estimated project overheads in detail and the rest of them estimated it as a percentage of total value of work.
2. 65% of estimators used their experience and professional judgment to estimate the project overheads. Only 35% of them estimated the items based on cost data. Most of the estimators referred to in-house cost database and / or quotations from subcontractors if they estimated the project overheads based on cost data.
3. Around 65% of the respondents thought that setting profit margin was the most important estimating task. Only 9% thought that estimating project overheads was the most important. However, all of them assigned only senior estimators to estimate project overhead costs. Here indicates a misalignment between the perceived importance of project overheads estimation and the actual resource allocated to it.
4. Twenty-three project overhead items were shortlist by the respondents as important items that must be included for study or pricing, e.g. site management, safety precautions, etc. The list was used in the subsequent study of estimation inaccuracy of project overhead items.

9.1.1.2 Estimating Inaccuracy of Project Overhead Items

Out of the one hundred and nineteen number of Group C contractors, forty responses were received in the opinion survey on estimating inaccuracy of project overheads.

1. Around 40% of the respondents commented that over- or under-estimation of project overheads was common in their estimates. For those who commented that under-estimation was likely, majority of them said that the shortfall could not be recovered or could only be partly recovered by prolongation claims.

2. Among the twenty-three project overhead items, the top three items that were most likely to be over-estimated (in descending order) are : protection of finished works; cleaning and removal of rubbish; and site offices, stores and latrines.
3. The items that were most likely to be under-estimated (in descending order) were : protection of finished works; cleaning and removal of rubbish; and restrictions to noise and nuisance.
4. The rank agreement of relative rankings of over-estimation, under-estimation and inaccurate estimation among different types of companies was low. This implied that different types of projects (building or civil engineering works) carried out by different types of company demanded different requirements in project overheads. Therefore, attention should be draw when sampling project data in the study of project overheads.

9.1.1.3 Distribution of Project Overhead Costs

Project overheads cost breakdown from twenty completed building projects were collected from the Group C contractors. Observations from the data included:

1. Project overhead cost accounted for 11% to 19% of the total project cost.
2. There were seven items accounting for more than 80% of the total project cost. They are listed in descending order : site management (35.9%), mechanical plants (12.4%), cleaning and removal of rubbish (10%), insurances and surety bond (7.4%), setting out (6.2%), temporary works e.g. hoardings, temporary roads, signboard (5.9%), and scaffolding (5.9%). Staffing and plants on site accounted for almost 50% of the total project overhead cost.
3. Misalignment was observed between the average contribution of project overhead cost and the relative estimating accuracy of the item. In other words, some items were high in cost impact but low in estimating accuracy or vice versa.
4. A decision-making matrix model was devised to allocate the estimating resources more efficiently across different project overhead items (details of the model can be referred to Section 8.4). Project overhead items were proposed to be divided into four categories for prioritizing estimating resources (in descending order) : 1) high in both potential cost and likelihood of inaccuracy in estimation; 2) high in potential cost but low in likelihood of estimation inaccuracy; 3) low in potential cost but high in likelihood of estimation inaccuracy; 4) low in both potential cost and likelihood of estimation inaccuracy.

5. The proposed matrix is helpful to direct estimating resources in case if itemized project overhead cost is required to estimate.

9.1.2 Development of Project Overheads Estimating Model

Seventy-one project data sets were collected from large contractors in Hong Kong. The projects were all constructed during 1997 to 2004.

Extraction of Inputs

1. From the literature review and focus group discussion, forty-seven factors were shortlist as factors affecting project overheads. They were used for opinion survey to analyze their respective significance.
2. Exploratory factor analysis of the opinion survey extracted fourteen factors as principal factors affecting project overheads. These fourteen factors were used as input variables to build the artificial neural network estimating model for project overheads.

ANN Model Prediction

3. GMDH network was chosen to predict project overhead cost as it is self-organized in nature, and being able to automatically identify the significant factors and the best formula after training.
4. The best model generated by the GMDH network using FCPSE selection criterion was :

   $Y = -0.92*X_3 + 0.21*X_4 + 0.17*X_5 + 3.7*X_3*X_5 - 0.44*X_1*X_3$

   Where Y = (OUT_COST-427,000)/129,373,000; X_1 = (ECON-85.32)/26.43; X_3 = (SIZE-3,500)/179,500; X_4 = (RELAT-1)/4 and X_5 = (STA_PLT-1)/4
5. The R squared of the best GMDH model was 0.8818 for the training set. The correlation coefficient was 0.939. This represents a strong relationship between the predicted and actual values.

Significant Variables for Project Overheads Estimation

6. Five variables were identified as significant after three stages of re-training to remove the non-significant factors from the inputs. These five variables were: *economic condition during construction stage, procurement nature, project size, contractor's past relationship with project parties* and *extent of staff and plants assigned to the project.*

Model Validation

7. The best GMDH model was validated by eight project cases (production data) which were not used in training the model before. The prediction was much better than that in the training set, with R squared = 0.9754. The correlation coefficient was very close to 1 (0.9885).
8. The GMDH model was cross-validated by dividing the sample cases into nine sub-groups; with training and production repeated eight times (each time using eight sets for training and one set for production in turn). The results of cross-validation were satisfactory with R squared of all trials exhibited above 0.7 (except for one trial in the production set produced only 0.493). The highest R squared produced across all the trials was 0.9883.
9. GMDH network outperformed the multiple linear regression model. The latter produced only 0.7961 in R squared value and identified *project size* and *extent of staff and plants assigned to the project* as significant variables.
10. The five significant variables identified by the GMDH network were experimented with multi-layer feed-forward with backpropagation network and general regression neural network. The results were also satisfactory, with R squared above 0.65 and 0.5 for MLFF and GRNN respectively (for all trials in training, testing and production). This verified the reliability of using the five variables to predict project overhead cost.
11. After the various testing and validations of the model and variables, GMDH network was proved to be a suitable and reliable method to predict project overheads.
12. A focus group of seven senior quantity surveyors from large contractors unanimously agreed that the ANN estimating model was efficient and accurate. Majority of the group thought that the model was applicable to contractors / consultants / project owners. Only one member held an opposite view concerning resistance from practitioners due to the lack of knowledge in ANN modelling. Nevertheless, all of the members conceded that the model or equation did provide a satisfactory estimate with minimal estimating effort. The technique should worth development into a tailored model in contractors or project consultants in order to enhance the estimating accuracy further.

9.2 Achievement of Research Objectives

9.2.1 First Objective

The findings above indicated satisfactory achievement of the research objectives. Referring to the Introduction Chapter, the first objective was "to explore the nature of project overheads and the associated estimation methods in use by practitioners in the Hong Kong construction industry". This was achieved by the first stage survey; which had successfully collected and analyzed forty-nine responses from large contractors in Hong Kong (out of one hundred and nineteen in the registration list). The survey results depicted the common practice of estimating project overheads which was based on professional judgment. Although majority thought project overheads were not important compared with other estimating tasks, project overhead estimation demanded the most expertise and time from the senior estimators.

9.2.2 Second Objective

The second objective was "to identify critical project overhead items that possess higher risk of genuine under- or over-estimation by tenderers in Hong Kong". It was achieved by the first and second stage surveys. The first stage survey identified a representative list of project overhead items which covered twenty-three items that had financial impact. The list was then used in the questionnaire of the second stage survey to study the likelihood of under- or over-estimation. This opinion survey identified items like protecting finished works, cleaning and removal of rubbish had the highest likelihood of under- or over-estimation. Besides, different types of company (civil / building / both) responded differently in terms of the likelihood of inaccurate estimation. This implied different project overhead requirements generates different difficulty in estimation.

9.2.3 Third Objective

The achievement of the third objective was evidenced in the project overhead cost data collected from the large contractors in the last part of the exploratory study. From the analysis of the project overhead cost breakdown of twenty projects collected, the overall pattern of project expenses amongst projects in Hong Kong was identified. The cost data collected was extremely sensitive as it was the internal cost data of the project; and thus study of this nature has not been done in the past. Overall percentage and distribution of the project overhead cost was analyzed. The project overhead cost spent was within a range of 11% to 10% of the total project cost, with site management and mechanical plants already accounted for 48% of the total project overhead cost. The rest of the 52% was attributed by nineteen items like cleaning, insurances, etc.

9.2.4 Fourth Objective

The fourth and the fifth objectives were also satisfactorily achieved. Having identified the input factors by exploratory factor analysis, the ANN cost estimating model was successfully developed with GMDH network using FCPSE selection criterion. Based on sixty-three training data set, the model could predict the project overhead cost with an R squared of 0.8818. Prediction on 8 sets of new production data even produced 0.9754 R squared value, indicating very accurate prediction (R squared is 1 in case of perfect prediction). The ANN model was efficient with high accuracy; and thus fulfillment of the fourth objective (to develop an estimating model to predict project overheads in an accurate and efficient manner using artificial intelligence).

9.2.5 Fifth Objective

Besides predicting the project overhead cost, the GMDH network also identified five variables (out of fourteen principal factors extracted by exploratory factor analysis) as significant variables affecting project overheads. They were 1) *economic condition during construction stage,* 2) *procurement nature,* 3) *project size,* 4) *contractor's past relationship with project parties* and 5) *extent of staff and plants assigned to project.* These five variables had proved to be significant variables to predict project overheads. Validation by MLFF and GRNN also produced acceptable results (R squared above 0.65 and 0.5 respective for all training, testing and production trials). Thus, the fifth objective, "to identify significant factors affecting project overhead expenses in Hong Kong" was satisfied.

9.3 Testing of Hypotheses

There were three hypotheses described in the Introduction Chapter, including:

1. There is substantial inaccuracy in the estimation of project overheads by the large contractors in Hong Kong.
2. Project overhead cost can be estimated more accurately by artificial intelligent modelling than conventional methods such as multiple regression.
3. Project overhead cost can be satisfactorily predicted by simple parameters.

Testing of these hypotheses was also satisfactorily accomplished.

9.3.1 First Hypothesis

The first hypothesis was proved by the Stage 2 opinion survey on the likelihood of inaccurate estimation of project overheads. It was evidenced that the likelihood of inaccurate estimation was substantial. Ten out of twenty-three items were commented as likely to be inaccurately estimated by more than 45% of the respondents. From the analysis, likelihood of under-estimation was slightly more common than over-estimation, probably due to the intuitive tendering decision under the competitive bidding environment.

9.3.2 Second Hypothesis

From the validation results of the GMDH model, it was proved that ANN model outperformed the traditional multiple linear regression method. The regression equation gave an R squared value of 0.7961 whereas the GMDH network provided much better results (R squared values were 0.8818 and 0.9754 in training and production sets respectively). In other words, the second hypothesis was confirmed – project overhead cost can be estimated more accurately by GMDH network than multiple regression.

9.3.3 Third Hypothesis

The third hypothesis was proved as a result of the successful identification of the significant variables affecting project overheads by the GMDH model. The 5 variables namely: 1) *economic condition during construction stage,* 2) *procurement nature,* 3) *project size,* 4) *contractor's past relationship with project parties* and 5) *extent of staff and plants assigned to project* could predict project overheads satisfactorily using GMDH network (R squared = 0.8818 in training set and 0.9754 in production set). In case of validation using other ANN architectures, acceptable prediction was also obtained (R squared were above 0.6 and 0.5 in training, testing and production sets for MLFF and GRNN networks respectively).

On the other hand, in terms of data collection, the five significant variables were either measured by opinion in likert scale (in case of *procurement nature, extent of staff and plants assigned* and *contractor's past relationship with project parties*) or by easily accessible ratio data (in case of gross floor area and government indices for *economic condition during construction stage*). The measurement of these parameters is rather straight-forward and thus fulfills the hypothesis.

9.4 Contribution to Knowledge

A few past studies suggested that project overheads estimation should be the most important estimating task when most of measured work is sublet as offsetting bids will be obtained from subcontractors. In other words, project overheads are critical to the success in bidding. Nevertheless, despite the abundant researches in construction cost estimation, the studies in relation to project overheads are very scarce. The main reason is probably due to the difficulty to collect the sensitive cost data from contractors. Unlike cost of measured works which can be built up from published unit cost data, project overheads involved a lot of internal cost like salaries of staff, insurance premium, etc. As a result, the existing knowledge in project overheads remains merely on the theoretical level.

9.4.1 Reveal the Practice of Project Overheads Estimation

This research started with an exploratory study to reveal the practice of project overheads estimation amongst the large contractors in Hong Kong. It verified the estimating resources, difficulty and inaccuracy involved in estimating project overheads. The results not only prove the importance of project overheads, but also provide necessary data for developing an efficient and accurate model to estimate project overheads.

9.4.2 Improve the Estimation of Itemized Project Overhead Costs

In terms of improving the estimating efficiency and accuracy of project overheads, both the overall project overhead estimation and itemized project overhead costs were considered and contributed. Unlike the few past studies related to itemized project overhead costs which only relied on tender estimates from consultant quantity surveying firms, the cost data used in this research was the actual expenditure of contractors. Such analysis avoids all possible errors and bias of the estimators, and thus produces more realistic and meaningful findings. By studying the actual expenditure of project overheads, priority of estimating resources allocation was suggested according to the likely cost and estimating accuracy of each item. This can improve the utilization of estimating resources within the limited tendering period. The savings in contractor's bidding process can in turn reflect in their tender bids which benefit the project owners.

9.4.3 Improve the Estimation of Total Project Overhead Cost

If project overhead cost is to be estimated as a total sum, this research also provided an effective solution by using artificial neural network. The research demonstrated a satisfactory model to estimate total project overhead cost using real project data. Adopting

GMDH technique in the ANN model, the prediction of project overheads was proved to be accurate and time-saving. In terms of industrial application, the technique should produce more accurate predictions if project data was coming from the single company environment with a large database of past data. Practitioners agreed that the calculation was simple regardless of the complicated mathematics behind the ANN principles. The predicted project overhead cost can act as a good indication for estimators to check their estimates, as well as an estimate for project owners' budget in the design stage.

9.4.4 Identification of Significant Parameters Affecting Project Overhead Costs

After three stages of training and retraining, five significant variables were scrutinized, namely: 1) *project size*; 2) *extent of staff and plants assigned to project*; 3) *economic condition during construction stage*, 4) *procurement nature*; and 5) *contractor's past relationship with project parties*. These were identified as the most significant variables by most selection criteria. When feeding these five variables to the GMDH model to train again, five out of six selection criteria selected all of them to build the best model with satisfactory accuracy produced. Due to this reason, the input variables could not be reduced further and the model using five variables as input was regarded as the best and simplest one modelled by the GMDH algorithm.

The first factor, *project size*, is one of the commonly cited factors affecting project cost including project overhead cost. Larger construction project certainly requires more resources and management effort. Therefore, project overhead is likely to be dependent on project size.

Staff and plant assigned to project is in fact the largest cost centre of project overhead cost. From the current exploratory study of project overheads and similar study by Solomon (1993), management staff and mechanical plants accounted for around 50% of the total project overheads cost. If the extent of staff and plants assigned to the site is high, the overall project overhead cost should undoubtedly increase.

The construction market has a close association with the economy. When the economy is prosperous, more construction works will be executed (especially in the private sector) as evidenced by the trends of the value of construction works and the consumer price index

(Table 7-23). Under that situation, construction cost including project overhead costs will increase as contractors have to compete for resources. Therefore, it is reasonable to see *economic condition during construction stage* being as one of the significant factors affecting project overheads cost.

For *procurement*, different arrangements can lead to different impacts on project overheads cost. For instance, if the contractor is procured under a design and build contract, it can have higher autonomy to determine the site accommodation quality, adopt simplified procedures for sample or drawings submission, etc. If a developer's in-house contract is used instead of standard form of contract, more onerous terms which are in favour of the developer are usually found, e.g. provision of non-standard insurance package to works, lesser entitlement to loss and expenses claim, etc. Therefore, the impact of this factor to project overheads should not be understated.

The last factor, *contractor's past relationship with project parties*, revealed the importance of good business relationship. The three parties involved in this factor are the major partners who exhibit either full or partial influence on the work of the contractor. Imagine if the contractor maintained good relationships with the client, designer and even consultant QS in the past, approval of works and submissions in the current project is likely to be smoother. In that case, contractor's expenses in project overheads like drawings and mock-up can be very much reduced.

When using the proposed GMDH model to predict project overhead cost in practice, contractors should collect, as much as possible, the past project data related to the five input variables and output to build an in-house database. With these training and production data from the contractor's database, the best polynomial equation generated by the GMDH is likely to be different from the one presented in the Data Analysis Chapter. Therefore, training and testing should be done by all the six selection criteria sequentially in order to determine the best model with the highest predictive accuracy. Estimation of the project overhead cost for new projects can then be executed by the chosen selection criteria using the GMDH algorithm.

To sum up, this research is build upon empirical cost data from real projects which is not found in other related studies. Due to this reason, the applicability of the proposed model and research findings is higher than other similar studies.

9.5 Future Study

This research provides a thorough review and analysis of project overheads estimation in construction projects in Hong Kong. Further researches can be done to expand the body of knowledge in project overheads estimation. Some of the possible areas are discussed below.

9.5.1 Larger Sample Size

In statistical terms, larger sample size can lead to a better representation of the population by the sample results. Especially when adopting AI models, more samples can generate more accurate results. Therefore, similar research can be done using larger sample size to validate the findings obtained from this research.

9.5.2 Wider Spectrum of Sample

As mentioned in earlier section, different countries and cultures may have different impact to the project overheads expenditure. It is worthwhile to collect project data from other countries to verify the generalizability of the proposed model as well as the significant factors. Furthermore, different project nature (e.g. civil engineering projects) can also be sampled to verify if there is any difference in the significant factors when compared with those identified in the building projects.

9.5.3 ANN Estimating Model Using Other Algorithms

GMDH algorithm was used in the model architecture for the cost estimating model in this research. The rationale was based on the self-organizing and transparent nature of the GMDH algorithm. Although the significant factors and cost data were also validated in other ANN algorithms, no attempt was made to compare the performance of different ANN architectures. Researchers interested in ANN may conduct comparative study further to search for the "best ANN model" to estimate project overheads.

9.5.4 Testing of Significant Factors by Other Models

The significant factors identified by the GMDH network are useful parameters to strive for cost savings in overheads. However, due to the scope and nature of this study, validation of the significant factors was limited to the tests by MLFF network, GRNN and linear regression method. More comprehensive testing of these factors could be done to verify the significance level of them. Through such study, better understanding of the link between these factors to project overhead expenditure could also be established.

9.5.5 GMDH as an Alternative Algorithm

As ANN is gaining popularity in construction management researches, most researchers only adopt their familiar architectures like multi-layered feedforward and general regression models in their studies. In fact, more applications of the GMDH algorithm can be made to fully exploit its power in forecasting, modelling and optimization.

CHAPTER 10

References and Bibliography

10 References and Bibliography

10.1 References

1. Adeli, H. and Wu, M. (1998). "Regularization Neural Network for Construction Cost Estimation." *Journal of Construction Engineering and Management*, Jan/Feb, 18 – 24.

2. Adrian, J. J. (1982). *Construction Estimating*, Reston, VA: Reston Publishing Co.

3. Afifi, A., Clark, V.A., May, S. (2004). "Factor Analysis." *Computer-Aided Multivariate Analysis*, 4th Ed., Boca Raton, Fla.: Chapman & Hall/CRC.

4. Ahmad, I. U. and Ommi, P. K. (1996). "Parametric Estimating: An Object-oriented Approach." *Proc., 3rd Congress on Computing in Civil Engineering*, American Society of Civil Engineers, 254 – 260.

5. Ahmad, I. U. and Rahman, S. T. (1994). "Refinement of Cost Estimates with Artificial Neural Nets." *Proc., 1st Congress on Computing in Civil Engineering*, American Society of Civil Engineers, 1373 – 1380.

4. Akintoye, A. and Fitzgerald, E. (2000). "A survey of current cost estimating practices in the UK." *Construction Management and Economics*, 18, 161 – 172.

6. Al Khalil, M., Assaf, S., Rahman, W.A., Asfoor, M. (1999). "A Conceptual Cost Estimating Model for Water Reservoirs." Cost Engineering, 41(5), 38-43.

7. Ashworth, A. and Skitmore, R.M. (1982). "Accuracy in Estimating." *Occasional Paper no. 27, Chartered Institute of Building*, Ascot: Chartered Institute of Building, London, U.K.

8. Assaf, S.A., Bubshait, A.A., Atiyah, S. and Al-Shahri, M. (1999). "Project Overhead Costs in Saudi Arabia." *Cost Engineering*, 41(4), Apr., 33 -38.

9. Bartholomew, S. H. (2000). *Estimating and Bidding for Heavy Construction*, Upper Saddle River, N.J. : Prentice Hall, 43-46, 301-315.

10. Bhokha, S. and Ogunlana, S.O. (1999). "Application of Artificial Neural Network to Forecast Construction Duration of Buildings at the Predesign Stage." *Engineering, Construction and Architectural Management*, 6(2), 133-144.

11. Boussabaine, A.H. (1996). "The Use of Artificial Neural Networks in Construction Management: A Review." *Construction Management and Economics*, 14, 427-436.

12. Boussabaine, A.H and Elhag T.M.S. (1999). "Tender Price Estimation Using ANN Methods." *EPSRC Research Report No. 3*, University of Liverpool, Liverpool, U.K.

13. Boussabaine, A.H. and Kaka, A.P. (1998). "A Neural Networks Approach for Cost Flow Forecasting." *Construction Management and Economics*, 16, 471-479.

14. Brook, M. (1998). *Estimating and Tendering for Construction Work*, 2nd Ed., Oxford; Boston: Butterworth-Heinemann, 189 – 203.

15. Census and Statistics Department, Hong Kong SAR Government. (2005). "Principal Statistics for All Building and Civil Engineering Establishments & Consumer Price Indices." Available from <http://www.info.gov.hk/censtatd/eng/hkstat/index1.html> (3 October 2005).

16. The Chartered Institute of Building. (1997). "Project Overheads." *Code of Estimating Practice*, 6th Ed., Addison Wesley Longman Ltd.

17. Chan, A.P.C., Chan, D.W.M., Chiang, Y.H., Tang, B.S., Chan, E.H.W. (2004). "Exploring Critical Success Factors for Partnering in Construction Projects." *Journal of Construction Engineering and Management*, Mar/Apr, 2004, 188-198.

18. Chan, N.H. and Genovese, C.R. (2001). "A Comparison of Linear and Nonlinear Statistical Techniques in Performance Attribution." *IEEE Transactions on Neural Networks*, 12, 922-928.

19. Chan, D.W.M. and Kumaraswamy, M.M. (1998). "Contributors to Construction Delays." *Construction Management and Economics*, 16(1), 17 – 29.

20. Chan, D.W.M. and Kumaraswamy, M.M. (1999). "Forecasting Construction Durations for Public Housing Projects: A Hong Kong Perspective." *Building and Environment*, 34(5), 633-646.

21. Charlton, M. (2001). "Risk Allocation in Construction Contracts in East Asia." The Chartered Institute of Building (HK) Journal, 3, 2001, 13-17.

22. Cheong, P.F. (1991). "Accuracy of Design Stage Cost Estimating." MSc Thesis, National University of Singapore, Singapore.

23. Cooke, B. (1984). *Contract Planning and Contractual Procedures*, 2nd Ed., Basingstoke, Hants: Macmillan Educational, 50 – 58.

24. Dagostino, F. and Feigenbaum, L. (2003). *Estimating in Building Construction*, 6th Ed., Upper Saddle River, N.J.: Prentice Hall, 51-62.

25. Datta, S. and Mukherjee, S.K. (2001). "Developing a Risk Management Matrix for Effective Project Planning – An Empirical Study." *Project Management Journal*, 32(2), 45-57.

26. Davis, J. J. (1998). "Parametric Estimates Made Easy." *Proc., 42nd Annual Meeting of AACE International,* EST.01.1 – 01.5.

27. Dissanayaka, S.M. and Kumaraswamy, M.M. (1999). "Evaluation of Factors Affecting Time and Cost Performance in Hong Kong Building Projects." *Engineering, Construction and Architectural Management,* 6(3), 287-298.

28. Elhag, T.M.S. and Boussabaine, A.H. (1998). "Factors Affecting Cost and Duration of Construction Projects." *EPSRC Research Report Phase (1),* University of Liverpool, Liverpool, U.K.

29. Elhag, T.M.S. and Boussabaine, A.H. (1998). "Statistical Analysis and Cost Models Development." *EPSRC Research Report No. (2),* EPSRC Research Grant ref: GR/K/85001, University of Liverpool, Liverpool, U.K.

30. Elkington, P and Smallman, C. (2002). "Managing Project Risks: A Case Study From the Utilities Sector." *International Journal of Project Management,* 20(1), 49-57.

31. Emsley, M.W., Lowe, D.J., Duff, A.R., Harding, A. and Hickson, A. (2002). "Data Modelling and the Application of a Neural Network Approach to the Prediction of Total Construction Costs." *Construction Management and Economics,* 20(6), September, 465-472.

32. Environment, Transport and Works Bureau, Hong Kong SAR Government (2005). "List of Approved Contractors and Suppliers for Public Works." Available from <http://www.etwb.gov.hk/consultants_and_contractors/contractors/approve_contractors/index.aspx?lstCategory=All&lstGroup=GroupC&nodeID=622> (Jul 31, 2005).

33. Fang, D.P., Xie, F., Huang, X.Y. and Li, H. (2004). "Factor Analysis-based Studies on Construction Workplace Safety Management in China." *International Journal of Project Management,* 22, 43-49.

34. Fayek, A., Ghoshal, I., & AbouRizk, S. (1999). "A survey of the bidding practices of Canadian civil engineering construction contractors." *Canadian Journal of Civil Engineering, 26,* 13 – 25.

35. Fayek, A., Young, D.M., Duffield, C.F. (1998). "Survey of Tendering Practices in the Australian Construction Industry." *Engineering Management Journal,* 10(4), 29-34.

36. Fellows, R. and Liu, A. (2003). "Data Collection." *Research Methods for Construction,* 2nd Ed., Oxford; Malden, MA: Blackwell Publishing, 137-160.

37. Gallant, S.I. (1993). "Neural Networks At Work." *IEEE Spectrum*, Jun., 26-32.

38. Garza, J. M. de la and Rouhana, K. G. (1995). "Neural Networks Versus Parameter-Based Applications in Cost Estimating." *Cost Engineering*, 37(2), 14 – 18.

39. Geddes, S. (1996). *Estimating for Building and Civil Engineering Works*, 9th Ed., Oxford; Boston: Butterworth-Heinemann, 36-47.

40. George, D. and Mallery, P. (1999). "Factor Analysis." *SPSS for Windows Step by Step: A Simple Guide and Reference*, USA : Allyn and Bacon, 231-246.

41. Ghosh, S. and Jintanapakanont, J. (2004). "Identifying and Assessing the Critical Risk Factors in an Underground Rail Project in Thailand: A Factor Analysis Approach." *International Journal of Project Management*, 22, 633-643.

42. Gorsuch, R.L. (1983). *Factor Analysis,* 2nd Ed., Hillsdale, N.J.: L. Erlbaum Associates, 1–6, 175–212, 369-381.

43. Grimm, L.G. and Yarnold, P.R. (1995). "Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis." *Reading and Understanding Multivariate Statistics,* Washington, D.C.: American Psychological Association, 99-135.

44. Gunner, J. and Skitmore, M. (1999). "Comparative Analysis of Pre-bid Forecasting of Building Prices Based on Singapore Data." *Construction Management and Economics,* 17, 635-646.

45. Hall, D. and Hulett, D. (2002). "Universal Risk Project Risk Project Final Report." Available from <http://www.risksig.com/articles/UR%20Project%20Report.doc> (Feb. 20, 2002).

46. Hegazy, T. and Ayed, A. (1998). Neural Network Model for Parametric Cost Estimation of Highway Projects." *Journal of Construction Engineering and Management*, May/Jun., 210-218.

47. Hegazy, T. and Moselhi, O. (1995). "Elements of Cost Estimation: A Survey in Canada and the United States." *Cost Engineering*, 37(5), 27 – 33.

48. Holm, L., Schaufelberger, J. E., Griffin, D., Cole, T. (2005). *Construction Cost Estimating – Process and Practices*, Upper Saddle River, N.J.: Pearson/Prentice Hall, 6-13, 169.

49. Ivakhnenko, G. and Cerda R.H. (2002). "Inductive Self-Organizing GMDH Algorithms for Complex Systems Modeling and Forecasting." Available from <http://www.gmdh.net/GMDH.ch.files/frame.html> (May 15, 2006)

50. Ivakhnenko, A. G. and Ivakhnenko, G. A. (1995). "The Review of Problems Solvable by Algorithms of the Group Method of Data Handling (GMDH)." *Pattern Recognition and Image Analysis*, 5(4), 527-535.

51. Jensen, P. (1993). "Cost-efficient Programming of Road Projects Using a Statistical Appraisal Method." *Transportation Research Record*, 1400, 18-26.

52. Jupp, B.C. and McMillan, V. (1981). "The Reliability of Cost Data." *Quantity Surveyors Research and Development Forum*, 16 June 1981, South Bank Polytechnic, London, U.K.

53. Kaming, P.F., Olomolaiye, P.O., Holt, G.D. and Harris, F.C. (1997). "Factors Influencing Construction Time and Cost Overruns on High-Rise Projects in Indonesia." *Construction Management and Economics*, 15, 83-94.

54. Khosrowshahi, F. (1997). "The Optimum Project Duration and Cost Curve for Hong Kong Public Housing Projects." *Engineering, Construction and Architectural Management*, 4(4), 249-269.

55. Kim, D.W. and Park, G.T. (2006). "A Design of EA-based Self-Organized Polynomial Neural Networks Using Evolutionary Algorithm for Nonlinear System Modeling." Available from <http://www.gmdh.net> (May 15, 2006).

56. Kim, K.J. and Han, I. (2003). "Application of a Hybrid Genetic Algorithm and Neural Network Approach in Activity-Based Costing." *Expert Systems with Applications*, 24, 73-77.

57. Kometa, S.T., Olomolaiye, P.O. and Harris, F.C. (1995). "Quantifying Client-generated Risk by Project Consultants." *Construction Management and Economics*, 13(2), 137-147.

58. Kondo, T. and Pandya, A. S. (2000). "GMDH-type Neural Networks with a Feedback Loop and Their Application to the Identification of Large-spatial Air Pollution Patterns." *Proc., 39th SICE Annual Conference International Session Papers*, 112A-4, 1-6.

59. Kumaraswamy, M.M. and Chan, D.W.M. (1998). "Contributors to Construction Delays." *Construction Management and Economics*, 16, 17-29.

60. Kwakye, A. A. (1994). *Understanding Tendering and Estimating*, Aldershot, Hampshire, England: Gower Pub., 252-272.

61. Lattin, J., Carroll, J.D. and Green, P.E. (2003). "Principal Components Analysis" and "Exploratory Factor Analysis". *Analysing Multivariate Data,* Pacific Grove, CA: Thomson Learning, 108-165.

62. Li, H. (1995). "Neural Networks for Construction Cost Estimation." *Building Research and Information*, 23(5), 279 – 284.

63. Ling, Y.Y. and Boo, J.H.S. (2001). "Improving the Accuracy of Approximate Estimates of Building Projects." *Building Research and Information,* 29(4), 312-318.

64. Lowe, D. and Skitmore, M. (1994). "Experiential Learning in Cost Estimating." *Construction Management and Economics*, 12, 423-431.

65. Mansfield, G. L. (1983), Bidding and Estimating Procedures for Construction, 187 – 214.

66. Mason, A. K. and Kahn Sr., D. J. (1997). "Estimating Costs with Fuzzy Logic." *Proc., 41st Annual Meeting of AACE International*, EST.03.1 – 03.6.

67. Minsky, M. and Pappert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*, Cambridge, Mass.: MIT Press.

68. Morrison, N. (1984). "The Accuracy of Quantity Surveyors' Cost Estimating." *Construction Management and Economics*, 2, 57-75.

69. Moselhi, O. (1997). "Risk Assessment and Contingency Estimating." *Proc., 41st Annual Meeting of AACE International,* D&RM/A.06.1-6.

70. Moselhi, O. (1998). "Estimating the Cost of Change Orders." *Proc., 42nd Annual Meeting of AACE International*, EST.06.1 – 06.5.

71. Moselhi, O., Hegazy, T. and Fazio, P. (1991). "Neural Networks As Tools in Construction." *Journal of Construction Engineering and Management*, 117(4), 606-623.

72. Moselhi, O. and Siqueira, I. (1998). "Neural Network for Cost Engineering of Structural Steel Buildings." *Proc., 42nd Annual Meeting of AACE International,* IT/IM.06.1 – 06.4.

73. Nelson, C., Powell, D, and Federle, M. O. (1998). "Computer Assistance in Conceptual Estimating for Buildings." *Proc., 42nd Annual Meeting of AACE International*, IT/IM.04.1 – 04.5.

74. Nkado, R.N. (1995). "Construction Time-Influencing Factors. The Contractor's Perspective." *Construction Management and Economics*, 13(1), 81-89.

75. Norusis, M.K. (1993). "Factor Analysis." *SPSS for Windows, Professional Statistics, Release 6.0.* Statistical Package for Social Sciences, Inc., Chicago.

76. Ogunlana, S. and Thorpe, A. (1991). "The Nature of Estimating Accuracy: Developing Correct Associations." *Building and Environment*, 26(2), 77-86.

77. Oh, S.K. and Pedrycz, W. (2002). "The Design of Self-Organizing Polynomial Neural Networks." *Information Sciences*, 141, 237-258.

78. Okpala, D.C. and Aniekwu, A.N. 1988. Causes of High Costs of Construction in Nigeria. *Journal of Construction Engineering and Management,* 114 (2), 233 – 244.

79. Ostwald, P. F. (2001). "Estimating Methods." *Construction Cost Analysis and Estimating*, Upper Saddle River., N.J.: Prentice Hall, 190-227.

80. Oteifa, S. and Baldwin, A. (1991). "Estimators' Tasks and Computer-aided Estimating Systems: A Survey of FCEC Member Companies." *Construction Management and Economics*, 9, 543–552.

81. Park, W.R. (1992). "Risk and Uncertainty." *Construction Bidding: Strategic Pricing for Profit*, 2nd Ed., John Wiley & Sons, 161-175.

82. Peurifoy, RL and Oberlender, GD (2002) "Overhead." *Estimating Construction Cost*, 5th Ed., McGraw-Hill.

83. Phaobunjong, K. and Popescu, C.M. (2003). "Parametric Cost Estimating Model for Buildings." *Proc., 2003 Annual Meeting of AACE International*, EST.13.1- 13.12.

84. Poon, A.C.H. (2003). "Risk Allocation – An Equitable or Realistic Approach for the Public Sector of Hong Kong? The West Rail Experience." *Hong Kong Institute of Surveyors Journal*, Dec, 2003, 38-48.

85. Popham, K. (1996). "Cost Estimating Using Historical Costs." *Proc., 40th Annual Meeting of AACE International*, EST.14.1-14.5.

86. Pratt, D. (2004). Fundamentals of Construction Estimating, 2nd Ed., Thomas Delmar Learning.

87. Rosenblatt, F. (1958). "The Perception: A Probabilistic Model for Information Storage and Organisation in the Brain." *Psychological Review*, 65, 386 – 408.

88. RIBA Enterprises, National Building Specification, Preliminaries and General Conditions. Available from <http://www.thenbs.com/technical/preliminariesGeneralConditions.asp> (February 16, 2006).

89. Rosenbaum, D.B. (1997). "Risk Questions Hamper Deals." *Engineering News Record*, 21, 11.

90. The Royal Institute of Chartered Surveyors (Hong Kong Branch) and the Building Contractors' Association of Hong Kong. (1979). *Hong Kong Standard Method of Measurement of Building Works*, 3rd Ed., Hong Kong: Green Pagoda Press.

91. R.S. Means Co. Inc. (2001). Square Foot and Assemblies Estimating Methods, 3rd Ed., Kingston, Mass. : R.S. Means Co., 14-19.

92. Shash, A.A. (1993). "Factors Considered in Tendering Decisions by Top UK Contractors." *Construction Management and Economics,* 11(2), 111-118.

93. Shen, L.Y. (1997). "Project Risk Management in Hong Kong." *International Journal of Project Management,* 15(2), 101-105.

94. Skitmore, M. and Wilcock, J. (1994). "Estimating Processes of Small Builders." *Construction Management and Economics*, 12, 139-154.

95. Smith, A. E. and Mason, A. K. (1997). "Cost Estimation Predictive Modeling: Regression Versus Neural Network." *The Engineering Economist*, 42 (2), 137 – 161.

96. Smith, N. J., Merna, T., Jobling, P. (1999). *Managing Risk in Construction Projects*, Blackwell Science, 72-95.

97. Solomon, G. (1993). "Cost Analysis, Preliminaries, The Marketplace Effect." *Chartered Quantity Surveyors*, October, 9 – 11.

98. Steward, R. D., Wyskida, R. M., & Johannes, J. D. (1995). "Construction cost estimating." In R. D. Steward, (Ed.), *Cost estimator's reference manual,* 2nd Ed., New York: John Wiley & Sons, Inc., 353-406.

99. Tah, JHM, Thorpe, A and McCaffer, R (1994). "A Survey of Indirect Cost Estimating in Practice." *Construction Management and Economics*, 12, 31-36.

100. Tam, C. M. and Harris, F. (1996). "Model for Assessing Building Contractors' Project Performance." *Engineering, Construction and Architectural Management*, 3, 187 – 203.

101. Tam, C. M., Tong, T. K. L., Lau, T. C. T. and Chan, K. K. (2005). "Selection of Vertical formwork System by Probabilistic Neural Networks Models." *Construction Management and Economics*, 23(3), 245-254.

102. Tam, C. M., Tong, T. K. L. and Tse, S. L. (2004). "Modelling Hook Times of Mobile Cranes using Artificial Neural Networks." *Construction Management and Economics*, 22(10), 839-849.

103. Taylor, G. R. (1994). "The Importance of Estimating Your Overhead." *Cost Engineering*, 36(2), 15 – 18.

104. Tikhonov, A.N. and Arsenin, V.Y. (1977). *Solution of Ill-Posed Problems*, W.H. Winston, Washington, D.C.

105. Trost, S.M. and Oberlender, G.D. (2004). "Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression." *Journal of Construction Engineering and Management*, 129(2), 198-204.

106. Ulug, S. Erol. (1993). "Water distribution network cost estimates for small urban areas." *International journal of environmental studies*, 44(1), 63-75.

107. Vojinovic, Z. and Kecman, V. (2001). "Modelling Empirical Data to Support Project Cost Estimating: Neural Networks Versus Traditional Methods." *Construction Innovation*, 1, 227-243.

108. Ward Systems Group, Inc. "Build Neural Network Reference." and "Apply Neural Networks." *NeuroShell 2 User's Manual*, Ward Systems Group Inc., Frederick, MD, USA, 91-178.

109. Weinbach, R. M. and Grinnell, R. M. (2001). *Statistics for Social Workers*, 5th Ed., Boston, Mass. : Allyn and Bacon, 10-17.

110. Wong, C.F., Lam, K.C., Cheung, S.O. and Gao, H. (2002). "A Model of Contractual Risk in Construction Projects." *Journal of Building and Construction Management*, 7(1), 42-49.

111. Wright, M.G. and Williams, T.P. (2001). "Using Bidding Statistics to Predict Completed Construction Cost." *The Engineering Economist*, 46, 114-128.

112. Yegidis, B.L. and Weinbach, R.W. (1996). "Methods for Acquiring Research Data." *Research Methods for Social Workers*, 2nd Ed., Boston: Allyn and Bacon, 159-178.

113. Zadeh, L.A. (1965). "Fuzzy Sets." *Information and Control*, 8, 338-353.

114. Zadeh, L.A. (1994). "Fuzzy Logic, Neural Networks, and Soft Computing." *Communications of the ACM*, 37(3), 77-84.

115. Zhou, Z., Li, T., Xu, Y., Bode, J., Ren, S. (1996). "Neural Engineering in Management Application." *Proc., 1996 IEEE International Conference on Systems, Man & Cybernetics*, Part 2 (of 4), 1399 – 1403.

10.2 Bibliography

1. Arditi, D., Alnajjar, A. and Vingert, N. (2002). "Construction Cost Estimating Support System." *Cost Engineering*, 44(10), 17-25.
2. Baloi, D. and Price, A.D.F. (2003). "Modelling Global Risk Factors Affecting Construction Cost Performance." *International Journal of Project Management*, 21, 261-269.
3. Besston, D. T. (1983). *Statistical Methods for Building Price Data*, London; New York: E. & F.N. Spon, 114 -126.
4. Bledsoe, J. D. (1998). "Statistical Cost Estimating for Public Works." *Proc., 42nd Annual Meeting of AACE International*, RISK.03.1 – 03.4.
5. Brown, J. A. (1998). "15 Keys for Successful Construction, Estimating, and Bidding." *AACE International Transactions of the Annual Meeting*, EST.08.1-3.
6. Elhag, T.M.S., Boussabaine, A.H. and Ballal, T.M.A. (2005). "Critical Determinants of Construction Tendering Costs: Quantity Surveyors' Standpoint." *International Journal of Project Management*, Article in Press,1-8.
7. Foster, J.J. (2001). "Factor Analysis." *Data Analysis. Using SPSS for Windows*, New Edition: Versions 8-10. U.K.: Sage Publications.
8. Friedman, L. (1956). "A Competitive Bidding Strategy." *Operations Research*, IV, 104 – 112.
9. Griffis, F. H. (1992). "Bidding Strategy: Winning Over Key Competitors." *Journal of Construction Engineering and Management*, 118(1), Mar., 151 – 165.
10. Hanssmann, F. and Rivett, B. H. P. (1959). "Competitive bidding." *Operational Research Quarterly*, 10(1), 49 – 55.
11. Hegazy, T. (1998). "Neural Network Model for Parametric Cost Estimation of Highway Projects." *Journal of Construction Engineering and Management*, May/June, 210 – 218.
12. Herbsman, Z. (1995). "A + B Bidding Method – Hidden Success Story for Highway Construction." *Journal of Construction Engineering and Management*, 121(4), Dec., 430 – 437.
13. Herbsman, Z., and Ellis, R. (1992). "Multiparameter Bidding System – Innovation in Contract Administration." *Journal of Construction Engineering and Management*, 118(1), Mar., 142 -149.

14. Ioannou, P. G., Leu, S.S. (1993). "Average-bid Method – Competitive Bidding Strategy." *Journal of Construction Engineering and Management*, 119(1), Mar., 131 – 147.
15. Jrade, A. and Alkass, S. (2001). "A Conceptual Cost Estimating Computer System for Building Projects." *AACE International Transactions; 2001*, IT.09.1-8.
16. Khosrowshahi, F. (1999). "Neural Network Model for Contractors' Prequalification for Local Authority Projects." *Engineering, Construction and Architectural Management*, 6(3), 315-328.
17. Lam, K. C., Hu, T., Cheung, S. O., Yuen, R. K. K., Deng, Z. M. (2001). "Multi-Project Cash Flow Optimization: Non-inferior Solution Through Neuro-multiobjective Algorithm." *Engineering, Construction and Architectural Management*, 8(2), 130-144.
18. Leo, D. W. (1998). "Project Estimating for the 21st Century." *Proc., 42nd Annual Meeting of AACE International*, AACE, EST.09.1 – 09.4.
19. National Institute for Strategic Studies, International Center for Information Technologies and Systems of the National Academy of Sciences of Ukraina. (2006). "GMDH". Available from <http://www.gmdh.net> (May 15, 2006).
20. Park, W. R. and Chapin, W. B. (1992). *Construction Bidding: Strategic Pricing for Profit*, 2nd Ed., New York: Wiley, 95-123.
21. Rencher, A.C. (2002). "The Multivariate Normal Distribution." and "Factor Analysis." *Methods of Multivariate Analysis*, 2nd Ed., Wiley-Science.
22. Shtub, A. and Zimmerman, Y. (1993). "Neural-network-based Approach for Estimating the Cost of Assembly Systems." *International Journal of Production Economics*, 32(2), 189-207.
23. Skitmore, R. M. (1986). "A Model of the Construction Project Selection and Bidding Decision." PhD Thesis, University of Salford, U.K.
24. Skitmore, M. and Patchell, B. (1990). "Contract Price Forecasting and Bidding Techniques." In Brandon, P.S. (Ed.), *Quantity Surveying Techniques: New Directions*, Oxford: Blackwell Scientific Publications.
25. Smith, A. J. (1995). *Estimating, Tendering and Bidding for Construction: Theory and Practice*, London: Macmillan, 173 – 190.
26. Tam, C.M. and Tong, T.K.L. (2003). "GA-ANN Model for Optimizing the Locations of Tower Crane and Supply Points for High-Rise Public Housing Construction." *Construction Management and Economics*, 21, 257-266.

27. Woodward, J. F. (1975). *Quantitative Methods in Construction Management and Design*, London: Macmillan.
28. Zhong, L. Z., Tao, L., Yong, Q. X., Bode, J. and Shou, J. R. (1996). "Neural Engineering in Management Application." *Proc., 1996 IEEE International Conference on Systems, Man and Cybernetics*, Part 2 (of 4), 1399 – 1403.

CHAPTER 11

Appendix

APPENDIX A

Questionnaire for Stage 1 Survey

This questionnaire studies the practice and difficulties of **project overheads estimation** during the tender stage from contractors' viewpoint. The eventual purpose of this study is to generate data for a PhD research to develop a project overheads estimation model.

The questionnaire would only take you 2 minutes to complete. All the information collected will be treated in strict confidentiality.

Name of your Company : ____

____

Please put a " ✓ " in the box provided against the appropriate answer.

1.0 Nature of construction works engaged by your Company (you can tick more than one) :
Building works - new works ☐ alterations ☐ maintenance works ☐
Civil Engineering works – foundation ☐ infrastructural works ☐
Specialist works (like Building services engineering works) ☐
Others ☐

2.0 How do your company allow for project overheads in tenders ?
a. Not allowed at all ☐
b. Estimate as a percentage of total value of work ☐
c. Estimate as a lump sum allowance ☐
d. Estimate in detail with reference to tender documents ☐

Go directly to Question 4.0

3.0 How does the detailed estimate of project overheads calculated ?
a. Based on estimator's professional knowledge and experience ☐
b. Build up plant, labour, materials, subcontractor and/or other associated cost for each project overheads item ☐
c. Estimate using sophisticated estimation model, e.g. simulation model, expert system ☐

4.0 What is the major source of cost data for estimating project overheads ?
a. Cost indices / schedule of rates published by the Hong Kong Government or consultant QS firms like Davis Langdon and Seah, Levett & Bailey ☐
b. Quotations from subcontractors or suppliers for each relevant item, e.g. scaffolding ☐
c. In-house cost database ☐
d. Others ☐ ____

5.0 If there is an cost database maintained in your Company, is there ay feedback / review system on the estimating accuracy?
a. Yes ☐
b. No ☐
c. Not applicable ☐

* * * The End * * *

Thank you very much for your kind assistance.

Please kindly return this questionnaire by fax **to 2788 9716, <u>Attn : Ms Caroline Chan</u>**.
For any queries, please contact Caroline Chan at 2788 9798 or email to bstwchan@cityu.edu.hk

APPENDIX B

Structured Interview Questions for Follow-up Interview After Stage 1 Survey

Follow-up Interview after Stage 1 Survey on General Estimating Practice of Project Overheads
Structured Interview Questions

Name of Company : ____

Name of Interviewee / Position : ____________________________________

Date / Time : ____________________

Area of questions	Expected outcomes from the questions
1. Estimation method used by the company	- Identify :- Major component in estimating, - Staff involved in estimating project overheads, - Time spent in estimating, - Perceived risk transferability of estimates, - General accuracy of project overheads estimation
2. Project overhead items likely to be under- or over-estimated	- Identify at most 3 items likely to be under- or over-estimated by each interviewee
3. List of project overheads to be included for pricing/ study	- Identify the project overhead items that should be priced / studied

Introduction :
Thank you for sparing your time to participate this interview. The purpose of this interview is to collect more information about the estimation practice of your company, and your professional view on accuracy of project overheads estimation.

Interview Questions :

1. **Can you tell me which component in the tender estimate is the most important ?**
   Project overheads ______
   Measured work ______
   Profit margin ______
   Attendance ______

   Supplement by interviewee : ____

   ____

2. **Which grade of staff in your company will be involved in estimating? And what their responsible part (in the tender bid preparation) ?**
   Senior estimators _____Responsible for :______________________________

   ____

   ____

   Junior estimators _____Responsible for :______________________________

   ____

   ____

3. **Estimation error in which part of the tender do you think is the most difficult to transfer ?**
   Project overheads ______
   Measured work ______
   Profit margin ______
   Attendance ______

   Supplement by interviewee : ______________________________

   ____

4. **Does over- or under-estimation of project overheads common in your company estimates ?**
   Do you think it is acceptable? Why ?

   ____

   ____

   ____

   ____

   ____

   ____

5. **Can you identify 3 items (or less) most likely to be under- or over-estimated ?**

   1. ______________________________

   2. ______________________________

   3. ______________________________

6. **Which items would you select if a (representative & typical) list of project overheads items to be prepared ?** Give out the HKSMM list of project overhead items → **You can pick from the HKSMM list or suggest any item to add / combine / delete.**

HKSMM list of project overhead items :

	HKSMM project overheads items	Suggested view		
		Include	**Combine with**	**delete**
1	Employer, Architect and Quantity Surveyor			
2	Description of works			
3	Site and inspection			
4	Division of work into sections			
5	Form of Contract			
6	Particulars to be inserted in Appendix to Schedule of Conditions			
7	Working hours, rates of wages, etc.			
8	Plant, tools, sheds, etc.			
9	Notices and fees			
10	Safety precautions			
11	Industrial Training Levy			
12	Setting out			
13	Foreman			
14	Protection of public property, etc.			
15	Sub-letting			
16	Artists or Tradesmen not Sub-contractors			
17	Injury to persons and damage to property			
18	Insurances, etc.			
19	Provisional and Prime Cost Sums			
20	Conditions of payment			
21	Surety or bond			
22	Watching			
23	Protection			
24	Treasure trove, coins, etc.			
25	Variations and methods of measuring and valuing			
26	Samples			
27	Testing of materials			
28	Water			
29	Lighting and power			
30	Attendance			
31	Hoardings, etc.			
32	Works by Public Authorities			
33	Temporary roads			
34	Drying the building			
35	Removal of rubbish			
36	Defects after completion			

Do you have any suggestion to the interview ?

____

____

____

____

____

____

APPENDIX C

Questionnaire for Stage 2 Survey

This questionnaire studies the **likelihood of inaccurate estimation of project overheads** during the tender stage from contractors' viewpoint. The eventual purpose of this study is to generate data for a PhD research to develop a project overheads estimation model. The questionnaire would only take you about 3 minutes to complete. All the information collected will be treated in strict confidentiality.

Section 1 General Information

Name of your Company :____

Please put a " ✓ " in the box provided against the appropriate answer.

1.0 Nature of construction works engaged by your Company (you can tick more than one) :
General contractor works :
Building works - new works ☐ alterations ☐ maintenance works ☐
Civil Engineering works – foundation ☐ infrastructural works ☐
Others ☐

Specialist works :
Building Services Engineering works - new works ☐ alterations ☐ maintenance works☐
Builder's specialist works e.g. waterproofing, cladding,… ☐
Others ☐

2.0 How did your company allow for project overhead costs in tenders ?
Estimate in detail ☐
Estimate as a percentage of total value of work ☐
Estimate as a lump sum ☐
Not allowed at all ☐

3.0 How accurate is your method of estimating Project overhead costs ?
Actual project overhead costs are generally under-estimated ☐
About the same as actual project overhead costs ☐ (if ✓ this answer, please go directly to question 5.0)
Actual project overhead costs are generally over-estimated ☐

4.0 Can any of the shortfall / surplus of the Project overhead costs allowance as mentioned in your answer to 3.0 be adjusted later (after the contract is being awarded) in the Final Account, say by loss and expenses claim for the extended contract period ?

Yes ☐ Partly ☐ No ☐

Section 2 Project Overhead Costs Estimation

5.0 According to your experience with this Company, please indicate the general extent of under / over estimation of the financial impact of the following project overhead cost items during the tender stage, by circling at the appropriate rating.

Note that rating "3" (neutral) means that the Project overhead cost item estimated during the tender stage was about the same when compared with the actual expenditure of Project overhead cost (or the actual financial impact of the Project overhead cost) during the construction stage.

		Nature of Estimation				
	Project Overhead Items	over-estimate		*neutral*		under-estimate
1	Definitions of various contractual parties	1	2	*3*	4	5
2	Description of works	1	2	*3*	4	5
3	Nature of Site and site inspection	1	2	*3*	4	5
4	Site possession and completion	1	2	*3*	4	5
5	Site management, watchman and attendance to NSCs	1	2	*3*	4	5
6	Principles of measurement of the Bills of Quantities	1	2	*3*	4	5
7	Drawings to be prepared by the Main Contractor	1	2	*3*	4	5
8	Contract conditions and amendments	1	2	*3*	4	5
9	Methods of measuring and valuing variations	1	2	*3*	4	5
10	Fees and levies	1	2	*3*	4	5
11	Restrictions to noise and dust nuisance	1	2	*3*	4	5
12	Insurances	1	2	*3*	4	5
13	Protection of finished works	1	2	*3*	4	5
14	Protection of adjacent / existing works	1	2	*3*	4	5
15	Safety precautions	1	2	*3*	4	5
16	Facilities to be provided by Main Contractor : plants, scaffolding	1	2	*3*	4	5
17	Site offices, stores, latrines	1	2	*3*	4	5
18	Setting out	1	2	*3*	4	5
19	Samples, mock ups	1	2	*3*	4	5
20	Testing	1	2	*3*	4	5
21	Power and water supply	1	2	*3*	4	5
22	Temporary works e.g. hoardings, temporary roads, signboard	1	2	*3*	4	5
23	Cleaning and removal of rubbish	1	2	*3*	4	5

6.0 Any other comments regarding to estimation of project overhead costs during tender stage :

____

____

____

____

____

____

____

____

* * * The End * * *

Thank you for your kind assistance.

APPENDIX D

Questionnaire for Opinion Survey on Impact of Factors Affecting Project Overheads

This questionnaire aims to collect views on the **factors affecting the expenditure of project overheads like site management, site offices, hoarding, etc.**The eventual purpose of this study is to generate data for a PhD research to develop a project overheads estimation model.

The questionnaire would only take you about 5 minutes to complete. All the information collected will be treated in strict confidentiality.

Section 1 General Information

Name of your Company : ______________________________

Please put a " ✓ " in the box provided against the appropriate answer.

1.0 Nature of your company business (you can ✓ more than one) :
Building works - new works ☐ alterations ☐ maintenance works ☐
Civil Engineering works – foundation ☐ infrastructural works ☐
Specialist works (like Building services engineering works) ☐
Others ☐

6.0 Your tenure in this company : ______________ years

Section 2 Factors Affecting Project Overheads Expenditure

3.0 According to your experience, please indicate your views on the external factors affecting project overheads **expenditure** (not project overheads pricing), by circling at the appropriate rating.

	External Factors affecting project overheads expenditure	**Extent of influence to the expenditure of project overheads** Significant influence		Slight influence	***Nil/Negligible influence***
	Information available at tender stage				
1.	Estimated value of project works	3	2	1	***0***
2.	Duration of Project	3	2	1	***0***
3.	Type of Project (e.g. commercial, residential, industrial)	3	2	1	***0***
4.	Size of Project (gross floor area of proposed building)	3	2	1	***0***
5.	No. of storey of proposed structure	3	2	1	***0***
6.	No. of level of basement (of proposed structure)	3	2	1	***0***
7.	Complexity of project	3	2	1	***0***
8.	Shape of site	3	2	1	***0***
9.	Shape of building	3	2	1	***0***
10.	Site coverage of proposed building	3	2	1	***0***
11.	Accessibility of site	3	2	1	***0***
12.	Required quality level of project	3	2	1	***0***
13.	Selection method by client (i.e. negotiation or competitive tendering)	3	2	1	***0***
14.	Procurement method of project (e.g. traditional, design and build, management contract)	3	2	1	***0***
15.	Type of contract (e.g. standard form, government form)	3	2	1	***0***
16.	Payment terms	3	2	1	***0***
17.	Completeness of project information and drawings from architect / engineer (drawings, design, specifications)	3	2	1	***0***
18.	Extent of bond / warranty requirement from the project	3	2	1	***0***

	External Factors affecting project overheads expenditure	Extent of influence to the expenditure of project overheads Significant influence	Slight influence		*Nil/Negligible influence*
	Information unavailable at tender stage				
19.	Soil conditions	3	2	1	*0*
20.	Promptness of response from designers	3	2	1	*0*
21.	Extent of variations	3	2	1	*0*
22.	Strictness in supervision by client / client's representatives	3	2	1	*0*
23.	Punctuality of payment by client	3	2	1	*0*
24.	Economic conditions of Hong Kong at construction stage	3	2	1	*0*
25.	Economic conditions of the construction market at construction stage	3	2	1	*0*
26.	Interest rate during the construction stage	3	2	1	*0*
27.	Inflation rate during the construction stage	3	2	1	*0*

4.0 According to your experience, please indicate your views on the internal factors affecting project overheads **expenditure** (not project overheads pricing), by circling at the appropriate rating.

	Internal Factors affecting project overheads expenditure	Extent of influence to the expenditure of project overheads Significant influence	Slight influence		*Nil/Negligible influence*
	Information available at tender stage				
1.	Extent of off-site fabrication or prefabrication in the project works	3	2	1	*0*
2.	General salary level of project management staff	3	2	1	*0*
3.	Extent of sub-letting of project works	3	2	1	*0*
4.	Technical competence of sub-contractors	3	2	1	*0*
5.	Extent of standardization in site management documentation (e.g. standard policy and forms used for inspection, query, etc.)	3	2	1	*0*
6.	Degree of computerization in site office	3	2	1	*0*
7.	Extent of mechanical plants owned by the contractor	3	2	1	*0*
8.	Size of the company (the contractor)	3	2	1	*0*
9.	Contractor is client's subsidiary firm	3	2	1	*0*
10.	Relationship with client in past projects	3	2	1	*0*
11.	Relationship with designer in past projects	3	2	1	*0*
12.	Relationship with PQS in past projects	3	2	1	*0*
	Information unavailable at tender stage				
13.	Past experience of project manager in similar project	3	2	1	*0*
14.	Claims consciousness of project manager who will be assigned to the project	3	2	1	*0*
15.	Interpersonal skills of project manager who will be assigned to the project	3	2	1	*0*
16.	Technical competence of site management staff	3	2	1	*0*
17.	No. of management staff assigned on site for the project	3	2	1	*0*
18.	Extent of mechanical plants used for the project	3	2	1	*0*
19.	Financial stability and cash flow of sub-contractors at construction stage	3	2	1	*0*
20.	Workload of subcontractors	3	2	1	*0*

5.0 Is there any other factor that you think would affect the expenditure of project overheads but not included in the above list? If yes, please state.

____

____

* * * The End * * *

Please kindly return this questionnaire by fax **to 2788 9716**, **Attn : Ms Caroline Chan**.

For any queries, please contact Caroline Chan at 2788 9798 or email to bstwchan@cityu.edu.hk

* * * Thank you for your kind assistance. * * *

APPENDIX E

Proforma to Collect Project Information for ANN Model

Project information sheet
Please use this sheet for one project.

The information collected in this proforma is used to conduct a research study on estimation of project overheads. All information collected will be treated in strict confidentiality.

Project scope

Please give a brief description on the scope of work in this project (you can extract this information from the Preliminaries Section of the contract document normally).
E.g. A 35-storey high-class residential building including 3 levels of basement.

Project details
Please put a " √ " in the appropriate box or fill in the information in the space provided.

Project type: ☐ 1. Foundation work, site formation, etc.

☐ 2. Renovation or alteration works

☐ 3. Factory or domestic housing which required a little amount of M&E co-ordination

☐ 4. Deluxe housing project or office building which required more subcontracting and M&E co-ordination

☐ 5. Hotel or high-class office building

☐ 6. Hospital or complicated project

Is your company the subsidiary of the project client ? ☐ 1. Yes or ☐ 2. No

No. of storeys (superstructure) : [ ]

No. of levels (basement) : [ ]

Gross floor area : [m^2]

Project commencement date (mm/yy): [/]

Project completion date (mm/yy) : [/]

Final project overhead cost : [HK$]

<u>Your assessment of the project</u>

*Please read the explanation of each item carefully and evaluate the financial impact of these attributes to the **PROJECT OVERHEADS EXPENSES** of this project by circling the appropriate answer.*

		Small amount			Large amount
1.	**Design and variations** *Refers to the completeness of design information, promptness of response from designers and extent of variations.*	1	2	3	4
2.	**Bonds and warranties** *Refers to the extent of bond / warranty required in this project.*	1	2	3	4
3.	**Procurement nature** *Refers to the complexity of procurement system adopted for the main contract in terms of the selection method (e.g. negotiation vs competitive tendering), contractual arrangement (e.g. D&B, management contracting) and type of contract (e.g. HKIA standard form vs client's in-house form)*	1	2	3	4
4.	**Project complexity** *Refers to the complexity of the project in terms of the use of special techniques / plants / technicians.*	1	2	3	4
5.	**Site and building shape** *Refers to the irregularity of site and building shape.*	1	2	3	4
6.	**Location** *Refers to the accessibility of the site and the extent of supervision by client or his representatives.*	1	2	3	4
7.	**Strength of management team** *Refers to the overall strength of the contractor's project team : PM's past experience in similar project, PM's interpersonal skills and the technical competence of the team.*	1	2	3	4
8.	**Contractor's past relationship with project parties** *Refers to the external relationship with client, designers and PQS in past projects.*	1	2	3	4
9.	**Extent of staff and plants assigned to project** *Refers to the amount of management staff and mechanical plants used in this project.*	1	2	3	4
10.	**Financial strength of sub-contractors** *Refers to the overall financial strength of sub-contractors with consideration of sub-contracting extent.*	1	2	3	4
11.	**Sub-contractors performance** *Refers to the combined effect of the technical competence of the sub-contractors and the extent of standardization in site management documentation.*	1	2	3	4

APPENDIX F

Calculation of Maximum Absolute Difference in Rank and RA_{max} in Stage 2 Survey

Table F-1 Calculation of Maximum Absolute Difference in Rank and RA_{max} for 23 items in the Stage 2 survey

Item	Ranking by group 1, R_{i1}	Ranking by group 2, R_{j2}	Maximum absolute difference in rank
A	1	23	22
B	2	22	20
C	3	21	18
D	4	20	16
E	5	19	14
G	6	18	12
H	7	17	10
I	8	16	8
J	9	15	6
K	10	14	4
L	11	13	2
M	12	12	0
N	13	11	2
O	14	10	4
P	15	9	6
Q	16	8	8
R	17	7	10
S	18	6	12
T	19	5	14
U	20	4	16
V	21	3	18
W	22	2	20
X	23	1	22

Hence, the Maximum rank agreement factor (RA_{max}) for 23 items (from A to X) in the Stage 2 survey is :

$$RA_{\max} = \frac{\sum_{i=1}^{N} \left| R_{i1} - R_{j2} \right|}{N} = \frac{264}{23}$$

$$= 11.4783$$

APPENDIX G

Rotated Component Matrices and Correlation Matrices of Factors

Table G-1 Rotated Component Matrix for the External Factors

Ref.	Component							
	1	2	3	4	5	6	7	8
E.17	0.800	(0.021)	0.098	0.124	0.180	0.177	0.010	(0.092)
E.20	0.714	0.203	0.142	0.172	0.136	0.040	0.340	(0.116)
E.21	0.600	0.442	0.104	(0.067)	0.119	(0.040)	0.064	0.046
E.24	0.126	0.901	0.130	0.055	0.056	0.048	0.072	0.030
E.25	0.095	0.888	0.132	0.079	0.100	(0.097)	0.157	(0.015)
E.18	(0.038)	(0.234)	0.732	0.202	0.010	(0.056)	0.013	(0.205)
E.27	0.169	0.314	0.713	(0.041)	0.111	0.122	0.126	(0.017)
E.3	0.199	0.226	(0.097)	0.656	0.100	0.058	0.084	(0.154)
E.5	0.169	(0.256)	0.177	0.615	(0.044)	0.236	0.127	0.326
E.6	(0.192)	(0.045)	0.177	0.672	0.053	0.165	0.021	0.164
E.13	0.112	0.098	(0.013)	(0.065)	0.731	0.058	0.077	(0.090)
E.14	0.390	(0.059)	0.150	0.249	0.629	(0.330)	(0.009)	0.006
E.15	(0.051)	0.032	0.033	(0.017)	0.741	0.119	0.131	0.144
E.8	(0.035)	0.090	0.130	0.128	0.134	0.752	0.085	0.221
E.9	0.143	0.075	0.233	0.304	(0.030)	0.721	0.066	(0.053)
E.11	0.219	(0.013)	0.226	0.107	0.010	0.076	0.708	0.141
E.12	0.048	0.052	(0.117)	0.291	0.215	(0.028)	0.685	(0.337)
E.4	(0.088)	0.113	(0.075)	0.325	0.099	0.090	0.096	0.746
E.1	0.361	0.243	0.263	(0.198)	(0.032)	(0.352)	0.168	0.353
E.2	(0.259)	0.207	0.163	0.230	(0.136)	(0.310)	(0.059)	0.103
E.7	0.120	0.065	0.019	0.315	(0.075)	(0.012)	0.125	0.031
E.10	0.294	(0.086)	(0.026)	0.079	(0.166)	0.485	0.386	0.429
E.16	0.283	0.109	0.285	0.029	0.373	0.068	0.225	(0.109)
E.19	0.403	0.103	0.185	0.152	0.086	0.107	0.246	(0.350)
E.22	(0.006)	0.271	0.003	0.017	0.273	0.094	0.721	0.084
E.23	0.217	0.136	0.465	0.038	0.392	0.106	0.422	0.007
E.26	0.174	0.349	0.396	0.031	0.105	0.057	(0.013)	0.088

Rotation Method: Varimax with Kaiser Normalization.

Table G-2 Rotated Component Matrix for the Internal Factors

Ref.	Component					
	1	2	3	4	5	6
I.13	0.680	0.280	0.042	0.360	(0.040)	(0.020)
I.15	0.776	0.233	(0.032)	0.155	0.164	0.107
I.16	0.773	0.195	0.141	0.013	0.126	(0.023)
I.10	0.160	0.728	(0.194)	0.083	0.110	0.431
I.11	0.284	0.852	(0.035)	0.065	0.053	0.091
I.12	0.193	0.834	(0.046)	0.182	0.068	(0.022)
I.17	0.261	(0.170)	0.677	0.184	(0.246)	(0.136)
I.18	(0.092)	0.098	0.799	(0.076)	0.218	(0.014)
I.3	0.024	(0.054)	0.304	0.670	0.404	0.045
I.19	0.245	0.160	0.009	0.730	0.280	0.179
I.4	0.195	0.199	(0.147)	0.261	0.682	0.040
I.5	0.503	0.102	(0.053)	(0.030)	0.667	0.030
I.9	0.070	0.155	(0.113)	0.277	0.094	0.682
I.1	0.020	(0.110)	0.321	0.236	0.371	0.071
I.2	0.066	(0.137)	0.397	0.033	(0.111)	0.121
I.6	0.385	(0.329)	0.086	0.134	0.254	0.370
I.7	0.012	(0.101)	0.345	(0.125)	0.078	0.378
I.8	(0.016)	0.264	0.403	0.084	(0.124)	0.523
I.14	0.376	0.407	(0.012)	0.341	0.150	(0.060)
I.20	0.233	0.192	(0.050)	0.437	(0.085)	0.114

Rotation Method: Varimax with Kaiser Normalization.

Table G-3 Correlation Matrix of External Factors

Factor	E.1	E.2	E.3	E.4	E.5	E.6	E.7	E.8	E.9	E.10	E.11	E.12	E.13	E.14	E.15	E.16	E.17	E.18	E.19	E.20	E.21	E.22	E.23	E.24	E.25	E.26	E.27
E.1	1.000																										
E.2	0.217	1.000																									
E.3	-0.086	0.055	1.000																								
E.4	0.084	0.105	0.202	1.000																							
E.5	-0.071	-0.022	0.153	0.321	1.000																						
E.6	-0.178	0.036	**0.560**	0.272	**0.541**	1.000																					
E.7	-0.002	0.079	0.418	0.193	0.406	0.262	1.000																				
E.8	-0.235	-0.113	0.070	0.253	0.317	0.183	0.171	1.000																			
E.9	-0.243	-0.221	0.270	0.130	0.342	0.326	0.170	**0.546**	1.000																		
E.10	-0.035	-0.231	0.086	0.309	0.403	0.103	0.190	0.415	0.319	1.000																	
E.11	0.203	-0.038	0.220	0.141	0.234	0.149	0.175	0.163	0.219	0.404	1.000																
E.12	0.000	-0.060	0.287	-0.071	0.197	0.159	0.184	0.086	0.088	0.105	**0.642**	1.000															
E.13	0.051	-0.067	0.003	-0.064	-0.023	-0.077	0.017	0.216	0.014	0.012	0.105	0.233	1.000														
E.14	0.218	0.007	0.201	0.039	0.127	0.089	0.158	-0.124	-0.066	-0.049	0.094	0.226	0.430	1.000													
E.15	0.013	-0.125	0.121	0.170	0.028	0.031	-0.030	0.230	0.026	0.017	0.248	0.167	**0.654**	**0.530**	1.000												
E.16	0.128	-0.093	0.230	0.007	0.077	0.101	0.003	0.042	0.231	0.045	0.307	0.247	0.401	0.352	0.367	1.000											
E.17	0.182	-0.261	0.322	-0.083	0.222	-0.049	0.145	0.155	0.254	0.253	0.178	0.194	0.181	0.372	0.182	0.285	1.000										
E.18	0.064	0.064	0.001	-0.113	0.207	0.152	0.161	0.047	0.203	-0.070	0.125	0.060	0.046	0.202	0.008	0.077	0.093	1.000									
E.19	-0.024	-0.053	0.240	-0.218	0.081	-0.063	0.162	0.068	0.196	0.068	0.263	0.341	0.198	0.205	0.107	0.329	0.407	0.209	1.000								
E.20	0.241	-0.091	0.294	-0.006	0.175	0.039	0.221	0.103	0.328	0.187	0.442	0.318	0.216	0.405	0.074	0.454	0.540	0.074	0.470	1.000							
E.21	0.340	0.013	0.084	0.015	0.029	-0.036	-0.033	0.095	0.097	0.046	0.200	0.143	0.204	0.244	0.059	0.348	**0.565**	0.011	0.253	**0.586**	1.000						
E.22	0.130	-0.094	0.187	0.213	0.068	0.020	0.112	0.206	0.150	0.232	0.353	0.474	0.227	0.182	0.265	0.237	0.138	-0.023	0.137	0.352	0.177	1.000					

Factor	E.1	E.2	E.3	E.4	E.5	E.6	E.7	E.8	E.9	E.10	E.11	E.12	E.13	E.14	E.15	E.16	E.17	E.18	E.19	E.20	E.21	E.22	E.23	E.24	E.25	E.26	E.27
E.23	0.211	-0.119	0.082	0.112	0.121	0.147	0.168	0.194	0.298	0.142	0.327	0.285	0.283	0.346	0.210	0.437	0.296	0.257	0.285	0.484	0.304	0.489	1.000				
E.24	0.295	0.062	0.188	0.097	-0.049	0.015	0.133	0.106	0.154	0.070	0.143	0.123	0.180	0.098	0.087	0.204	0.119	-0.062	0.190	0.276	0.427	0.266	0.249	1.000			
E.25	0.335	0.138	0.170	0.075	-0.093	0.055	0.146	0.005	0.056	-0.005	0.139	0.200	0.174	0.148	0.137	0.207	0.105	-0.010	0.210	0.333	0.452	0.351	0.302	**0.882**	1.000		
E.26	0.311	0.054	0.142	0.049	0.141	0.092	0.099	0.171	0.200	0.060	0.221	-0.057	0.087	0.228	0.152	0.327	0.263	0.404	0.213	0.297	0.270	0.174	0.450	0.439	0.408	1.000	
E.27	0.229	0.064	0.163	-0.015	0.042	0.064	0.018	0.225	0.193	0.126	0.296	0.092	0.106	0.205	0.163	0.361	0.293	**0.684**	0.286	0.302	0.272	0.217	0.412	0.363	0.338	0.341	1.000

Table G-4 Correlation Matrix of Internal Factors

Factor	I.1	I.2	I.3	I.4	I.5	I.6	I.7	I.8	I.9	I.10	I.11	I.12	I.13	I.14	I.15	I.16	I.17	I.18	I.19	I.20
I.1	1.000																			
I.2	0.263	1.000																		
I.3	0.433	0.268	1.000																	
I.4	0.274	(0.091)	0.348	1.000																
I.5	0.211	(0.052)	0.268	**0.593**	1.000															
I.6	0.293	0.115	0.264	0.221	0.390	1.000														
I.7	0.323	0.381	0.124	(0.079)	(0.020)	0.259	1.000													
I.8	0.169	0.260	0.136	0.076	0.009	0.085	0.320	1.000												
I.9	0.097	0.032	0.222	0.204	0.176	0.148	0.133	0.228	1.000											
I.10	(0.075)	(0.188)	0.081	0.297	0.234	0.078	0.032	0.238	0.444	1.000										
I.11	(0.036)	(0.081)	0.048	0.245	0.263	0.060	(0.067)	0.244	0.163	0.407	1.000									
I.12	(0.006)	(0.129)	0.079	0.299	0.185	(0.035)	(0.110)	0.123	0.198	**0.614**	**0.728**	1.000								
I.13	0.088	0.052	0.250	0.269	0.288	0.277	(0.017)	0.088	0.188	0.324	0.411	0.381	1.000							
I.14	0.098	(0.013)	0.267	0.306	0.486	0.192	(0.056)	0.093	0.224	0.349	0.525	0.433	0.477	1.000						
I.15	0.180	(0.020)	0.176	0.324	0.445	0.415	0.054	0.059	0.195	0.386	0.398	0.429	0.401	0.447	1.000					
I.16	0.169	0.088	0.136	0.321	0.399	0.306	0.060	0.080	0.152	0.218	0.355	0.293	**0.532**	0.481	**0.578**	1.000				
I.17	0.238	0.403	0.225	(0.114)	(0.029)	0.216	0.120	0.260	(0.047)	(0.188)	(0.143)	(0.084)	0.130	0.049	0.096	0.208	1.000			
I.18	0.421	0.345	0.209	(0.070)	0.088	0.060	0.478	0.159	(0.007)	(0.071)	(0.000)	0.005	0.003	(0.015)	(0.058)	0.060	**0.523**	1.000		
I.19	0.353	(0.036)	**0.514**	0.374	0.278	0.334	0.078	0.160	0.293	0.320	0.285	0.318	0.457	0.485	0.444	0.248	0.028	0.049	1.000	
I.20	0.080	(0.047)	0.317	0.265	0.129	0.222	0.043	0.100	0.246	0.249	0.310	0.344	0.379	0.405	0.302	0.236	0.098	(0.033)	0.430	1.000

APPENDIX H

Project Data Collected for ANN Modelling

Table H-1 Project Data Collected for ANN Modeling

Ref	INPUTS														OUTPUT
	DES_VO	ECON	BOND	COMP	PROCU	SHAPE	LOCAT	SIZE	TEAM	RELAT	STA_PLT	SC_FIN	SC_PER	SUBSID	OUT_COST
1	3	100	2	3.76	1	1	3	12,232.00	2	3	2	3	3	1	16,543,000.00
2	3	111.74707	2	1.44	1	1	3	4,300.00	2	1	1	4	3	1	427,000.00
3	3	103.1056	3	3.36	1	4	4	9,300.00	2	3	2	3	3	1	25,128,000.00
4	4	96.072123	4	2.73	2	5	3	8,700.00	1	1	4	4	4	1	16,432,000.00
5	3	100	4	3.44	1	3	4	8,500.00	1	1	3	3	3	1	9,970,000.00
6	3	111.29732	4	3.11	1	2	4	8,769.00	2	3	2	4	2	1	13,996,000.00
7	5	106.21121	3	2.71	1	3	3	24,000.00	3	3	3	3	3	1	35,022,000.00
8	2	109.60195	4	10.36	1	2	3	32,000.00	1	1	4	2	4	1	50,236,000.00
9	3	100	2	2.65	1	3	4	8,000.00	2	2	3	3	3	1	11,913,000.00
10	5	108.97914	3	3.42	1	4	4	25,060.00	3	3	3	3	3	1	40,076,000.00
11	2	109.60195	4	11.51	1	2	3	40,000.00	2	2	4	3	2	1	82,450,000.00
12	3	109.60195	3	9.21	2	2	4	39,075.00	3	2	4	3	2	0	42,721,000.00
13	5	110.84756	2	7.49	5	2	4	37,000.00	3	3	4	2	5	1	65,403,000.00
14	3	98.036061	3	7.31	1	2	2	28,000.00	3	2	3	3	2	1	31,260,000.00
15	4	110.84756	3	11.4	1	2	4	49,000.00	2	3	4	2	2	1	70,380,000.00
16	3	108.97914	2	7.26	1	2	4	35,095.00	3	3	4	3	3	1	55,301,000.00
17	2	108.97914	3	10.65	1	2	4	47,540.00	2	2	4	3	3	1	68,072,000.00
18	3	105.98609	3	10.3	1	2	3	75,000.00	3	2	3	3	3	1	48,461,000.00
19	5	106.21121	3	12.07	1	2	3	43,000.00	2	1	4	2	2	1	48,140,000.00
20	3	100	3	10.76	1	2	3	120,000.00	2	2	3	3	2	1	78,390,000.00
21	3	100.76111	3	10.76	1	2	3	60,000.00	3	3	3	3	3	1	45,320,000.00
22	2	98.036061	2	2.02	1	3	3	4,000.00	3	2	2	3	3	1	4,982,000.00
23	3	98.036061	3	6.16	1	3	3	22,500.00	3	3	3	3	3	1	47,216,000.00
24	4	93.890079	3	9.54	2	2	4	74,000.00	2	2	4	3	3	1	128,044,000.00
25	5	109.60195	3	12.3	1	2	4	50,000.00	2	1	4	3	3	1	83,510,000.00
26	2	98.036061	3	3.05	1	2	1	8,900.00	2	2	2	3	3	1	9,870,000.00
27	1	108.97914	2	8.69	1	1	2	30,000.00	2	1	3	2	3	0	33,050,000.00
28	2	111.29732	2	9.84	1	2	2	35,000.00	2	2	3	2	3	1	38,730,000.00
29	2	103.1056	3	2.82	1	4	4	8,900.00	2	3	2	4	2	1	16,450,000.00

Table H-1 Project Data Collected for ANN Modeling (cont'd)

Ref	INPUTS														OUTPUT
	DES_VO	ECON	BOND	COMP	PROCU	SHAPE	LOCAT	SIZE	TEAM	RELAT	STA_PLT	SC_FIN	SC_PER	SUBSID	OUT_COST
30	4	93.890079	4	2.25	3	4	3	7,500.00	2	1	4	4	4	1	23,316,000.00
31	3	98.036061	3	6.85	1	2	3	35,000.00	2	3	3	3	2	1	44,280,000.00
32	5	100	2	3.82	5	2	2	13,000.00	5	3	2	3	4	1	20,555,000.00
33	3	111.74707	2	1.73	1	1	2	5,000.00	3	2	1	4	3	1	5,527,000.00
34	3	111.74707	3	3.88	1	2	3	9,500.00	3	3	2	3	3	1	15,549,000.00
35	3	98.036061	2	2.5	1	3	3	5,100.00	3	2	2	2	2	1	3,580,000.00
36	4	85.315378	3	10	2	2	4	75,000.00	1	2	4	2	2	1	129,800,000.00
37	5	108.97914	2	4.13	1	3	2	27,000.00	2	2	3	3	3	1	42,790,000.00
38	3	110.84756	4	11.78	1	2	3	43,000.00	2	2	4	2	2	1	89,020,000.00
39	4	110.84756	3	9.92	3	2	4	42,000.00	3	2	4	1	3	0	44,631,000.00
40	2	103.1056	3	2.36	1	3	3	8,000.00	2	2	2	3	2	1	14,977,000.00
41	4	91.708035	4	2.25	3	3	3	7,000.00	2	1	3	3	4	1	21,520,000.00
42	3	106.21121	3	11.3	5	2	5	77,000.00	3	2	3	5	3	1	68,000,000.00
43	2	103.1056	3	6.39	2	2	3	23,000.00	2	3	3	2	3	1	26,520,000.00
44	3	103.1056	3	11.01	1	2	3	124,000.00	2	2	3	3	3	1	52,460,000.00
45	3	96.072123	3	10.76	1	2	3	183,000.00	3	2	3	3	3	1	120,273,000.00
46	3	100	3	10.76	1	2	3	65,000.00	3	3	3	3	3	1	46,000,000.00
47	3	103.1056	2	7.25	1	1	3	18,000.00	2	2	3	2	2	0	26,093,000.00
48	3	108.97914	2	7.71	1	1	2	21,000.00	2	3	3	3	3	1	31,180,000.00
49	5	110.84756	3	12.53	1	2	4	54,000.00	2	1	4	2	2	1	83,975,000.00
50	4	100	3	7.85	1	2	3	37,000.00	2	3	3	2	2	1	63,836,000.00
51	3	100	2	3.28	1	1	2	11,000.00	2	3	2	3	3	1	15,386,000.00
52	3	110.84756	2	1.44	1	1	2	3,800.00	2	1	1	3	3	1	3,275,000.00
53	3	110.84756	4	2.65	1	2	3	8,500.00	2	2	2	3	2	1	12,840,000.00
54	4	103.1056	5	2.25	5	3	4	13,500.00	4	3	3	3	5	1	37,490,000.00
55	3	91.708035	3	10.76	1	2	2	167,000.00	3	2	3	3	2	1	115,390,000.00
56	3	100	2	2.42	1	2	3	7,200.00	2	2	2	3	3	1	10,862,000.00
57	3	93.890079	2	6.1	1	1	3	16,000.00	3	2	3	3	2	1	22,830,000.00
58	3	100	3	8.52	1	3	3	34,000.00	2	3	3	3	3	1	51,003,000.00

Table H-1 Project Data Collected for ANN Modeling (cont'd)

Ref	INPUTS														OUTPUT
	DES_VO	ECON	BOND	COMP	PROCU	SHAPE	LOCAT	SIZE	TEAM	RELAT	STA_PLT	SC_FIN	SC_PER	SUBSID	OUT_COST
59	3	109.60195	3	8.75	2	2	3	37,900.00	3	1	4	2	2	0	35,865,000.00
60	3	106.21121	2	6.55	1	2	3	34,000.00	3	3	4	2	2	1	56,490,000.00
61	2	106.21121	3	9.48	1	2	3	46,600.00	2	2	4	2	3	1	62,093,000.00
62	3	105.98609	3	9.61	1	2	3	68,000.00	3	2	3	2	3	1	41,200,000.00
63	3	103.1056	3	10.76	1	2	3	120,000.00	2	2	3	3	3	1	85,486,000.00
64	3	93.890079	3	10.76	1	2	3	180,000.00	3	2	3	3	3	1	118,494,000.00
65	3	100	3	10.76	1	2	3	55,000.00	3	3	3	3	2	1	37,317,000.00
66	2	96.072123	2	2.02	1	2	2	3,500.00	3	2	2	2	2	1	4,337,000.00
67	3	96.072123	3	5.7	1	3	2	21,300.00	3	2	3	3	3	1	40,196,000.00
68	2	111.29732	3	2.02	2	3	3	3,600.00	4	3	1	2	2	1	3,541,000.00
69	1	98.036061	3	3.34	1	1	2	10,000.00	1	1	2	2	2	1	5,742,000.00
70	4	100	3	6.87	1	4	3	24,000.00	3	3	3	3	3	1	48,570,000.00
71	2	111.29732	3	9.38	1	2	3	38,000.00	2	2	3	3	3	1	33,840,000.00

APPENDIX I

GMDH Modelling and Cross-validation Results

Table I-1 Training, Testing and Production Results of ANN Models Using GMDH Algorithm (Based on All 14 Inputs)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Number of inputs	14	14	14	14	14	14
Number of outputs	1	1	1	1	1	1
Number of training patterns	63	63	63	63	63	56
Number of test patterns	N/A	N/A	N/A	N/A	N/A	7
Number of production patterns	8	8	8	8	8	8
Layers constructed	5	6	7	6	3	14
Best criterion value	0.043418	0.051365	0.027129	0.019772	0.037208	0.000014
Training						
R squared	0.9192	0.929	0.9545	**0.9558**	0.9269	0.7951
Mean squared error	71,565,167,299,589	62,897,364,691,072	40,296,650,178,143	39,129,916,156,740	64,755,456,052,339	139,989,028,595,817
Mean absolute error	6,763,250	5,750,872	4,855,273	4,796,762	6,234,221	8,891,612
Min. absolute error	500,963	176,246	314,402	31,469	2	604,084
Max. absolute error	26,751,798	22,516,311	18,202,490	16,450,441	21,988,601	37,516,138
Correlation coefficient r	0.9589	0.9646	0.977	0.9777	0.9628	0.8917
Testing						
R squared						1
Mean squared error						83,779,361,619
Mean absolute error						237,245
Min. absolute error						25,811
Max. absolute error						463,295
Correlation coefficient r						1
Production						
R squared	0.9202	0.7998	0.858	**0.9574**	0.7597	0.8369
Mean squared error	146,365,800,133,674	367,430,206,673,466	260,523,154,998,481	78,229,762,816,407	441,010,672,806,768	299,345,864,926,514
Mean absolute error	8,365,369	11,814,313	8,981,302	5,267,334	14,928,188	12,157,300
Min. absolute error	295,401	352,835	573,067	99,545	1,011,265	2,151,742
Max. absolute error	23,749,323	49,997,400	43,570,882	23,328,637	43,692,982	38,649,688
Correlation coefficient r	0.9648	0.9036	0.9382	0.9853	0.8718	0.9162

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Best formula	Y=-1.3*X8+0.15-0.5*X2+0.19*X10+0.16*X11+3.4*X8*X11+0.38*X2^2+0.57*X5^2-0.65*X2*X5+1.2*X6*X8	Y=-0.5*X5+0.35-1.2*X2+3.1*X8+0.83*X2^2-4.5*X8^2+0.45*X10^2+3*X8^3-1.6*X8*X10+0.6*X5^2+0.32*X11^3-0.83*X8*X9+0.16*X6*X10	Y=-8.4E-002*X4+0.76*X11+0.24*X8^2-0.21*X11^3-0.33*X5*X11+0.13*X6*X11-3.1*X2*X11+15*X8*X11+4.9*X2^2*X11-2.3*X4^2*X11-2.4*X2^3*X11+2.4*X4^3*X11+1.1*X2*X4*X11-13*X2*X8*X11-16*X4*X8*X11+14*X2*X4*X8*X11+0.48*X5^2*X11+0.27*X4*X10+0.15*X1^2-0.35*X1*X11	Y=0.26*X9-9.7E-002*X13+0.14-0.12*X5+4.6E-002*X6-1.2*X2+5.1*X8+1.9*X2^2-1.3*X4^2-0.94*X2^3+1.5*X4^3-0.48*X8^3+0.31*X2*X4-4.3*X2*X8-9.8*X4*X8+4.7*X2*X4*X8+0.17*X5^2+0.87*X11^2-0.12*X10^3-0.57*X11^3+0.17*X10*X11+0.14*X10+0.36*X12-0.35*X3+0.88*X4-0.23*X11+1.6*X8^2-1.3*X4*X11+2.8*X8*X11+3*X4*X8*X11+0.29*X3^2+0.37*X3*X4-0.69*X12^2+0.52*X12^3-0.2*X10*X12-0.73*X9^2+0.11*X13^2+0.45*X9^3	Y=0.35-2.6*X3-0.28*X5-4.5*X8^2+3.4*X8^3+4.4*X8*X9+5*X8*X11+0.5*X9*X11-11*X8*X9*X11+5.4*X3^2+0.41*X5^2-3.3*X3^3	Y=6.6E-002*X12+5.4E-002-7.5E-002*X6+6.5E-002*X13-0.19*X2-2.6*X8^2+2.4*X8*X9^2+25*X8^2*X11-11*X8^2*X9*X11-0.54*X9^4-11*X8*X9^2*X11+5*X8*X9^3*X11-59*X8^2*X11^2+53*X8^2*X9*X11^2+46*X8^2*X9^2*X11^2-0.12*X1^3+4.9E-002*X1*X2+0.95*X2*X8-0.43*X2*X9^2-4.5*X2*X8*X11+2*X2*X8*X9*X11-0.26*X1*X2*X8+0.12*X1*X2*X9^2+1.3*X1*X2*X8*X11-0.56*X1*X2*X8*X9*X11-1.6*X8+0.71*X9^2+7.5*X8*X11-3.3*X8*X9*X11+9.9E-003*X2^2+2.3*X10^2+0.24*X2^3-2.3*X10^3-1.9*X8^3+2.6*X8^2*X9^2+27*X8^3*X11-12*X8^3*X9*X11-1.1*X8*X9^4-24*X8^2*X9^2*X11+11*X8^2*X9^3*X11-1.3E+002*X8^3*X11^2+1.1E+002*X8^3*X9*X11^2-25*X8^3*X9^2*X11^2+0.17*X9^6+5.5*X8*X9^4*X11-2.4*X8*X9^5*X11-51*X8^2*X9^3*X11^2+11*X8^2*X9^4*X11^2+2.E+002*X8^3*X11^3-2.7E+002*X8^3*X9*X11^3+1.2E+002*X8^3*X9^2*X11^3-18*X8^3*X9^3*X11^3-0.59*X2*X10+2.4*X2*X8^2-2.1*X2*X8*X9^2-23*X2*X8^2*X11+10*X2*X8^2*X9*X11+0.48*X2*X9^4+10*X2*X8*X9^2*X11-4.5*X2*X8*X9^3*X11+54*X2*X8^2*X11^2-48*X2*X8^2*X9*X11^2-26*X2*X8^2*X9^2*X11^2+0.37*X1^3*X2-0.15*X1*X2^2-0.42*X2^2*X8+0.19*X2^2*X9^2+2*X2^2*X8*X11-0.89*X2^2*X8*X9*X11+0.8*X1*X2^2*X8-0.36*X1*X2^2*X9^2-3.8*X1*X2^2*X8*X11+1.7*X1*X2^2*X8*X9*X11-1.5*X2*X10^2-5.9E-002*X2^4+1.5*X2*X10^3+1.2*X2*X8^3-1.6*X2*X8^2*X9^2-17*X2*X8^3*X11+7.7*X2*X8^3*X9*X11+0.74*X2*X8*X9^4+16*X2*X8^2*X9^2*X11-6.9*X2*X8^2*X9^3*X11+82*X2*X8^3*X11^2-73*X2*X8^3*X9*X11^2+16*X2*X8^3*X9^2*X11^2-0.11*X2*X9^6-3.5*X2*X8*X9^4*X11+1.6*X2*X8*X9^5*X11+33*X2*X8^2*X9^3*X11^2-7.3*X2*X8^2*X9^4*X11^2-1.3E+002*X2*X8^3*X11^3+1.7E+002*X2*X8^3*X9*X11^3-77*X2*X8^3*X9^2*X11^3+11*X2*X8^3*X9^3*X11^3+0.38*X2^2*X10+0.11*X6^2+0.23*X2^2*X8^2-0.21*X2^2*X8*X9^2-2.2*X2^2*X8^2*X11+0.96*X2^2*X8^2*X9*X11+4.6E-002*X2^2*X9^4+0.98*X2^2*X8*X9^2*X11-0.43*X2^2*X8*X9^3*X11+5.1*X2^2*X8^2*X11^2-4.6*X2^2*X8^2*X9*X11^2-2.7*X2^2*X8^2*X9^2*X11^2+3.4E-002*X1^3*X2^2-1.3E-002*X1*X2^3-4.5E-002*X2^3*X8+2.1E-002*X2^3*X9^2+0.22*X2^3*X8*X11-9.6E-002*X2^3*X8*X9*X11+7.2E-002*X1*X2^3*X8-3.2E-002*X1*X2^3*X9^2-0.34*X1*X2^3*X8*X11+0.15*X1*X2^3*X8*X9*X11-0.15*X2^2*X10^2-6.3E-003*X2^5+0.15*X2^2*X10^3+0.12*X2^2*X8^3-0.16*X2^2*X8^2*X9^2-1.7*X2^2*X8^3*X11+0.76*X2^2*X8^3*X9*X11+7.4E-002*X2^2*X8*X9^4+1.6*X2^2*X8^2*X9^2*X11-0.69*X2^2*X8^2*X9^3*X11+8.2*X2^2*X8^3*X11^2-7.3*X2^2*X8^3*X9*X11^2+1.6*X2^2*X8^3*X9^2*X11^2-1.1E-002*X2^2*X9^6-0.35*X2^2*X8*X9^4*X11+0.16*X2^2*X8*X9^5*X11+3.3*X2^2*X8^2*X9^3*X

Best formula (cont'd)	11^2-0.73*X2^2*X8^2*X9^4*X11^2- 13*X2^2*X8^3*X11^3+17*X2^2*X8^3*X9*X11^3- 7.6*X2^2*X8^3*X9^2*X11^3+1.1*X2^2*X8^3*X9^3*X11^3+3.8E- 002*X2^3*X10-3.3E-002*X2*X6+7.2E-002*X2^2*X6+0.99*X2*X6*X8^2- 0.89*X2*X6*X8*X9^2- 9.4*X2*X6*X8^2*X11+4.2*X2*X6*X8^2*X9*X11+0.2*X2*X6*X9^4+4.2*X2*X 6*X8*X9^2*X11-1.9*X2*X6*X8*X9^3*X11+22*X2*X6*X8^2*X11^2- 20*X2*X6*X8^2*X9*X11^2-17*X2*X6*X8^2*X9^2*X11^2+4.6E- 002*X1^3*X2*X6-1.8E-002*X1*X2^2*X6- 0.41*X2^2*X6*X8+0.19*X2^2*X6*X9^2+2*X2^2*X6*X8*X11- 0.87*X2^2*X6*X8*X9*X11+1.E-001*X1*X2^2*X6*X8-4.5E- 002*X1*X2^2*X6*X9^2- 0.47*X1*X2^2*X6*X8*X11+0.21*X1*X2^2*X6*X8*X9*X11+0.6*X2*X6*X8- 0.27*X2*X6*X9^2-2.8*X2*X6*X8*X11+1.3*X2*X6*X8*X9*X11-4.E- 002*X2^3*X6-0.87*X2*X6*X10^2-5.3E- 002*X2^4*X6+0.88*X2*X6*X10^3+0.71*X2*X6*X8^3- 0.96*X2*X6*X8^2*X9^2- 10*X2*X6*X8^3*X11+4.5*X2*X6*X8^3*X9*X11+0.43*X2*X6*X8*X9^4+9.1* X2*X6*X8^2*X9^2*X11- 4.1*X2*X6*X8^2*X9^3*X11+48*X2*X6*X8^3*X11^2- 43*X2*X6*X8^3*X9*X11^2+9.5*X2*X6*X8^3*X9^2*X11^2-6.5E- 002*X2*X6*X9^6- 2.1*X2*X6*X8*X9^4*X11+0.91*X2*X6*X8*X9^5*X11+19*X2*X6*X8^2*X9^3 *X11^2-4.3*X2*X6*X8^2*X9^4*X11^2- 76*X2*X6*X8^3*X11^3+1.E+002*X2*X6*X8^3*X9*X11^3- 45*X2*X6*X8^3*X9^2*X11^3+6.7*X2*X6*X8^3*X9^3*X11^3+0.22*X2^2*X6* X10-0.98*X2^2*X6*X8^2+0.89*X2^2*X6*X8*X9^2+9.3*X2^2*X6*X8^2*X11- 4.1*X2^2*X6*X8^2*X9*X11-0.2*X2^2*X6*X9^4- 4.2*X2^2*X6*X8*X9^2*X11+1.9*X2^2*X6*X8*X9^3*X11- 22*X2^2*X6*X8^2*X11^2+20*X2^2*X6*X8^2*X9*X11^2+11*X2^2*X6*X8^2* X9^2*X11^2-0.14*X1^3*X2^2*X6+5.7E-002*X1*X2^3*X6+0.2*X2^3*X6*X8- 8.8E-002*X2^3*X6*X9^2-0.93*X2^3*X6*X8*X11+0.41*X2^3*X6*X8*X9*X11- 0.31*X1*X2^3*X6*X8+0.14*X1*X2^3*X6*X9^2+1.5*X1*X2^3*X6*X8*X11- 0.65*X1*X2^3*X6*X8*X9*X11+0.63*X2^2*X6*X10^2+2.7E-002*X2^5*X6- 0.64*X2^2*X6*X10^3- 0.52*X2^2*X6*X8^3+0.7*X2^2*X6*X8^2*X9^2+7.4*X2^2*X6*X8^3*X11- 3.3*X2^2*X6*X8^3*X9*X11-0.32*X2^2*X6*X8*X9^4- 6.7*X2^2*X6*X8^2*X9^2*X11+3*X2^2*X6*X8^2*X9^3*X11- 35*X2^2*X6*X8^3*X11^2+31*X2^2*X6*X8^3*X9*X11^2- 6.9*X2^2*X6*X8^3*X9^2*X11^2+4.8E- 002*X2^2*X6*X9^6+1.5*X2^2*X6*X8*X9^4*X11- 0.67*X2^2*X6*X8*X9^5*X11- 14*X2^2*X6*X8^2*X9^3*X11^2+3.1*X2^2*X6*X8^2*X9^4*X11^2+56*X2^2* X6*X8^3*X11^3- 74*X2^2*X6*X8^3*X9*X11^3+33*X2^2*X6*X8^3*X9^2*X11^3- 4.9*X2^2*X6*X8^3*X9^3*X11^3-0.16*X2^3*X6*X10+0.15*X12^3- 0.17*X12*X13-2.5E-002*X6^3-0.11*X12^2+0.15*X6*X12

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	ECON	ECON	DES_VO	ECON	BOND	DES_VO
	PROCU	PROCU	ECON	BOND	PROCU	ECON
	SHAPE	SHAPE	COMP	COMP	SIZE	SHAPE
	SIZE	SIZE	PROCU	PROCU	TEAM	SIZE
	RELAT	TEAM	SHAPE	SHAPE	STA_PLT	TEAM
	STA_PLT	RELAT	SIZE	SIZE		RELAT
		STA_PLT	RELAT	TEAM		STA_PLT
			STA_PLT	RELAT		SC_FIN
				STA_PLT		SC_PER
				SC_FIN		
				SC_PER		

Note : Legend for the best formula:

X1 = (DES_VO-1)/4
X2 = (ECON-85.32)/26.43
X3 = (BOND-1)/4
X4 = (COMP-.75)/12.16
X5 = (PROCU-1)/4
X6 = (SHAPE-1)/4
X7 = (LOCAT-1)/4
X8 = (SIZE-3500)/179500
X9 = (TEAM-1)/4
X10 = (RELAT-1)/4
X11 = (STA_PLT-1)/4
X12 = (SC_FIN-1)/4
X13 = (SC_PER-1)/4
X14 = SUBSID
Y = (OUT_COST-427000)/129373000

Table I-2 Training, Testing and Production Results of ANN Models Using GMDH Algorithm (Based on 12 Inputs)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Number of inputs	12	12	12	12	12	12
Number of outputs	1	1	1	1	1	1
Number of training patterns	63	63	63	63	63	56
Number of test patterns	N/A	N/A	N/A	N/A	N/A	7
Number of production patterns	8	8	8	8	8	8
Layers constructed	3	2	6	9	3	7
Best criterion value	0.043802	0.045333	0.027667	0.016724	0.040992	0.000013
Training						
R squared	0.9122	0.9048	0.953	0.9613	0.9168	0.8239
Mean squared error	77,822,488,114,465	84,378,826,385,230	41,675,472,174,332	34,304,952,949,026	73,721,937,462,796	120,280,602,101,630
Mean absolute error	6,879,704	7,278,854	5,134,915	4,492,623	6,457,866	8,315,946
Min. absolute error	30,185	106,597	48,043	542,533	23,878	98,868
Max. absolute error	27,545,755	22,663,835	15,480,630	20,513,019	22,131,386	36,884,311
Correlation coefficient r	0.9551	0.9515	0.9762	**0.9805**	0.9575	0.9077
Testing						
R squared						1
Mean squared error						81499004571
Mean absolute error						254723.8413
Min. absolute error						116917.6952
Max. absolute error						553582.6949
Correlation coefficient r						1
Production						
R squared	0.9211	0.8684	0.8106	0.7056	**0.9538**	0.7194
Mean squared error	144,709,064,117,343	241,578,807,256,808	347,626,106,860,228	540,207,935,607,255	84,785,036,327,094	514,915,600,804,365
Mean absolute error	8,362,608	10,864,589	10,119,304	14,241,066	7,698,939	14,982,979
Min. absolute error	139,158	451,935	165,326	1,163,743	2,224,680	2,526,993
Max. absolute error	22,823,119	37,055,994	50,161,086	57,841,660	17,410,313	55,791,150
Correlation coefficient r	0.9647	0.9346	0.9217	0.8443	0.9774	0.8565

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Best formula	Y=0.15-0.5*X2-0.98*X7+0.19*X9+0.16*X10+3.4*X7*X10+0.38*X2^2+0.57*X5^2-0.66*X2*X5	Y=-4.5*X7^2+3.5*X7^3+4.5*X7*X8+5.1*X7*X10+0.51*X8*X10-12*X7*X8*X10	Y=-0.1*X4+3.8E-002+X7+0.15*X7^3+15*X7*X10-0.28*X5*X10+0.63*X10-3.5*X2*X10+5.6*X2^2*X10-2.6*X4^2*X10-2.7*X2^3*X10+2.7*X4^3*X10+1.3*X2*X4*X10-15*X2*X7*X10-18*X4*X7*X10+16*X2*X4*X7*X10+0.46*X5^2*X10+0.28*X4*X9-0.2*X8^2+0.23*X6*X8	Y=0.3+0.74*X8-1.2*X3+0.33*X4-0.27*X6+4*X7+0.17*X10+0.8*X6^2-1.5*X7^2-0.57*X6^3+1.4*X7^3-0.61*X10^3+0.29*X6*X10+4.1*X7*X10+0.62*X5^2-2.5*X8^2+2.5*X8^3-0.77*X5*X8+1.7*X3^2-1.6*X4^2+1.1*X4^3+0.66*X3*X4-0.12*X5-1.2*X2+1.9*X2^2-0.92*X2^3+0.3*X2*X4-4.2*X2*X7-5.6*X4*X7+4.6*X2*X4*X7-0.76*X3^3-0.52*X7*X8+3.5E-002*X9	Y=0.12-0.31*X2-2.5*X7+0.32*X9^2+1.7*X10^2+0.27*X7^3-2.2*X10^3+5.9*X7*X10+0.22*X2^3+0.43*X5^3-0.46*X2*X5	Y=-3.8E-002*X12-2.E-002+0.16*X1-2.4*X7+1.1*X8^2+12*X7*X10-14*X7*X8*X10-3.3*X7^2+3*X7*X8^2+31*X7^2*X10-14*X7^2*X8*X10-0.67*X8^4-10*X7*X8^2*X10+6.3*X7*X8^3*X10-74*X7^2*X10^2+66*X7^2*X8*X10^2+58*X7^2*X8^2*X10^2+2.2E-002*X1^3-2.4*X7^3+3.2*X7^2*X8^2+34*X7^3*X10-15*X7^3*X8*X10-1.5*X7*X8^4-31*X7^2*X8^2*X10+14*X7^2*X8^3*X10-1.6E+002*X7^3*X10^2+1.4E+002*X7^3*X8*X10^2-32*X7^3*X8^2*X10^2+0.22*X8^6+6.9*X7*X8^4*X10-3.1*X7*X8^5*X10-64*X7^2*X8^3*X10^2+14*X7^2*X8^4*X10^2+2.5E+002*X7^3*X10^3-3.4E+002*X7^3*X8*X10^3+1.5E+002*X7^3*X8^2*X10^3-22*X7^3*X8^3*X10^3-0.55*X1*X8+0.94*X1*X7-0.43*X1*X8^2-4.5*X1*X7*X10+19*X1*X7*X8*X10+1.9*X7*X8-0.87*X8^3-3.7*X1*X7*X8+1.6*X1*X8^3-7.7*X1*X7*X8^2*X10+0.21*X9^2+0.16*X9^3-2.4E-002*X1^2*X9+0.32*X1*X7*X9-0.15*X1*X8^2*X9-1.6*X1*X7*X9*X10+1.4*X1*X7*X8*X9*X10+0.48*X1*X7^2*X9-0.44*X1*X7*X8^2*X9-4.6*X1*X7^2*X9*X10+2*X1*X7^2*X8*X9*X10+9.8E-002*X1*X8^4*X9+1.8*X1*X7*X8^2*X9*X10-0.92*X1*X7*X8^3*X9*X10+11*X1*X7^2*X9*X10^2-9.7*X1*X7^2*X8*X9*X10^2-8.5*X1*X7^2*X8^2*X9*X10^2-3.3E-003*X1^4*X9+0.35*X1*X7^3*X9-0.47*X1*X7^2*X8^2*X9-5*X1*X7^3*X9*X10+2.2*X1*X7^3*X8*X9*X10+0.21*X1*X7*X8^4*X9+4.5*X1*X7^2*X8^2*X9*X10-2*X1*X7^2*X8^3*X9*X10+24*X1*X7^3*X9*X10^2-21*X1*X7^3*X8*X9*X10^2+4.7*X1*X7^3*X8^2*X9*X10^2-3.2E-002*X1*X8^6*X9-X1*X7*X8^4*X9*X10+0.45*X1*X7*X8^5*X9*X10+9.4*X1*X7^2*X8^3*X9*X10^2-2.1*X1*X7^2*X8^4*X9*X10^2-37*X1*X7^3*X9*X10^3+50*X1*X7^3*X8*X9*X10^3-22*X1*X7^3*X8^2*X9*X10^3+3.3*X1*X7^3*X8^3*X9*X10^3+8.1E-002*X1^2*X8*X9-0.14*X1^2*X7*X9+6.2E-002*X1^2*X8^2*X9+0.66*X1^2*X7*X9*X10-2.8*X1^2*X7*X8*X9*X10-0.15*X1*X7*X8*X9+6.7E-002*X1*X8^3*X9+0.54*X1^2*X7*X8*X9-0.24*X1^2*X8^3*X9+1.1*X1^2*X7*X8^2*X9*X10+3.E-003*X12^3+1.3E-003*X9-6.1E-003*X1*X9+9.E-002*X7*X9-4.1E-002*X8^2*X9-0.44*X7*X9*X10+0.53*X7*X8*X9*X10+0.12*X7^2*X9-0.11*X7*X8^2*X9-1.2*X7^2*X9*X10+0.51*X7^2*X8*X9*X10+2.5E-002*X8^4*X9+0.37*X7*X8^2*X9*X10-0.23*X7*X8^3*X9*X10+2.8*X7^2*X9*X10^2-2.4*X7^2*X8*X9*X10^2-2.1*X7^2*X8^2*X9*X10^2-8.3E-004*X1^3*X9+8.8E-002*X7^3*X9-0.12*X7^2*X8^2*X9-1.3*X7^3*X9*X10+0.56*X7^3*X8*X9*X10+5.4E-002*X7*X8^4*X9+1.1*X7^2*X8^2*X9*X10-0.5*X7^2*X8^3*X9*X10+5.9*X7^3*X9*X10^2-5.3*X7^3*X8*X9*X10^2+1.2*X7^3*X8^2*X9*X10^2-8.E-003*X8^6*X9-0.25*X7*X8^4*X9*X10+0.11*X7*X8^5*X9*X10+2.4*X7^2*X8^3*X9*X10^2-0.53*X7^2*X8^4*X9*X10^2-9.4*X7^3*X9*X10^3+13*X7^3*X8*X9*X10^3-

Best formula (cont'd)	5.6*X7^3*X8^2*X9*X10^3+0.82*X7^3*X8^3*X9*X10^3+2.E-002*X1*X8*X9-7.1E-002*X7*X8*X9+3.2E-002*X8^3*X9-7.E-004*X9^4+1.3E-003*X1^2*X9^2-1.9E-002*X1*X7*X9^2+8.4E-003*X1*X8^2*X9^2+9.E-002*X1*X7*X9^2*X10-0.11*X1*X7*X8*X9^2*X10-2.5E-002*X1*X7^2*X9^2+2.3E-002*X1*X7*X8^2*X9^2+0.24*X1*X7^2*X9^2*X10-0.11*X1*X7^2*X8*X9^2*X10-5.1E-003*X1*X8^4*X9^2-7.7E-002*X1*X7*X8^2*X9^2*X10+4.8E-002*X1*X7*X8^3*X9^2*X10-0.57*X1*X7^2*X9^2*X10^2+0.5*X1*X7^2*X8*X9^2*X10^2+0.44*X1*X7^2*X8^2*X9^2*X10^2+1.7E-004*X1^4*X9^2-1.8E-002*X1*X7^3*X9^2+2.5E-002*X1*X7^2*X8^2*X9^2+0.26*X1*X7^3*X9^2*X10-0.11*X1*X7^3*X8*X9^2*X10-1.1E-002*X1*X7*X8^4*X9^2-0.23*X1*X7^2*X8^2*X9^2*X10+0.1*X1*X7^2*X8^3*X9^2*X10-1.2*X1*X7^3*X9^2*X10^2+1.1*X1*X7^3*X8*X9^2*X10^2-0.24*X1*X7^3*X8^2*X9^2*X10^2+1.7E-003*X1*X8^6*X9^2+5.3E-002*X1*X7*X8^4*X9^2*X10-2.3E-002*X1*X7*X8^5*X9^2*X10-0.49*X1*X7^2*X8^3*X9^2*X10^2+0.11*X1*X7^2*X8^4*X9^2*X10^2+1.9*X1*X7^3*X9^2*X10^3-2.6*X1*X7^3*X8*X9^2*X10^3+1.1*X1*X7^3*X8^2*X9^2*X10^3-0.17*X1*X7^3*X8^3*X9^2*X10^3-4.2E-003*X1^2*X8*X9^2+7.2E-003*X1^2*X7*X9^2-3.2E-003*X1^2*X8^2*X9^2-3.4E-002*X1^2*X7*X9^2*X10+0.15*X1^2*X7*X8*X9^2*X10+1.5E-002*X1*X7*X8*X9^2-6.6E-003*X1*X8^3*X9^2-2.8E-002*X1^2*X7*X8*X9^2+1.3E-002*X1^2*X8^3*X9^2-5.9E-002*X1^2*X7*X8^2*X9^2*X10-2.3E-003*X1*X5+3.5E-002*X5*X7-1.6E-002*X5*X8^2-0.16*X5*X7*X10+0.2*X5*X7*X8*X10+4.7E-002*X5*X7^2-4.2E-002*X5*X7*X8^2-0.44*X5*X7^2*X10+0.2*X5*X7^2*X8*X10+9.5E-003*X5*X8^4+0.14*X5*X7*X8^2*X10-8.9E-002*X5*X7*X8^3*X10+1.1*X5*X7^2*X10^2-0.94*X5*X7^2*X8*X10^2-0.82*X5*X7^2*X8^2*X10^2-3.2E-004*X1^3*X5+3.4E-002*X5*X7^3-4.6E-002*X5*X7^2*X8^2-0.48*X5*X7^3*X10+0.21*X5*X7^3*X8*X10+2.1E-002*X5*X7*X8^4+0.43*X5*X7^2*X8^2*X10-0.19*X5*X7^2*X8^3*X10+2.3*X5*X7^3*X10^2-2*X5*X7^3*X8*X10^2+0.45*X5*X7^3*X8^2*X10^2-3.1E-003*X5*X8^6-9.7E-002*X5*X7*X8^4*X10+4.3E-002*X5*X7*X8^5*X10+0.91*X5*X7^2*X8^3*X10^2-0.2*X5*X7^2*X8^4*X10^2-3.6*X5*X7^3*X10^3+4.8*X5*X7^3*X8*X10^3-2.1*X5*X7^3*X8^2*X10^3+0.32*X5*X7^3*X8^3*X10^3+7.8E-003*X1*X5*X8-1.3E-002*X1*X5*X7+6.E-003*X1*X5*X8^2+6.3E-002*X1*X5*X7*X10-0.27*X1*X5*X7*X8*X10-2.7E-002*X5*X7*X8+1.2E-002*X5*X8^3+5.2E-002*X1*X5*X7*X8-2.3E-002*X1*X5*X8^3+0.11*X1*X5*X7*X8^2*X10-2.2E-003*X11*X12-0.19*X7*X11*X12+1.3E-002*X1*X11*X12+8.9E-002*X8^2*X11*X12+0.94*X7*X10*X11*X12-1.2*X7*X8*X10*X11*X12-0.27*X7^2*X11*X12+0.24*X7*X8^2*X11*X12+2.5*X7^2*X10*X11*X12-1.1*X7^2*X8*X10*X11*X12-5.4E-002*X8^4*X11*X12-0.82*X7*X8^2*X10*X11*X12+0.51*X7*X8^3*X10*X11*X12-6*X7^2*X10^2*X11*X12+5.4*X7^2*X8*X10^2*X11*X12+4.7*X7^2*X8^2*X10^2*X11*X12+1.8E-003*X1^3*X11*X12-

Best formula (cont'd)	0.19*X7^3*X11*X12+0.26*X7^2*X8^2*X11*X12+2.8*X7^3*X10*X11*X12- 1.2*X7^3*X8*X10*X11*X12-0.12*X7*X8^4*X11*X12- 2.5*X7^2*X8^2*X10*X11*X12+1.1*X7^2*X8^3*X10*X11*X12- 13*X7^3*X10^2*X11*X12+12*X7^3*X8*X10^2*X11*X12- 2.6*X7^3*X8^2*X10^2*X11*X12+1.8E- 002*X8^6*X11*X12+0.56*X7*X8^4*X10*X11*X12- 0.25*X7*X8^5*X10*X11*X12- 5.2*X7^2*X8^3*X10^2*X11*X12+1.2*X7^2*X8^4*X10^2*X11*X12+21*X7^3* X10^3*X11*X12- 28*X7^3*X8*X10^3*X11*X12+12*X7^3*X8^2*X10^3*X11*X12- 1.8*X7^3*X8^3*X10^3*X11*X12-4.5E-002*X1*X8*X11*X12+7.7E- 002*X1*X7*X11*X12-3.4E-002*X1*X8^2*X11*X12- 0.36*X1*X7*X10*X11*X12+1.6*X1*X7*X8*X10*X11*X12+0.16*X7*X8*X11* X12-7.1E-002*X8^3*X11*X12- 0.3*X1*X7*X8*X11*X12+0.13*X1*X8^3*X11*X12- 0.62*X1*X7*X8^2*X10*X11*X12-6.5E-004*X1*X5*X11*X12+9.7E- 003*X5*X7*X11*X12-4.4E-003*X5*X8^2*X11*X12-4.6E- 002*X5*X7*X10*X11*X12+5.7E-002*X5*X7*X8*X10*X11*X12+1.3E- 002*X5*X7^2*X11*X12-1.2E-002*X5*X7*X8^2*X11*X12- 0.12*X5*X7^2*X10*X11*X12+5.5E-002*X5*X7^2*X8*X10*X11*X12+2.6E- 003*X5*X8^4*X11*X12+4.E-002*X5*X7*X8^2*X10*X11*X12-2.5E- 002*X5*X7*X8^3*X10*X11*X12+0.29*X5*X7^2*X10^2*X11*X12- 0.26*X5*X7^2*X8*X10^2*X11*X12-0.23*X5*X7^2*X8^2*X10^2*X11*X12- 8.9E-005*X1^3*X5*X11*X12+9.4E-003*X5*X7^3*X11*X12-1.3E- 002*X5*X7^2*X8^2*X11*X12-0.13*X5*X7^3*X10*X11*X12+5.9E- 002*X5*X7^3*X8*X10*X11*X12+5.7E- 003*X5*X7*X8^4*X11*X12+0.12*X5*X7^2*X8^2*X10*X11*X12-5.4E- 002*X5*X7^2*X8^3*X10*X11*X12+0.64*X5*X7^3*X10^2*X11*X12- 0.56*X5*X7^3*X8*X10^2*X11*X12+0.13*X5*X7^3*X8^2*X10^2*X11*X12- 8.6E-004*X5*X8^6*X11*X12-2.7E-002*X5*X7*X8^4*X10*X11*X12+1.2E- 002*X5*X7*X8^5*X10*X11*X12+0.25*X5*X7^2*X8^3*X10^2*X11*X12-5.6E- 002*X5*X7^2*X8^4*X10^2*X11*X12- X5*X7^3*X10^3*X11*X12+1.3*X5*X7^3*X8*X10^3*X11*X12- 0.59*X5*X7^3*X8^2*X10^3*X11*X12+8.8E- 002*X5*X7^3*X8^3*X10^3*X11*X12+2.2E-003*X1*X5*X8*X11*X12-3.7E- 003*X1*X5*X7*X11*X12+1.7E-003*X1*X5*X8^2*X11*X12+1.8E- 002*X1*X5*X7*X10*X11*X12-7.6E-002*X1*X5*X7*X8*X10*X11*X12-7.6E- 003*X5*X7*X8*X11*X12+3.4E-003*X5*X8^3*X11*X12+1.4E- 002*X1*X5*X7*X8*X11*X12-6.5E-003*X1*X5*X8^3*X11*X12+3.E- 002*X1*X5*X7*X8^2*X10*X11*X12+3.3E-002*X9^3*X11*X12-9.E-004*X12^2

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	ECON	SIZE	ECON	ECON	ECON	DES_VO
	PROCU	TEAM	COMP	BOND	PROCU	PROCU
	SIZE	STA_PLT	PROCU	COMP	SIZE	SIZE
	RELAT		SHAPE	PROCU	RELAT	TEAM
	STA_PLT		SIZE	SHAPE	STA_PLT	RELAT
			TEAM	SIZE		STA_PLT
			RELAT	TEAM		SC_FIN
			STA_PLT	RELAT		SC_PER
				STA_PLT		

Note : Legend for the best formula:

X1 = (DES_VO-1)/4
X2 = (ECON-85.32)/26.43
X3 = (BOND-1)/4
X4 = (COMP-.75)/12.16
X5 = (PROCU-1)/4
X6 = (SHAPE-1)/4
X7 = (SIZE-3500)/179500
X8 = (TEAM-1)/4
X9 = (RELAT-1)/4
X10 = (STA_PLT-1)/4
X11 = (SC_FIN-1)/4
X12 = (SC_PER-1)/4
Y = (OUT_COST-427000)/129373000

Table I-3 Training, Testing and Production Results of ANN Models Using GMDH Algorithm (Based on 8 Inputs)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Number of inputs	8	8	8	8	8	8
Number of outputs	1	1	1	1	1	1
Number of training patterns	63	63	63	63	63	56
Number of test patterns	N/A	N/A	N/A	N/A	N/A	7
Number of production patterns	8	8	8	8	8	8
Layers constructed	3	3	8	7	3	5
Best criterion value	0.043802	0.044968	0.027313	0.024231	0.037442	0.000702
Training						
R squared	0.9122	0.9118	**0.9523**	0.9394	0.9189	0.8
Mean squared error	77,822,488,114,465	78,170,758,193,515	42,225,889,930,043	53,656,127,136,067	71,870,180,215,038	136,601,714,453,291
Mean absolute error	6,879,704	6,935,211	5,079,790	5,348,331	6,535,824	9,108,978
Min. absolute error	30,185	163,096	89,237	175,689	174,204	25,126
Max. absolute error	27,545,755	21,720,127	18,397,422	26,472,488	22,132,755	34,254,865
Correlation coefficient r	0.9551	0.9549	0.9759	0.9692	0.9588	0.8944
Testing						
R squared						0.9976
Mean squared error						4,335,847,355,564
Mean absolute error						1,640,547
Min. absolute error						269,120
Max. absolute error						4,043,283
Correlation coefficient r						0.9989
Production						
R squared	**0.9211**	0.8383	0.874	0.8804	0.826	0.89
Mean squared error	144,709,064,117,343	296,796,808,239,102	231,230,554,118,774	219,440,462,308,019	319,218,514,112,100	201,776,558,812,512
Mean absolute error	8,362,608	12,043,370	7,055,653	9,289,936	11,951,075	9,386,480
Min. absolute error	139,158	57,699	41,721	159,193	1,808,509	1,137,730
Max. absolute error	22,823,119	37,488,733	42,542,597	37,368,309	41,934,351	34,699,619
Correlation coefficient r	0.9647	0.9178	0.9472	0.9723	0.9133	0.9529

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Best formula	Y=0.15-0.5*X1-0.98*X5+0.19*X7+0.16*X8+3.4*X5*X8+0.38*X1^2+0.57*X3^2-0.66*X1*X3	Y=-4.5*X5^2+3.5*X5^3+4.4*X5*X6+5*X5*X8+0.51*X6*X8-11*X5*X6*X8+8.8E-002*X3^2	Y=0.76*X8-0.13*X2+0.27*X5^2-0.49*X8^3-0.47*X3*X8-3.1*X1*X8+15*X5*X8+4.8*X1^2*X8-2.3*X2^2*X8-2.4*X1^3*X8+2.4*X2^3*X8+1.1*X1*X2*X8-13*X1*X5*X8-16*X2*X5*X8+14*X1*X2*X5*X8+0.61*X3^2*X8+0.26*X8^4+0.3*X2*X7+7.9E-002*X4^2	Y=1.9*X5+0.29*X6+0.14*X7+2.7E-002+4.2E-002*X2-0.59*X3+3*X3^2+4*X5^2+0.98*X2^3-3.1*X3^3-1.6*X5^3-3.5*X2*X3-5.9*X2*X5+24*X3*X5-16*X2*X3*X5-4.9E-002*X4^2+0.45*X2*X4+1.6*X5*X8-0.29*X1^2+0.25*X1^3-1.2*X6^2+1.2*X6^3	Y=-4.4*X5^2+3.4*X5^3+4.4*X5*X6+4.9*X5*X8+0.5*X6*X8-11*X5*X6*X8+0.25*X3^3+0.25*X7^3-0.41*X3*X7	Y=6.6E-002-1.4*X5+0.65*X6^2+6.8*X5*X8-4.7*X5*X6*X8+0.16*X7^2-2.2*X5^2+2*X5*X6^2+21*X5^2*X8-9.5*X5^2*X6*X8-0.46*X6^4-8.9*X5*X6^2*X8+4.3*X5*X6^3*X8-50*X5^2*X8^2+45*X5^2*X6*X8^2+38*X5^2*X6^2*X8^2+0.16*X2^3-1.5*X5^3+2.1*X5^2*X6^2+22*X5^3*X8-9.9*X5^3*X6*X8-0.96*X5*X6^4-20*X5^2*X6^2*X8+8.9*X5^2*X6^3*X8-1.1E+002*X5^3*X8^2+95*X5^3*X6*X8^2-21*X5^3*X6^2*X8^2+0.14*X6^6+4.5*X5*X6^4*X8-2*X5*X6^5*X8-42*X5^2*X6^3*X8^2+9.4*X5^2*X6^4*X8^2+1.7E+002*X5^3*X8^3-2.3E+002*X5^3*X6*X8^3+99*X5^3*X6^2*X8^3-15*X5^3*X6^3*X8^3+0.33*X7^3-0.13*X1^2*X5^2+0.11*X1^3*X5^2+0.14*X1*X5^3*X8+7.6E-002*X1^4*X5-0.13*X1^5*X5+0.87*X1^2*X5^2*X8-0.93*X1^3*X5^2*X8+5.9E-002*X1^6*X5+0.14*X1^4*X5^2*X8-0.93*X1*X5^3*X8^2+8.6E-002*X1^2*X5^3*X8^2-1.5E-002*X1^6+4.E-002*X1^7-0.26*X1^4*X5*X8+0.5*X1^5*X5*X8-3.5E-002*X1^8-0.28*X1^6*X5*X8-1.5*X1^2*X5^2*X8^2+1.8*X1^3*X5^2*X8^2-0.53*X1^4*X5^2*X8^2+1.E-002*X1^9+3.7E-002*X1^7*X5*X8+4.5E-002*X1^5*X5^2*X8^2+1.6*X1*X5^3*X8^3-0.29*X1^2*X5^3*X8^3+1.8E-002*X1^3*X5^3*X8^3+7.E-002*X5*X7-4.1E-002*X1^2*X7+3.6E-002*X1^3*X7-0.23*X5*X7*X8+4.4E-002*X1*X5*X7*X8+1.6E-004*X5^4-0.15*X5*X8^3+2.9E-004*X4*X5*X8-2.3E-004*X2*X5+0.11*X2*X5^2-2.4E-002*X2*X5*X8-5.2E-003*X2*X5^4-0.18*X2*X5*X8^3-9.1E-003*X2*X4*X5*X8-0.82*X2*X5^2*X8+5.2E-003*X5*X8^2+1.4E-003*X5^4*X8+0.61*X5*X8^4+2.4E-003*X4*X5*X8^2-3.3E-002*X5^4*X8^2-0.25*X5*X8^5-5.7E-002*X4*X5*X8^3-3.9*X5^2*X8^3+0.13*X5^4*X8^3-1.1*X5*X8^6+0.23*X4*X5*X8^4+9.5*X5^2*X8^4-0.78*X2*X5^3+3.7E-002*X2*X5^5-1.4*X2*X5^2*X8^3+6.5E-002*X2*X4*X5^2*X8+6.9*X2*X5^3*X8+0.16*X2*X5*X8^2+3.4E-002*X2*X5^4*X8-0.29*X2*X5*X8^4+6.E-002*X2*X4*X5*X8^2+1.7*X2*X5^2*X8^2-2.4E-002*X5^5*X8-4.2E-002*X4*X5^2*X8^2-0.24*X2*X5^5*X8+2*X2*X5^2*X8^4-0.41*X2*X4*X5^2*X8^2-19*X2*X5^3*X8^2-4.E-004*X6^2*X8-7.4E-005*X5^3*X6^2+4.9E-002*X6^2*X8^3-1.3E-004*X4*X6^2*X8+1.E-004*X2*X6^2-5.E-002*X2*X5*X6^2+1.1E-002*X2*X6^2*X8+2.3E-003*X2*X5^3*X6^2-2.E-002*X2*X6^2*X8^3+4.1E-003*X2*X4*X6^2*X8-1.7*X2*X5*X6^2*X8-4.3E-003*X6^2*X8^2-9.7E-004*X5^3*X6^2*X8-4.2E-002*X6^2*X8^4-1.7E-003*X4*X6^2*X8^2-0.21*X5*X6^2*X8^2-8.7E-002*X6^2*X8^5+1.8E-002*X4*X6^2*X8^3+0.76*X5*X6^2*X8^3+8.9E-002*X6^2*X8^6-1.8E-002*X4*X6^2*X8^4-0.65*X5*X6^2*X8^4+0.35*X2*X5^2*X6^2-1.7E-002*X2*X5^4*X6^2+0.14*X2*X5*X6^2*X8^3-2.9E-

Best formula (cont'd)	002*X2*X4*X5*X6^2*X8-1.5*X2*X5^2*X6^2*X8-2.1E- 002*X2*X6^2*X8^2-4.3E-003*X2*X5^3*X6^2*X8+3.7E- 002*X2*X6^2*X8^4-7.6E-003*X2*X4*X6^2*X8^2- 0.1*X2*X5*X6^2*X8^2+1.1E-002*X5^4*X6^2*X8+1.9E- 002*X4*X5*X6^2*X8^2+2.7E-002*X2*X5^4*X6^2*X8- 0.23*X2*X5*X6^2*X8^4+4.7E- 002*X2*X4*X5*X6^2*X8^2+1.5*X2*X5^2*X6^2*X8^2- 0.11*X5^4*X8^4+0.94*X5*X8^7-0.19*X4*X5*X8^5-6.9*X5^2*X8^5- 4.6E-002*X2*X5^4*X8^2+0.39*X2*X5*X8^5-8.E- 002*X2*X4*X5*X8^3+0.11*X5^5*X8^2+0.2*X4*X5^2*X8^3+0.28*X2 *X5^5*X8^2- 2.4*X2*X5^2*X8^5+0.5*X2*X4*X5^2*X8^3+15*X2*X5^3*X8^3+1.9E -003*X5*X6*X8^2+3.5E-004*X5^4*X6*X8-0.23*X5*X6*X8^4+6.1E- 004*X4*X5*X6*X8^2+4.1*X2*X5*X6*X8+0.23*X2*X5^2*X6*X8- 5.2E-002*X2*X5*X6*X8^2-1.1E-002*X2*X5^4*X6*X8+9.3E- 002*X2*X5*X6*X8^4-1.9E-002*X2*X4*X5*X6*X8^2- 0.64*X2*X5^2*X6*X8^2+2.E-002*X5*X6*X8^3+4.5E- 003*X5^4*X6*X8^2+0.2*X5*X6*X8^5+7.9E- 003*X4*X5*X6*X8^3+0.99*X5^2*X6*X8^3-4.7E- 002*X5^4*X6*X8^3+0.4*X5*X6*X8^6-8.3E-002*X4*X5*X6*X8^4- 3.6*X5^2*X6*X8^4+4.9E-002*X5^4*X6*X8^4- 0.42*X5*X6*X8^7+8.5E-002*X4*X5*X6*X8^5+3.1*X5^2*X6*X8^5- 1.6*X2*X5^3*X6*X8+7.8E-002*X2*X5^5*X6*X8- 0.67*X2*X5^2*X6*X8^4+0.14*X2*X4*X5^2*X6*X8^2+6.9*X2*X5^3 *X6*X8^2+9.8E-002*X2*X5*X6*X8^3+2.E-002*X2*X5^4*X6*X8^2- 0.17*X2*X5*X6*X8^5+3.5E- 002*X2*X4*X5*X6*X8^3+0.47*X2*X5^2*X6*X8^3-5.1E- 002*X5^5*X6*X8^2-8.9E-002*X4*X5^2*X6*X8^3- 0.13*X2*X5^5*X6*X8^2+1.1*X2*X5^2*X6*X8^5- 0.22*X2*X4*X5^2*X6*X8^3-6.8*X2*X5^3*X6*X8^3+1.E- 004*X3*X5-4.9E-002*X3*X5^2+1.1E-002*X3*X5*X8+2.4E- 003*X3*X5^4-2.E-002*X3*X5*X8^3+4.1E- 003*X3*X4*X5*X8+0.36*X3*X5^2*X8-4.6E-005*X3*X6^2+2.2E- 002*X3*X5*X6^2-5.2E-003*X3*X6^2*X8-1.1E- 003*X3*X5^3*X6^2+9.1E-003*X3*X6^2*X8^3-1.9E- 003*X3*X4*X6^2*X8-5.7E-002*X3*X5*X6^2*X8-5.4E- 002*X3*X5*X8^2-1.1E-002*X3*X5^4*X8+9.6E-002*X3*X5*X8^4- 2.E-002*X3*X4*X5*X8^2-0.6*X3*X5^2*X8^2+2.2E- 004*X3*X5*X6*X8-0.1*X3*X5^2*X6*X8+2.4E- 002*X3*X5*X6*X8^2+5.E-003*X3*X5^4*X6*X8-4.3E- 002*X3*X5*X6*X8^4+8.7E- 003*X3*X4*X5*X6*X8^2+0.27*X3*X5^2*X6*X8^2-0.35*X1+4.7E- 002*X6+0.7*X1^2-0.38*X1^3-7.3E-002*X1*X6- 0.36*X2*X6+0.36*X5*X6-0.16*X6^3-0.87*X2*X5*X6+0.39*X2*X6^3

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	ECON	PROCU	ECON	ECON	PROCU	ECON
	PROCU	SIZE	COMP	COMP	SIZE	COMP
	SIZE	TEAM	PROCU	PROCU	TEAM	PROCU
	RELAT	STA_PLT	SHAPE	SHAPE	RELAT	SHAPE
	STA_PLT		SIZE	SIZE	STA_PLT	SIZE
			RELAT	TEAM		TEAM
			STA_PLT	RELAT		RELAT
				STA_PLT		STA_PLT

Note : Legend for the best formula :

X1 = (ECON-85.32)/26.43

X2 = (COMP-.75)/12.16

X3 = (PROCU-1)/4

X4 = (SHAPE-1)/4

X5 = (SIZE-3500)/179500

X6 = (TEAM-1)/4

X7 = (RELAT-1)/4

X8 = (STA_PLT-1)/4

Y = (OUT_COST-427000) /129373000

Table I-4 Training, Testing and Production Results of ANN Models Using GMDH Algorithm (Based on 5 Inputs)

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Number of inputs	5	5	5	5	5	5
Number of outputs	1	1	1	1	1	1
Number of training patterns	63	63	63	63	63	56
Number of test patterns	N/A	N/A	N/A	N/A	N/A	7
Number of production patterns	8	8	8	8	8	8
Layers constructed	4	5	10	6	3	6
Best criterion value	0.048139	0.054191	0.034613	0.026185	0.040992	0.00304
Training						
R squared	0.8818	0.8911	0.9329	0.9323	0.9168	0.8221
Mean squared error	104,745,069,896,220	96,499,397,871,397	59,451,592,169,420	59,978,555,243,060	73,721,937,462,796	121,553,779,209,170
Mean absolute error	8,033,736	7,865,237	5,715,629	5,777,226	6,457,866	8,552,664
Min. absolute error	229,921	472,578	24,665	294,636	23,878	188,346
Max. absolute error	23,241,907	29,173,063	21,373,383	20,117,189	22,131,386	27,439,840
Correlation coefficient r	0.939	0.945	0.9659	0.9656	0.9575	0.9067
Testing						
R squared						0.9895
Mean squared error						18,768,174,479,400
Mean absolute error						3,391,587
Min. absolute error						1,015,045
Max. absolute error						8,298,346
Correlation coefficient r						0.995
Production						
R squared	0.9754	0.9451	0.8182	0.9152	0.9538	0.9919
Mean squared error	45,204,895,159,568	100,818,018,331,372	333,638,465,010,113	155,689,061,121,611	84,785,036,327,094	14,816,667,841,587
Mean absolute error	4,792,928	7,488,654	12,073,577	9,286,564	7,698,939	3,093,530
Min. absolute error	1,154,702	694,825	2,971,514	1,805,281	2,224,680	758,091
Max. absolute error	15,368,510	19,618,748	46,525,880	28,839,723	17,410,313	7,578,644
Correlation coefficient r	0.9885	0.9867	0.9062	0.9575	0.9774	0.996

Network type	1st - FCPSE	2nd - PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Best formula	Y=-0.92*X3+0.21*X4+0.17*X5+3.7*X3*X5-0.44*X1*X3	Y=-2*X3+0.52*X5+5.2E-002*X2^3-0.81*X5^3+5.3*X3*X5+0.34*X2^2-0.43*X1*X2+0.36*X4^3	Y=2.4*X3+1.8*X5^3+7.3E-002*X3*X5-6.4E-002*X5+0.23*X4*X5-14*X3*X5^2+0.67*X5^2-2.4*X1*X5^2+0.97*X2^2*X5^2-5.5*X3^2*X5^2+4.6*X3^3*X5^2-2.2*X5^5-1.2*X2*X5^3+18*X3*X5^3+1.7*X1^2*X5^2+X4^2*X5^2+4.4E-002*X1^2*X5+0.31*X2^2*X5-0.45*X1*X2*X5-1.6*X3*X4*X5	Y=3*X3+0.45*X5-1.1*X5^2+3.9*X5^3-8.8*X3*X5-0.69*X1*X5+1.8*X3^3*X5+0.76*X4^3*X5-3.5*X5^4+11*X3*X5^2+0.64*X4*X5^2-2.5*X3*X4*X5^2+0.48*X1^3*X5+0.76*X2^3*X5-0.82*X1*X2*X5-2.3*X3^2*X5-0.32*X1*X3*X5	Y=0.12-0.31*X1-2.5*X3+0.32*X4^2+1.7*X5^2+0.27*X3^3-2.2*X5^3+5.9*X3*X5+0.22*X1^3+0.43*X2^3-0.46*X1*X2	Y=0.4-2*X1+0.41*X3+3.4*X1^2-1.7*X1^3+0.54*X4^3+4.E-002*X3*X4+8.E-002*X5+4.E-003*X2^2+1.6*X3^3-3.2E-002*X5^3-2.3*X3*X5+5.8*X3^2-0.32*X2^2*X3-1.1*X3^4+4.9*X3*X5^3-26*X3^2*X5+0.23*X5^2+6.8E-002*X2^2*X5-3.5*X3^3*X5-0.92*X5^4+5.2*X3*X5^2+4.5E-003*X2^4+3.1E-002*X2^2*X3^3-0.12*X2^2*X5^3+0.7*X2^2*X3*X5+5.3E-002*X3^6-0.42*X3^3*X5^3+2.4*X3^4*X5+0.81*X5^6-9.4*X3*X5^4+27*X3^2*X5^2+0.12*X4-5.1E-002*X4*X5-6.E-003*X2^2*X4-0.43*X3^3*X4+8.8E-002*X4*X5^3-1.4*X3*X4*X5+0.63*X1*X3-0.39*X1*X5-0.6*X1*X4-5.6*X1*X3^2-3.7*X1*X3*X5-0.98*X1*X3*X4+1.2*X1*X5^2+0.25*X1*X4*X5-1.9E-002*X1*X2^2+0.17*X1*X2^2*X3+0.12*X1*X2^2*X5+3.E-002*X1*X2^2*X4-6.6E-002*X1*X3^3+0.59*X1*X3^4+0.4*X1*X3^3*X5+0.1*X1*X3^3*X4+0.37*X1*X5^3-1.4*X1*X3*X5^3-1.5*X1*X5^4-0.37*X1*X4*X5^3+14*X1*X3^2*X5+9.6*X1*X3*X5^2+1.8*X1*X3*X4*X5+2.1*X3^2*X4-4.5E-002*X2^2*X3^2-0.11*X2^2*X3*X4-0.16*X3^5-0.37*X3^4*X4+3.4*X3^2*X5^3+1.6*X3*X4*X5^3-9*X3^2*X4*X5-1.1*X1^2*X3+0.65*X1^2*X5+X1^2*X4+9.4*X1^2*X3^2+6.2*X1^2*X3*X5+1.6*X1^2*X3*X4-2*X1^2*X5^2-0.42*X1^2*X4*X5+3.2E-002*X1^2*X2^2-0.29*X1^2*X2^2*X3-0.19*X1^2*X2^2*X5-5.E-002*X1^2*X2^2*X4+0.11*X1^2*X3^3-0.99*X1^2*X3^4-0.67*X1^2*X3^3*X5-0.17*X1^2*X3^3*X4-0.63*X1^2*X5^3+2.3*X1^2*X3*X5^3+2.6*X1^2*X5^4+0.62*X1^2*X4*X5^3-23*X1^2*X3^2*X5-16*X1^2*X3*X5^2-3.1*X1^2*X3*X4*X5+0.54*X1^3*X3-0.33*X1^3*X5-0.5*X1^3*X4-4.7*X1^3*X3^2-3.1*X1^3*X3*X5-0.83*X1^3*X3*X4+X1^3*X5^2+0.21*X1^3*X4*X5-1.6E-002*X1^3*X2^2+0.14*X1^3*X2^2*X3+9.7E-002*X1^3*X2^2*X5+2.5E-002*X1^3*X2^2*X4-5.6E-002*X1^3*X3^3+0.5*X1^3*X3^4+0.34*X1^3*X3^3*X5+8.7E-002*X1^3*X3^3*X4+0.32*X1^3*X5^3-1.2*X1^3*X3*X5^3-1.3*X1^3*X5^4-0.31*X1^3*X4*X5^3+12*X1^3*X3^2*X5+8.1*X1^3*X3*X5^2+1.6*X1^3*X3*X4*X5-0.1*X3*X4^3+6.2E-002*X4^3*X5+9.5E-002*X4^4+0.84*X3^2*X4^3+0.6*X3*X4^3*X5+0.16*X3*X4^4-0.19*X4^3*X5^2-4.E-002*X4^4*X5+3.1E-003*X2^2*X4^3-2.6E-002*X2^2*X3*X4^3-1.8E-002*X2^2*X4^3*X5-4.8E-003*X2^2*X4^4+1.1E-002*X3^3*X4^3-8.8E-002*X3^4*X4^3-6.3E-002*X3^3*X4^3*X5-1.6E-002*X3^3*X4^4-5.9E-002*X4^3*X5^3+0.2*X3*X4^3*X5^3+0.25*X4^3*X5^4+5.9E-002*X4^4*X5^3-2.1*X3^2*X4^3*X5-1.5*X3*X4^3*X5^2-0.29*X3*X4^4*X5-2.3E-002*X3*X4^2-6.2E-002*X3^2*X4^2+1.6*X3*X4*X5^2-0.16*X3*X4^2*X5+1.3E-

Best formula (cont'd)	002*X2^2*X3^2*X4+0.22*X2^2*X3*X4*X5+1.9E- 003*X2^2*X3*X4^2+4.6E-002*X3^5*X4+0.76*X3^4*X4*X5+6.5E- 003*X3^4*X4^2+0.79*X3^2*X4*X5^3-3*X3*X4*X5^4-2.E- 002*X3*X4^2*X5^3+1.2*X3^3*X4*X5+9.3*X3^2*X4*X5^2+6.E- 002*X3^2*X4^2*X5+6.4E-003*X2^2*X5^2-0.34*X3^3*X5^2-8.7E- 002*X5^5+4.2E-004*X2^4*X5+2.9E-003*X2^2*X3^3*X5-1.1E- 002*X2^2*X5^4+6.9E-002*X2^2*X3*X5^2+5.E-003*X3^6*X5-3.9E- 002*X3^3*X5^4+0.24*X3^4*X5^2+7.7E-002*X5^7- 0.93*X3*X5^5+6.E-002*X4*X5^2+1.9E-002*X2^2*X4*X5- 0.26*X4*X5^4+0.38*X1*X4*X5^2+1.6E-002*X1*X2^2*X3*X5+1.1E- 002*X1*X2^2*X5^2+3.7E-002*X1*X2^2*X4*X5+5.6E- 002*X1*X3^4*X5+3.8E-002*X1*X3^3*X5^2+0.13*X1*X3^3*X4*X5- 0.22*X1*X3*X5^4-0.15*X1*X5^5- 0.5*X1*X4*X5^4+1.4*X1*X3^2*X5^2+3.1*X1*X3*X4*X5^2-4.3E- 003*X2^2*X3^2*X5-1.5E-002*X3^5*X5+5.7E-002*X3^2*X5^4- 0.64*X1^2*X4*X5^2-2.7E-002*X1^2*X2^2*X3*X5-1.9E- 002*X1^2*X2^2*X5^2-6.2E-002*X1^2*X2^2*X4*X5-9.3E- 002*X1^2*X3^4*X5-6.4E-002*X1^2*X3^3*X5^2- 0.21*X1^2*X3^3*X4*X5+0.37*X1^2*X3*X5^4+0.25*X1^2*X5^5+0.8 4*X1^2*X4*X5^4-2.3*X1^2*X3^2*X5^2- 5.2*X1^2*X3*X4*X5^2+0.32*X1^3*X4*X5^2+1.4E- 002*X1^3*X2^2*X3*X5+9.4E-003*X1^3*X2^2*X5^2+3.1E- 002*X1^3*X2^2*X4*X5+4.7E-002*X1^3*X3^4*X5+3.2E- 002*X1^3*X3^3*X5^2+0.11*X1^3*X3^3*X4*X5- 0.18*X1^3*X3*X5^4-0.13*X1^3*X5^5- 0.42*X1^3*X4*X5^4+1.2*X1^3*X3^2*X5^2+2.6*X1^3*X3*X4*X5^2- 6.E-002*X4^4*X5^2-2.4E-003*X2^2*X3*X4^3*X5-1.8E- 003*X2^2*X4^3*X5^2-5.9E-003*X2^2*X4^4*X5-8.3E- 003*X3^4*X4^3*X5-6.1E-003*X3^3*X4^3*X5^2-2.E- 002*X3^3*X4^4*X5+3.3E-002*X3*X4^3*X5^4+2.4E- 002*X4^3*X5^5+7.9E-002*X4^4*X5^4-0.21*X3^2*X4^3*X5^2- 0.49*X3*X4^4*X5^2+3.9E-002*X3*X4^2*X5^2+1.3E- 003*X2^2*X3^2*X4*X5+2.2E-002*X2^2*X3*X4*X5^2+5.5E- 003*X2^2*X3*X4^2*X5+4.3E-003*X3^5*X4*X5+7.5E- 002*X3^4*X4*X5^2+1.9E-002*X3^4*X4^2*X5-1.7E- 002*X3^2*X4*X5^4-0.29*X3*X4*X5^5-7.4E- 002*X3*X4^2*X5^4+0.11*X3^3*X4*X5^2+0.45*X3^2*X4^2*X5^2+1. 3E-003*X2^4*X4+9.E-003*X2^2*X3^3*X4-3.5E- 002*X2^2*X4*X5^3+1.5E-002*X3^6*X4- 0.12*X3^3*X4*X5^3+0.24*X4*X5^6-1.E-005*X4^2-2.2E- 002*X4^2*X5-2.1E-003*X2^2*X4^2-0.26*X3^3*X4^2+2.8E- 002*X4^2*X5^3+5.E-005*X1*X4^2-1.6*X1*X3^2*X4- 0.34*X1*X3*X4^2+0.11*X1*X4^2*X5+5.E- 002*X1*X2^2*X3*X4+1.E- 002*X1*X2^2*X4^2+0.17*X1*X3^4*X4+3.6E-002*X1*X3^3*X4^2- 0.4*X1*X3*X4*X5^3- 0.14*X1*X4^2*X5^3+4.1*X1*X3^2*X4*X5+0.85*X1*X3*X4^2*X5- 8.4E-005*X1^2*X4^2+2.7*X1^2*X3^2*X4+0.57*X1^2*X3*X4^2- 0.18*X1^2*X4^2*X5-8.4E-002*X1^2*X2^2*X3*X4-1.7E-

Best formula (cont'd)	002*X1^2*X2^2*X4^2-0.29*X1^2*X3^4*X4-6.E- 002*X1^2*X3^3*X4^2+0.67*X1^2*X3*X4*X5^3+0.23*X1^2*X4^2*X 5^3-6.9*X1^2*X3^2*X4*X5-1.4*X1^2*X3*X4^2*X5+4.2E- 005*X1^3*X4^2-1.4*X1^3*X3^2*X4-0.29*X1^3*X3*X4^2+9.E- 002*X1^3*X4^2*X5+4.2E-002*X1^3*X2^2*X3*X4+8.8E- 003*X1^3*X2^2*X4^2+0.15*X1^3*X3^4*X4+3.E- 002*X1^3*X3^3*X4^2-0.34*X1^3*X3*X4*X5^3- 0.12*X1^3*X4^2*X5^3+3.5*X1^3*X3^2*X4*X5+0.72*X1^3*X3*X4^ 2*X5-7.9E-006*X4^5+0.26*X3^2*X4^4+5.4E-002*X3*X4^5-1.7E- 002*X4^5*X5-7.9E-003*X2^2*X3*X4^4-1.6E-003*X2^2*X4^5-2.7E- 002*X3^4*X4^4-5.7E-003*X3^3*X4^5+6.3E- 002*X3*X4^4*X5^3+2.2E-002*X4^5*X5^3-0.65*X3^2*X4^4*X5- 0.13*X3*X4^5*X5+7.7E-003*X2^2*X3^2*X4^2+2.7E- 002*X3^5*X4^2-6.1E- 002*X3^2*X4^2*X5^3+0.63*X3^3*X4^2*X5+1.9E- 003*X2^2*X4*X5^2-2.5E-002*X4*X5^5+1.2E- 004*X2^4*X4*X5+8.5E-004*X2^2*X3^3*X4*X5-3.3E- 003*X2^2*X4*X5^4+1.5E-003*X3^6*X4*X5-1.1E- 002*X3^3*X4*X5^4+2.2E-002*X4*X5^7-2.E-003*X4^2*X5^2-2.E- 004*X2^2*X4^2*X5+2.7E-003*X4^2*X5^4+1.E- 002*X1*X4^2*X5^2+4.7E-003*X1*X2^2*X3*X4*X5+3.2E- 003*X1*X2^2*X4*X5^2+9.8E-004*X1*X2^2*X4^2*X5+1.6E- 002*X1*X3^4*X4*X5+1.1E-002*X1*X3^3*X4*X5^2+3.4E- 003*X1*X3^3*X4^2*X5-6.3E-002*X1*X3*X4*X5^4-4.3E- 002*X1*X4*X5^5-1.3E- 002*X1*X4^2*X5^4+0.4*X1*X3^2*X4*X5^2+8.3E- 002*X1*X3*X4^2*X5^2-1.7E-002*X1^2*X4^2*X5^2-7.9E- 003*X1^2*X2^2*X3*X4*X5-5.4E-003*X1^2*X2^2*X4*X5^2-1.6E- 003*X1^2*X2^2*X4^2*X5-2.7E-002*X1^2*X3^4*X4*X5-1.9E- 002*X1^2*X3^3*X4*X5^2-5.7E- 003*X1^2*X3^3*X4^2*X5+0.11*X1^2*X3*X4*X5^4+7.3E- 002*X1^2*X4*X5^5+2.2E-002*X1^2*X4^2*X5^4- 0.67*X1^2*X3^2*X4*X5^2-0.14*X1^2*X3*X4^2*X5^2+8.5E- 003*X1^3*X4^2*X5^2+4.E-003*X1^3*X2^2*X3*X4*X5+2.7E- 003*X1^3*X2^2*X4*X5^2+8.3E-004*X1^3*X2^2*X4^2*X5+1.4E- 002*X1^3*X3^4*X4*X5+9.4E-003*X1^3*X3^3*X4*X5^2+2.9E- 003*X1^3*X3^3*X4^2*X5-5.4E-002*X1^3*X3*X4*X5^4-3.7E- 002*X1^3*X4*X5^5-1.1E- 002*X1^3*X4^2*X5^4+0.34*X1^3*X3^2*X4*X5^2+7.1E- 002*X1^3*X3*X4^2*X5^2-1.6E-003*X4^5*X5^2-7.5E- 004*X2^2*X3*X4^4*X5-5.1E-004*X2^2*X4^4*X5^2-1.6E- 004*X2^2*X4^5*X5-2.6E-003*X3^4*X4^4*X5-1.8E- 003*X3^3*X4^4*X5^2-5.4E-004*X3^3*X4^5*X5+1.E- 002*X3*X4^4*X5^4+6.9E-003*X4^4*X5^5+2.1E-003*X4^5*X5^4- 6.4E-002*X3^2*X4^4*X5^2-1.3E-002*X3*X4^5*X5^2+7.3E- 004*X2^2*X3^2*X4^2*X5+5.E-004*X2^2*X3*X4^2*X5^2+2.5E- 003*X3^5*X4^2*X5+1.7E-003*X3^4*X4^2*X5^2-9.8E- 003*X3^2*X4^2*X5^4-6.7E-003*X3*X4^2*X5^5+6.2E- 002*X3^3*X4^2*X5^2

Network type	1st - FCPSE	2nd – PSE	3rd - MDL	4th - GCV	5th - FPE	6th - Regularity
Most significant variables	ECON	ECON	ECON	ECON	ECON	ECON
	SIZE	PROCU	PROCU	PROCU	PROCU	PROCU
	RELAT	SIZE	SIZE	SIZE	SIZE	SIZE
	STA_PLT	RELAT	RELAT	RELAT	RELAT	RELAT
		STA_PLT	STA_PLT	STA_PLT	STA_PLT	STA_PLT

Note : Legend for the best formula :

X1=(ECON-85.32)/26.43

X2=(PROCU-1.)/4.

X3=(SIZE-3500.)/179500.

X4=(RELAT-1.)/4.

X5=(STA_PLT-1.)/4.

Y=(OUT_COST-427000.)/129373000.

Table I-5 Cross-validation Results for GMDH Model Using FCPSE Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.9065	0.9198	0.889	0.9305	0.8998	0.9075	0.927	0.9137
Mean squared error	96,628,108,314,050	83,927,902,303,196	115,828,464,756,844	67,732,069,384,486	94,114,845,214,971	96,868,195,863,529	78,261,476,474,985	87,934,654,705,388
Mean absolute error	7,750,535	7,302,844	8,340,632	6,368,402	7,397,465	7,703,635	6,791,883	7,180,001
Min. absolute error	144,433	52,887	180,267	49,998	6,911	378,837	119,195	271,582
Max. absolute error	22,659,490	26,961,999	30,855,130	21,731,215	24,203,664	22,690,365	31,678,713	30,366,407
Correlation coefficient r	0.9522	0.9592	0.9429	0.9646	0.9486	0.9527	0.9629	0.956
Production								
R squared	0.828	0.8822	0.8396	0.8069	0.9405	0.8517	0.7046	0.8843
Mean squared error	118,190,420,042,827	63,242,678,545,074	104,412,855,742,974	203,238,437,174,950	87,799,635,894,723	103,407,491,626,698	147,674,522,945,427	101,376,675,959,642
Mean absolute error	9,818,486	6,691,384	5,545,989	11,683,792	7,311,915	9,245,994	11,309,203	9,150,638
Min. absolute error	1,353,755	1,739,505	336,894	1,322,596	2,033,485	2,640,984	1,839,300	2,880,544
Max. absolute error	14,131,199	14,199,580	27,209,695	23,880,706	19,261,929	16,046,228	18,867,893	14,068,119
Correlation coefficient r	0.9302	0.941	0.9366	0.9104	0.976	0.9252	0.8686	0.9499

Table I-6 Cross-validation Results for GMDH Models Using PSE Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.9146	0.9157	0.9194	0.9268	0.9105	0.9079	0.9023	0.8981
Mean squared error	88,265,904,636,427	88,299,376,832,754	84,123,151,267,378	71,251,291,359,856	84,110,761,711,154	96,481,600,907,377	104,713,078,125,164	103,764,386,770,077
Mean absolute error	7,045,217	7,444,038	7,156,103	6,707,452	7,227,005	7,580,252	7,566,966	7,624,351
Min. absolute error	56,476	44,630	27,110	212,900	29,074	83,391	181,410	396,315
Max. absolute error	24,873,728	25,515,876	25,227,442	22,304,257	25,431,088	24,386,541	31,561,024	29,814,800
Correlation coefficient r	0.9564	0.9571	0.9588	0.9628	0.9542	0.9528	0.9499	0.9477
Production								
R squared	0.7664	0.808	0.8852	0.7673	0.9448	0.8275	0.7112	0.8188
Mean squared error	160,549,153,814,202	103,054,068,731,545	74,700,866,744,471	244,952,550,337,235	81,493,629,271,844	120,290,862,930,733	144,379,422,498,112	158,778,346,365,835
Mean absolute error	10,644,625	7,391,494	5,851,929	13,439,311	6,418,665	9,331,632	10,774,962	10,673,332
Min. absolute error	3,010,337	155,141	162,546	2,330,104	362,392	1,879,576	5,421,788	2,519,700
Max. absolute error	24,592,317	20,258,861	20,406,387	24,701,216	20,273,442	20,783,604	20,942,935	24,391,989
Correlation coefficient r	0.8863	0.9029	0.9644	0.8837	0.9785	0.9108	0.853	0.9138

Table I-7 Cross-validation Results for GMDH Model Using MDL Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.9302	0.9356	0.9389	0.9392	0.9185	0.9345	0.937	0.9309
Mean squared error	72,161,874,290,064	67,397,355,281,117	63,707,166,947,642	59,175,382,438,581	76,618,537,608,851	68,605,957,133,251	67,505,188,265,713	70,336,739,438,234
Mean absolute error	6,481,934	6,279,491	6,282,306	6,010,714	6,826,787	6,448,541	6,329,852	6,421,901
Min. absolute error	185,919	54,120	84,110	52,768	96,753	107,742	92,864	78,910
Max. absolute error	21,788,201	21,205,903	20,656,178	22,151,312	21,289,053	24,190,304	22,225,652	20,648,844
Correlation coefficient r	0.9645	0.9673	0.9691	0.9691	0.9584	0.9667	0.968	0.9651
Production								
R squared	0.8659	0.8385	0.8737	0.7878	0.934	0.8759	0.7292	0.9296
Mean squared error	92,176,753,984,719	86,657,707,548,051	82,211,190,224,813	223,337,831,749,900	97,322,359,827,306	86,532,655,397,855	135,369,508,567,040	61,648,394,323,344
Mean absolute error	8,141,568	7,106,751	4,967,768	12,977,082	6,798,350	7,587,424	9,595,981	7,155,309
Min. absolute error	211,492	801,988	358,125	3,439,836	56,466	249,688	599,084	1,467,476
Max. absolute error	15,423,031	20,625,163	24,336,975	25,729,312	20,929,960	18,949,434	21,322,860	10,796,136
Correlation coefficient r	0.9427	0.9169	0.9391	0.8967	0.9823	0.936	0.8681	0.9699

Table I-8 Cross-validation Results for GMDH Model Using GCV Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.9417	0.9358	0.9407	0.9534	0.9345	0.9358	0.9477	0.9432
Mean squared error	60,264,152,045,823	67,241,588,790,417	61,832,728,345,976	45,346,273,115,585	61,539,582,227,150	67,202,633,441,772	56,098,606,022,636	57,857,372,289,132
Mean absolute error	5,832,177	6,090,362	5,895,982	4,928,787	5,931,372	6,289,637	5,559,289	5,607,015
Min. absolute error	48,788	79,714	131,754	0	366,766	46,922	40,007	297,498
Max. absolute error	20,643,961	22,347,404	21,549,300	23,807,982	20,647,444	24,234,000	22,179,956	21,099,041
Correlation coefficient r	0.9704	0.9674	0.9699	0.9765	0.9667	0.9674	0.9735	0.9712
Production								
R squared	0.7816	0.9047	0.8991	0.7139	0.9743	0.8693	0.7088	0.9307
Mean squared error	150,060,455,785,692	51,146,060,446,240	65,652,570,672,600	301,169,803,265,561	37,937,739,910,371	91,169,107,543,260	145,593,469,380,336	60,769,354,966,000
Mean absolute error	10,184,228	5,447,116	5,285,121	13,637,132	3,704,107	7,827,677	9,763,643	6,945,354
Min. absolute error	1,195,683	852,136	446,001	784,991	55,891	1,676,056	936,192	1,073,181
Max. absolute error	23,724,773	13,493,160	20,150,147	28,184,676	16,382,450	20,569,918	21,608,315	12,441,330
Correlation coefficient r	0.891	0.9598	0.966	0.8535	0.9977	0.9326	0.8711	0.9729

Table I-9 Cross-validation Results for GMDH Model Using FPE Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.9356	0.9119	0.9295	0.938	0.8923	0.9258	0.9346	0.9078
Mean squared error	66,594,400,284,232	92,248,632,609,073	73,580,641,994,730	60,405,012,938,895	101,192,898,332,396	77,750,293,230,246	70,116,975,725,816	93,925,980,452,790
Mean absolute error	6,372,931	7,388,831	6,893,770	6,053,983	7,775,816	6,602,703	6,440,270	7,491,222
Min. absolute error	55,243	42,365	140,386	43,956	45,829	2,821	84,481	196,888
Max. absolute error	21,041,168	24,096,567	22,129,655	23,971,984	25,175,005	24,938,695	25,523,338	25,043,036
Correlation coefficient r	0.9673	0.955	0.9641	0.9685	0.9449	0.9622	0.9667	0.9528
Production								
R squared	0.7274	0.9063	0.8752	0.8068	0.9012	0.8316	0.7016	0.8695
Mean squared error	187,331,138,967,875	50,299,237,699,637	81,247,415,141,426	203,378,053,656,008	145,759,953,244,996	117,450,438,894,590	149,186,701,841,711	114,330,101,667,652
Mean absolute error	12,085,351	5,707,220	5,073,410	11,454,530	10,072,500	9,533,584	10,481,263	9,101,942
Min. absolute error	5,512,521	1,680,376	73,241	928,596	759,889	3,433,988	219,668	862,306
Max. absolute error	26,345,763	14,824,087	24,036,488	25,923,963	22,986,786	20,174,363	20,402,058	20,890,635
Correlation coefficient r	0.8609	0.9636	0.9426	0.9028	0.9736	0.9148	0.8514	0.9445

Table I-10 Cross-validation Results for GMDH Model Using Regularity Selection Criteria

Network type	Set A	Set B	Set C	Set D	Set E	Set F	Set G	Set H
Training								
R squared	0.8806	0.8803	0.7778	0.8053	0.7392	0.7453	0.8817	0.9002
Mean squared error	125,502,469,902,277	127,767,162,302,539	235,831,929,360,008	189,462,692,210,215	246,397,922,743,118	270,513,267,372,763	128,911,077,601,678	87,382,438,078,582
Mean absolute error	8,272,743	7,952,472	12,025,332	10,601,225	11,469,322	12,497,721	9,035,736	7,032,199
Min. absolute error	5,347	151,121	78,141	135,425	0	3,518	96,880	1,276
Max. absolute error	32,036,170	33,078,463	36,644,983	40,650,660	47,894,976	52,331,179	33,506,956	23,561,546
Correlation coefficient r	0.9463	0.9445	0.8819	0.8974	0.868	0.8633	0.9392	0.9488
Testing								
R squared	0.9883	0.9875	0.9827	0.9681	0.9519	0.963	0.9866	0.9996
Mean squared error	8,900,196,902,363	9,467,407,996,855	13,123,778,528,802	24,200,134,397,770	36,503,060,885,279	28,048,672,967,408	10,133,824,581,575	695,837,342,632
Mean absolute error	2,305,949	2,436,864	2,867,141	3,736,350	5,439,402	4,366,754	2,330,643	603,994
Min. absolute error	782,779	933,155	87,556	395,990	708,229	1,155,635	509,514	113,127
Max. absolute error	6,586,369	6,463,981	6,230,165	8,079,221	7,771,331	10,180,544	7,059,453	1,869,245
Correlation coefficient r	0.9954	0.996	0.9924	0.9849	0.978	0.9909	0.9936	0.9998
Production								
R squared	0.8315	0.8266	0.606	0.493	0.9467	0.7357	0.6615	0.877
Mean squared error	115,797,491,618,688	93,047,596,134,121	256,455,219,617,689	533,695,386,874,855	78,566,489,421,195	184,349,845,350,472	169,211,015,012,182	107,787,365,838,374
Mean absolute error	9,278,768	7,790,395	12,186,028	17,436,494	6,663,307	10,232,762	9,554,898	8,563,385
Min. absolute error	2,831,298	83,931	711,146	2,761,759	239,615	2,811,980	1,023,722	1,790,633
Max. absolute error	18,921,601	20,294,817	36,118,913	40,259,921	21,015,096	29,184,421	31,282,598	21,918,970
Correlation coefficient r	0.9273	0.9147	0.8777	0.7721	0.9869	0.8745	0.8164	0.9484